Cartographier les données des Enquêtes Démographiques et de Santé à partir des coordonnées des zones d'enquête

Joseph Larmarange, Yaro Seydou, Vallo Roselyne, Msellati Philippe, Méda Nicolas, Benoît Ferry

▶ **To cite this version:**

Joseph Larmarange, Yaro Seydou, Vallo Roselyne, Msellati Philippe, Méda Nicolas, et al.. Cartographier les données des Enquêtes Démographiques et de Santé à partir des coordonnées des zones d'enquête. Chaire Quételet 2006, Dec 2006, Louvain la Neuve, Belgique. hal-04118141

HAL Id: hal-04118141
https://hal.science/hal-04118141v1

Submitted on 6 Jun 2023

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.

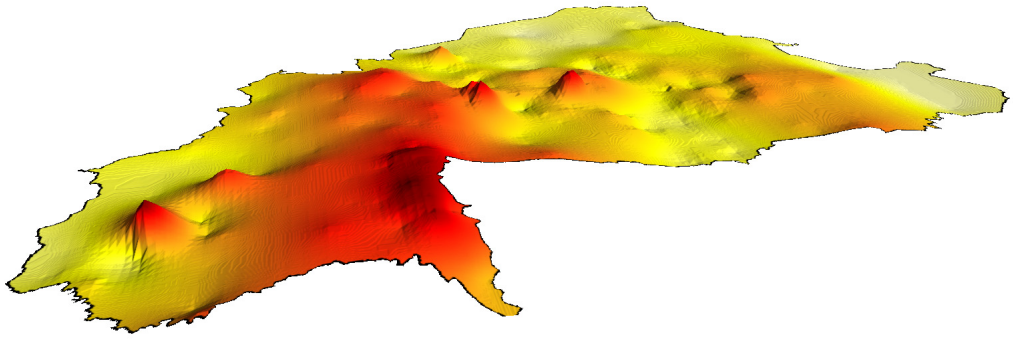
Chaire Quételet 2006

Cartographier les données des Enquêtes Démographiques et de Santé à partir des coordonnées des zones d'enquête

LARMARANGE Joseph, YARO Seydou, VALLO Roselyne, MSELLATI Philippe, MÉDA Nicolas, FERRY Benoît

Séance :

Les techniques récentes d'enquête, d'échantillonnage et de traitement



Coordonnées des auteurs :

LARMARANGE Joseph
Laboratoire POPINTER (Université Paris 5) / LPED UMR 151 (IRD / Université de Provence)
Popinter - 45 rue des Saints Pères – 75006 PARIS - France
Gsm : +33 6 62 06 51 82 – Fax : +33 1 42 86 33 50
joseph.larmarange@paris5.sorbonne.fr

YARO Seydou
Centre Muraz
01 BP 390 - Bobo-Dioulasso, Bukina Faso
Fax : +226 20 97 04 57
yaro_seydou@yahoo.com

VALLO Roselyne
Coopération Française / Centre de Calcul du Centre Muraz
roselyne_vallo@yahoo.fr

MSELLATI Philippe
IRD / Centre de Calcul du Centre Muraz
msellati@ird.bf

MÉDA Nicolas
Centre Muraz
nmeda.muraz@fasonet.bf

FERRY Benoît
LPED UMR 151 (IRD / Université de Provence) / CEPED
Ceped – 45 bis Avenue de la Belle Gabrielle – 94736 Nogent sur Marne Cedex – France
Tél. : +33 1 43 94 72 90 – Fax : +33 1 43 94 72 92
ferry@ceped.cirad.fr

Remerciements :

Les auteurs souhaitent remercier l'Agence Nationale de Recherche sur le Sida qui soutient financièrement le projet 12114 dans lequel s'inscrit le présent travail.

Introduction

Les Enquêtes Démographiques et de Santé (EDS) constituent une source de données standardisées en population générale bien connue des démographes. Conduites dans la majorité des pays en développement depuis le milieu des années 80, nombre d'entre elles incorporent depuis longtemps les coordonnées latitude/longitude des zones enquêtées (grappes). La collecte des coordonnées des grappes par GPS s'est standardisée dans les années 1990. Elle est aujourd'hui systématique pour toutes les enquêtes récentes.

Cependant, malgré l'utilisation de plus en plus fréquente des Systèmes d'Informations Géographiques (SIG), peu d'études portant sur les EDS ont eu recours à ce type d'information. Le plus souvent, les coordonnées géographiques ont été utilisées pour calculer des indicateurs spatiaux (tels qu'une distance à une route ou à une structure de santé), indicateurs réutilisés par la suite dans des analyses univariées ou multivariées.

Les travaux cartographiques réalisés à partir des EDS ont porté, pour la plupart, sur un échelon régional, comme c'est le cas du récent *Atlas Tanzanien des Indicateurs sur le VIH/Sida*[1] publié en janvier 2006. Cela s'explique en partie pour des raisons méthodologiques. En effet, le nombre de grappes enquêtées dans une EDS se situe généralement entre 400 et 500 et le nombre de personnes enquêtées par grappes est très faible (moins de quarante dans la majorité des cas). Ces effectifs sont trop réduits pour calculer, à partir des questionnaires individuels, des indicateurs statistiquement significatifs par grappe. On ne peut alors appliquer les méthodes d'interpolation spatiale classique[1] (telle que le krigeage ou la pondération par l'inverse de la distance) qui présuppose une mesure précise du phénomène étudié en chaque point connu.

Le présent travail a été initié pour représenter les variations spatiales de la prévalence du VIH à partir des EDS. Cependant, la méthodologie présentée peut être appliquée à d'autres indicateurs. Le terme de prévalence doit être considéré ici dans son sens le plus général, c'est-à-dire comme la proportion d'individus qui, dans une population donnée, présentent une caractéristique particulière. Ainsi, le quotient de mortalité infantile peut être considéré comme la prévalence du décès avant l'âge de un an exact dans la population des enfants étudiés.

Les Enquêtes Démographiques et de Santé

Depuis 1984, plus de 200 Enquêtes Démographiques et de Santé (EDS) ont été menées dans 75 pays à travers le monde [2]. Elles sont l'un des quatre composants du programme "Monitoring and Evaluation to Assess and Use Results" (MEASURE) fondé en 1997 par l'USAID et mis en œuvre par ORC Macro International.

Les EDS sont standardisées afin de permettre des comparaisons dans le temps et entre pays. Elles sont usuellement réalisées tous les cinq ans dans un même pays afin de pouvoir suivre les évolutions de la population. Il s'agit d'enquêtes auprès des ménages, représentatives au niveau national, par milieu de résidence et par région. Le nombre de ménages enquêtés par enquête se situe le plus souvent entre 5.000 et 30.000. Les EDS présentent un plan de sondage comparable dans chaque pays. Il s'agit d'enquêtes stratifiées avec un tirage à deux degrés. Le pays est divisé en plusieurs strates, une par région administrative et milieu de résidence. La

[1] L'interpolation spatiale est un processus consistant à estimer les valeurs d'une variable sur une grille régulière, à partir de points connus.

base de sondage des unités primaires est typiquement composée des zones de dénombrement au dernier recensement de la population effectué dans le pays. Au premier degré, les unités primaires ou grappes sont tirées au sort, séparément dans chaque strate, avec une probabilité proportionnelle au nombre de ménages de la grappe lors du dernier recensement de population. La répartition spatiale de ces grappes peut alors être considérée comme une approximation de la densité de la population. Après un recensement exhaustif des ménages de chaque grappe, un nombre prédéterminé de ménages est sélectionné au second degré, par tirage au sort simple, pour l'enquête ménage et le questionnaire individuel femmes (15-49 ans). Suivant le pays, seule une partie des ménages enquêtés est retenue pour le questionnaire hommes (15-59 ans) et le test de dépistage du VIH (lorsque ce module est inclus).

La majorité des EDS fournissent les coordonnées longitude/latitude du centre de chacune des grappes enquêtées. Depuis 2001, plusieurs EDS comportent un module de dépistage du VIH. Pour des raisons éthiques, il n'était pas possible de lier les résultats des tests VIH des premières enquêtes (Mali 2001, Zambie 2001-2002 et République Dominicaine 2002). Pour les enquêtes suivantes, Macro International a mis en place une procédure permettant de fournir les coordonnées géographiques des grappes enquêtées tout en garantissant l'anonymat des personnes testées. Ainsi, ces coordonnées sont décalées de manière aléatoire dans un rayon de 2 kilomètres pour les grappes urbaines et de 5 kilomètres pour les grappes rurales [3]. Ce biais peut être considéré comme négligeable à l'échelle d'un pays.

Le tableau 1 résume l'échantillonnage de plusieurs EDS récentes, réalisées en Afrique subsaharienne et comportant un module de dépistage du VIH.

Tableau 1. Échantillonnage de 13 enquêtes en Afrique subsaharienne avec dépistage du VIH

Enquête	Type*	Nombre de grappes	Personnes testées (15-49 ans)	Nombre moyen de pers. testées par grappe	Prévalence nationale du VIH (15-49 ans)
Burkina Faso 2003	EDS	400	7.151	17,9	1,8%
Cameroun 2004	EDS	466	9.900	21,2	5,5%
Ethiopie 2005	EDS	540	10.540	19,5	1,4%
Ghana 2003	EDS	412	9.144	22.2	2,2%
Guinée 2005 ‡	EDS	297	6.388	21,5	1,5%
Kenya 2003	EDS	400	6.001	15,0	6,7%
Lesotho 2004 §	EDS	405	5.043	12,5	23,5%
Malawi 2004 ‡	EDS	522	5.150	9,9	11,8%
Mali 2001 †	EDS	403	6.475	16,1	1,7%
Sénégal 2005 †† §	EDS	377	7.503	19,9	0,7%
Tanzanie 2003	AIS	345	10.747	31,2	7,0%
Uganda 2004 ‡	AIS	417	16.906	40,5	6,4%
Zambie 2001/02 †	EDS	320	3.807	11,9	15,6%

Sources : rapport final de chaque enquête, disponible sur http://www.measuredhs.com/.

* EDS : Enquête Démographique et de Santé – AIS : AIDS Indicators Survey

† Les données VIH ne peuvent être liées aux données individuelles et GPS.

†† Données GPS non collectées.

‡ Données non disponibles au 1er novembre 2006.

§ Données préliminaires disponibles au 1er novembre 2006.

Il apparaît clairement que le nombre de personnes testées en moyenne par grappe est extrêmement faible. Il en résulte que les prévalences observées par grappe traduisent essentiellement la variance de l'échantillonnage et non celle de l'épidémie. Il est donc nécessaire d'estimer la prévalence de chaque grappe avant de procéder à une interpolation spatiale.

Méthode

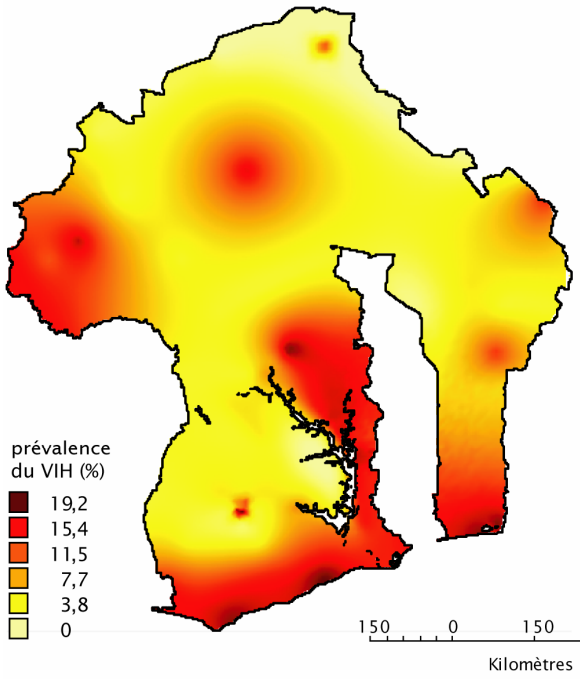
Dans le domaine du VIH, aucun pays ne dispose d'une enquête à grande échelle avec une mesure fine des prévalences locales. En l'absence de données de référence permettant de valider une approche méthodologique reconstituant les variations spatiales de la prévalence, nous avons eu recours à l'élaboration d'un pays modèle sur lequel il est possible de procéder à la simulation d'EDS et de comparer les écarts entre épidémie du modèle et épidémie reconstituée.

Élaboration d'un pays modèle

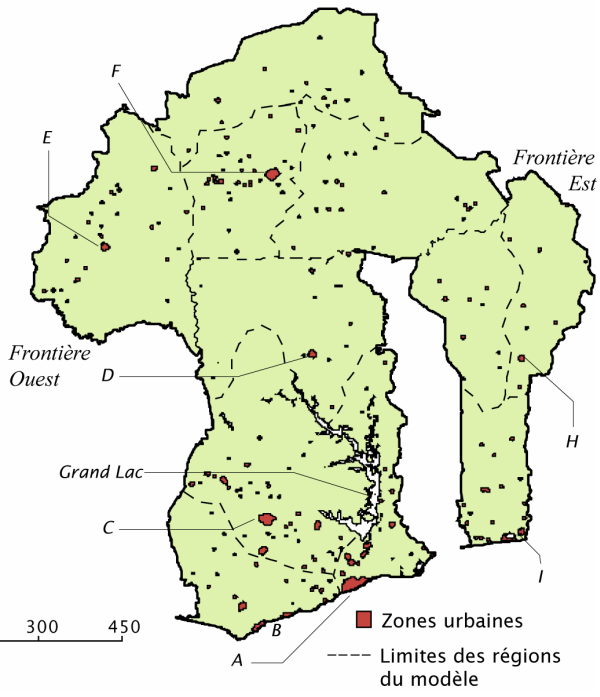
Ce pays résulte de l'agrégation du Bénin, du Burkina Faso et du Ghana, en intégrant leur densité de population en 2000 ainsi que le découpage en zones rurales et urbaines du projet GRUMP [4, 5]. Le modèle a été découpé en 9.137 grappes (7.818 grappes rurales et 1.319 grappes urbaines), avec une résolution moyenne de 2 arc-minutes en milieu urbain et de 5 arc-minutes en milieu rural. La superficie et la population de chaque grappe ont été calculées, puis le modèle a été divisé en 11 régions (figure 1.b).

Figure 1. Prévalence du modèle, régions et zones urbaines

a. Prévalence du VIH du modèle (niveau national de 10%)



b. Régions et zones urbaines



Dans un second temps, une épidémie fictive a été appliquée au modèle (figure 1.a), totalement indépendante de la situation des trois pays qui ont servi de base à la construction du modèle. Cette épidémie correspondant à une prévalence nationale de 10% et a été conçue de manière à présenter différents modèles de diffusion : ville importante avec une épidémie concentrée sur celle-ci et faible prévalence autour (C) ; ville importante (F) et moyenne (H) avec diffusion progressive ; pic localisé en zone rural dans le nord du pays ; rupture de continuité de part et d'autre d'un grand lac ; gradient des côtes vers l'intérieur des terres au sud du pays, avec deux grandes agglomérations (A et I) et deux villes moyennes (B) ; diffusion depuis la frontière ouest et depuis une ville située de l'autre côté d'une frontière (à l'est).

Pour la suite des analyses, il sera possible de faire varier la prévalence nationale en multipliant la prévalence de chaque grappe par un même facteur d'échelle.

Simulation des Enquêtes Démographiques et de Santé

La simulation des Enquêtes Démographiques et de Santé est réalisée à partir de trois paramètres : la prévalence nationale, le nombre total de personnes enquêtées et le nombre de grappes au premier degré.

Elle se déroule en trois étapes. Dans un premier temps, chacune des onze régions est divisée en deux strates, l'une urbaine et l'autre rurale. Le nombre de grappes tirées par strate est proportionnel à la population totale de celle-ci. Les grappes sont tirées aléatoirement, strate par strate, avec une probabilité de sondage proportionnelle à leur effectif de population. Dans un second temps, l'effectif de personnes testées par grappe est déterminé aléatoirement, selon une loi normale, afin de reproduire la variabilité du nombre de personnes enquêtées par grappe, variabilité que l'on peut observer dans les EDS[2]. Le nombre de personnes enquêtées par grappe est ensuite redressé pour que le total corresponde à l'effectif visé. Enfin, dans un dernier temps, le nombre de personnes présentant la caractéristique étudiée dans chaque grappe est déterminé aléatoirement selon une loi binomiale ayant pour paramètre le nombre de personnes enquêtées et la prévalence de cette grappe dans le modèle.

Un taux de pondération W_{hi} est calculé pour les individus de la grappe i de la strate h selon la formule suivante, comparable à celle utilisée dans les EDS :

Équation 1

$$W_{hi} = \frac{1}{P_{1hi} P_{2hi}}$$

$$P_{1hi} = \frac{a_h \cdot pop_{hi}}{POP_h}$$

$$P_{2hi} = \frac{n_{hi}}{pop_{hi}}$$

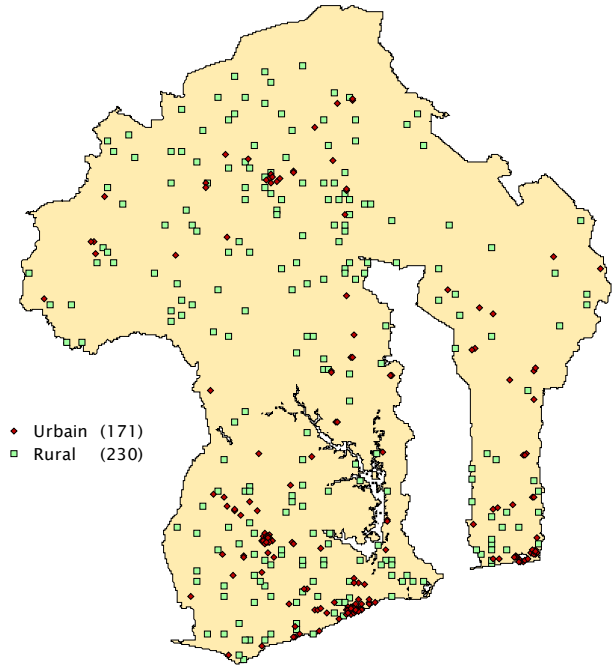
P_{1hi} représente la probabilité que la grappe hi soit incluse dans l'échantillon, P_{2hi} celle qu'un individu de la grappe hi soit enquêté, a_h le nombre de grappes sélectionnées dans la strate h, pop_{hi} l'effectif de population de la grappe hi, POP_h l'effectif total de la strate h et n_{hi} le nombre de personnes enquêtées dans la grappe hi. Ces taux de pondération sont ensuite corrigés afin que leur total corresponde à l'effectif total de l'enquête.

La figure 2 présente la répartition des grappes enquêtées pour une simulation d'une EDS avec une prévalence nationale de 10%, 8.000 personnes enquêtées et 400 grappes[3].

[2] Le nombre de personnes testées par grappe n'est pas autocorrélé spatialement (statistique de Geary non significative sur les données des dernières EDS du Burkina Faso, du Cameroun, du Kenya et de la Tanzanie, à l'exception du Ghana qui présente une autocorrélation spatiale positive).

[3] Le nombre total de grappes s'avère être de 401 en raison d'arrondis dans le calcul du nombre de grappes par strate.

Figure 2. Répartition des grappes d'une simulation d'une EDS



Estimation de la prévalence de chaque grappe

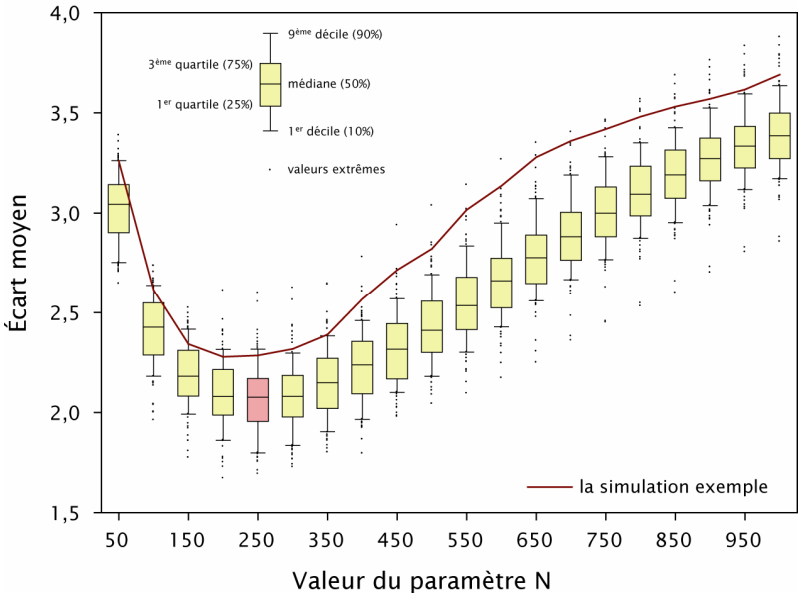
S'inspirant de techniques utilisées pour le calcul de tendances régionales[6-9], notre approche consiste à tracer un cercle autour de chaque grappe dont le rayon est tel que le nombre de personnes enquêtées[4] à l'intérieur de ce cercle atteigne un effectif minimum noté N. La prévalence de la grappe centrale est alors calculée sur cet ensemble d'individus, en tenant compte de leur pondération respective.

Lorsque le N paramètre augmente progressivement, les prévalences estimées sont lissées, atténuant ainsi les variations aléatoires de l'échantillon. Dans le même temps, elles tendent progressivement vers une valeur unique, la prévalence nationale, atteinte lorsque N est supérieur ou égale à la taille totale de l'échantillon. Il s'agit alors de déterminer un compromis minimisant suffisamment les aléas de l'échantillonnage tout en conservant une précision locale.

Dans le cadre du pays modèle, il est possible de calculer, pour une simulation donnée et une valeur de N, l'écart entre prévalence estimée et prévalence du modèle de chaque grappe, puis de calculer un écart moyen (en valeurs absolues). Pour une simulation donnée, lorsque N augmente, cet écart moyen diminue jusqu'à un minimum puis augmente à nouveau (voir courbe de la figure 3). En raison de légères variations de la valeur de N pour laquelle l'écart moyen se minimise d'une simulation à une autre, nous retiendrons la médiane de cette valeur sur 100 simulations comme valeur optimale du paramètre N, et ce pour une série donnée des paramètres de simulation, à savoir la prévalence nationale, l'effectif total et le nombre de grappes.

[4] Les données manquantes auront été préalablement supprimées.

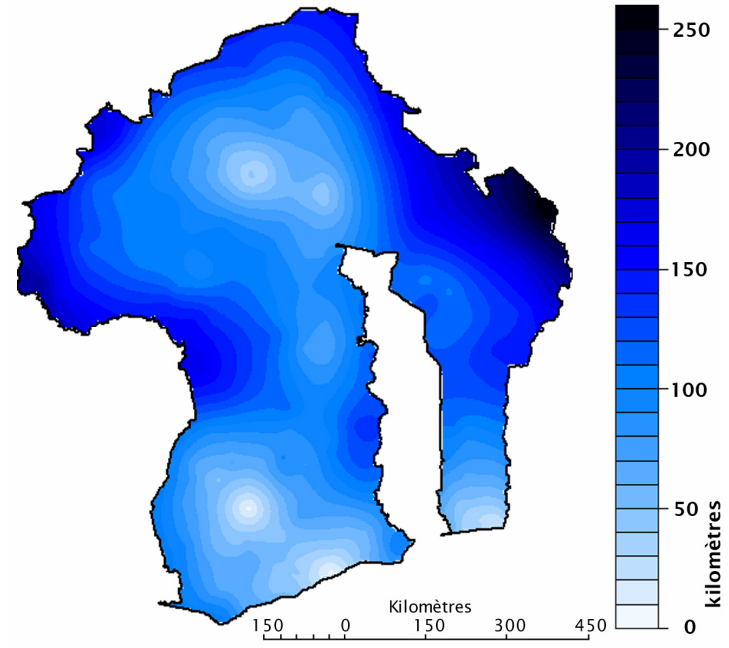
Figure 3. Écart moyen selon le paramètre N pour 100 simulations et pour celle de la figure 2



Simulations avec 8.000 personnes testées, réparties en 400 grappes, pour une épidémie globale de 10%

Le recours à un effectif minimum comme paramètre pour déterminer la taille des cercles permet à la fois de s'assurer que la prévalence de chaque grappe sera estimée sur un nombre de personnes suffisant et d'appliquer un niveau de lissage différent selon les zones enquêtées ; ce dernier étant déterminé par la superficie du cercle de lissage.

Figure 4. Rayon des cercles de lissage lorsque N vaut 250, pour la simulation de la figure 2



Ainsi, dans les régions denses et urbanisées, le rayon des cercles est faible (figure 4) et l'information représentée relativement localisée. Par contre, pour les zones peu enquêtées, notamment le long des frontières, le calcul des prévalences fait intervenir des grappes très

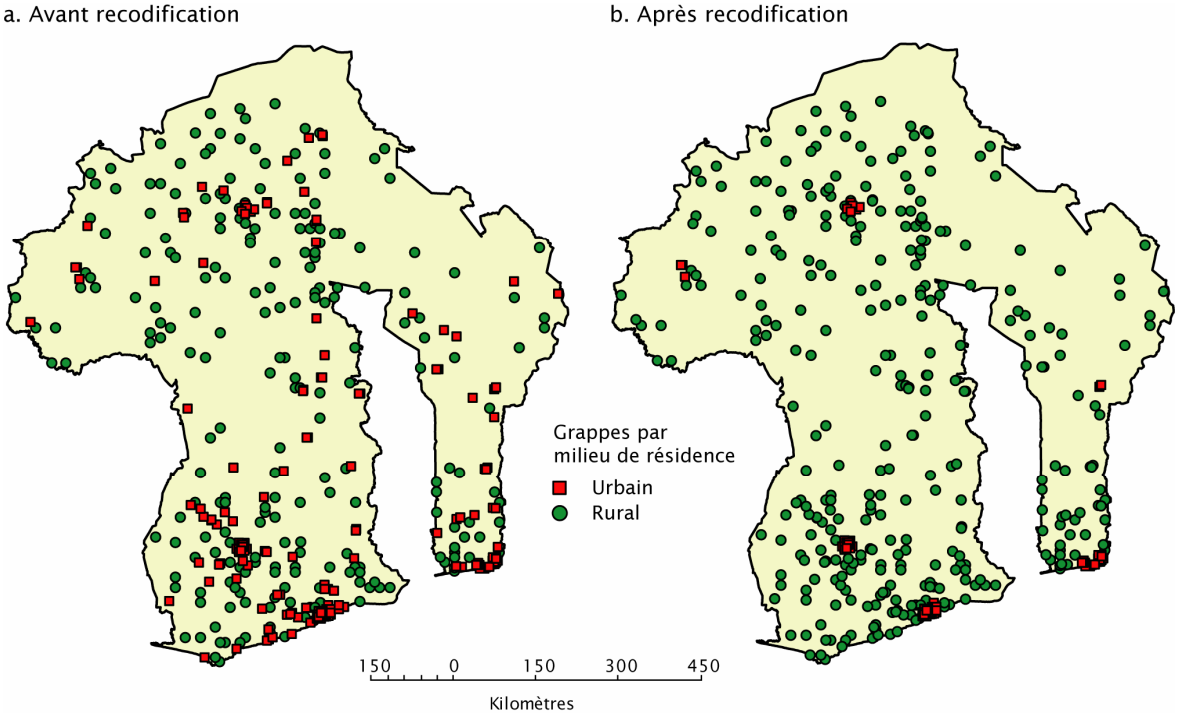
éloignées les unes des autres. Il est alors préférable de limiter le lissage à un cercle plus petit quitte à estimer les prévalences sur un nombre d'individus moindre. Afin de ne toucher que les grappes où le rayon du cercle de lissage est le plus élevé, nous limiterons la taille des cercles à la valeur du neuvième décile de l'ensemble des rayons lorsque seul le paramètre N est appliqué. Ce rayon maximum imposé aux cercles de lissage est noté R. Pour la simulation de la figure 2, R vaut 128 kilomètres.

Dans un dernier temps, il semble opportun de tenir compte des agglomérations urbaines, dans la mesure où de nombreux phénomènes présentent un différentiel marqué entre milieu urbain et milieu rural. La prévalence des grappes n'appartenant pas à une agglomération urbaine est alors calculée exclusivement à partir de grappes hors agglomérations ; à l'inverse, celle de grappes situées dans une agglomération repose uniquement sur des grappes de la même agglomération.

Dans les EDS, l'appartenance d'une grappe à une agglomération urbaine n'est pas renseignée (à l'exception des capitales usuellement considérées comme une région autonome). D'autre part, la définition du milieu de résidence des grappes est telle que des grappes isolées (correspondant à de petites villes et/ou à des chefs-lieux de subdivisions administratives) sont considérées comme urbaines. Or, il importe de ne retenir pour le lissage que des agglomérations urbaines suffisamment importantes pour que l'estimation des prévalences en leur sein ait un sens.

Pour procéder à la recodification du milieu de résidence[5], nous avons eu recours aux coordonnées longitude/latitude des principales villes fournies par le projet GRUMP[6] [10]. Après sélection des agglomérations urbaines suffisamment enquêtées, est considérée comme appartenant à une agglomération urbaine donnée, toute grappe urbaine au sens de l'EDS dont la distance entre ses coordonnées et celle de la ville considérée est inférieure à 25 kilomètres.

Figure 5. Recodification du milieu de résidence des grappes



[5] Le milieu de résidence sera alors entendu après recodification comme l'appartenance à une agglomération urbaine.
[6] Ces données peuvent être téléchargées gratuitement à l'adresse suivante : http://sedac.ciesin.columbia.edu/gpw/.

La valeur de 25 kilomètres représente une valeur seuil. Elle ne correspond pas à la taille des agglomérations mais à un compromis entre celle-ci et la distance qui sépare deux agglomérations. Elle peut être modifiée selon le contexte de chaque enquête.

Le nombre d'agglomérations urbaines considérées pour l'estimation des prévalences est noté U. La non prise en compte de ces dernières correspond donc au cas où U est nul. Plusieurs critères peuvent présider le choix des agglomérations urbaines retenues. D'une part, des agglomérations où l'effectif est relativement important et réparti sur plusieurs grappes, telles que A, I, C ou F (voir Tableau 2).

Tableau 2. Sélection des agglomérations urbaines pour la simulation de la figure 2

Agglomération	Nombre de grappes	Effectif	Prévalence observée
A	24	489	21,2 %
I	18	352	23,3 %
C	17	346	15,2 %
F	7	151	18,0 %
E	3	48	16,3 %
H	2	32	10,3 %

Le nom des agglomérations fait référence à la figure 1.b. Une grappe est considérée comme appartenant à une agglomération donnée si elle est considérée comme urbaine dans l'enquête et si elle est située à moins de 25 kilomètres du centre de l'agglomération considérée.

D'autre part, lorsque l'on dispose d'autres sources d'informations, comme par exemple la surveillance sentinelle des femmes enceintes pour la prévalence du VIH, il est envisageable d'inclure d'autres agglomérations qui, bien qu'ayant un effectif de personnes enquêtées faible, ont une prévalence observée proche de celle fournie par ces autres sources. C'est le cas par exemple de l'agglomération E (l'autre source étant dans le cas présent l'épidémie du modèle que nous connaissons par définition).

Enfin, des agglomérations où la prévalence observée dans l'EDS, bien qu'inférieure à celle fournie par d'autres sources, reste supérieure à celle estimée pour une valeur nulle de U (agglomération H par exemple). Le fait d'inclure ces agglomérations les fera ressortir par rapport à leur voisinage sur la carte finale produite, l'objectif de cette cartographie étant la mise en évidence des différentiels spatiaux et non la mesure fine des niveaux[7].

Interpolation spatiale

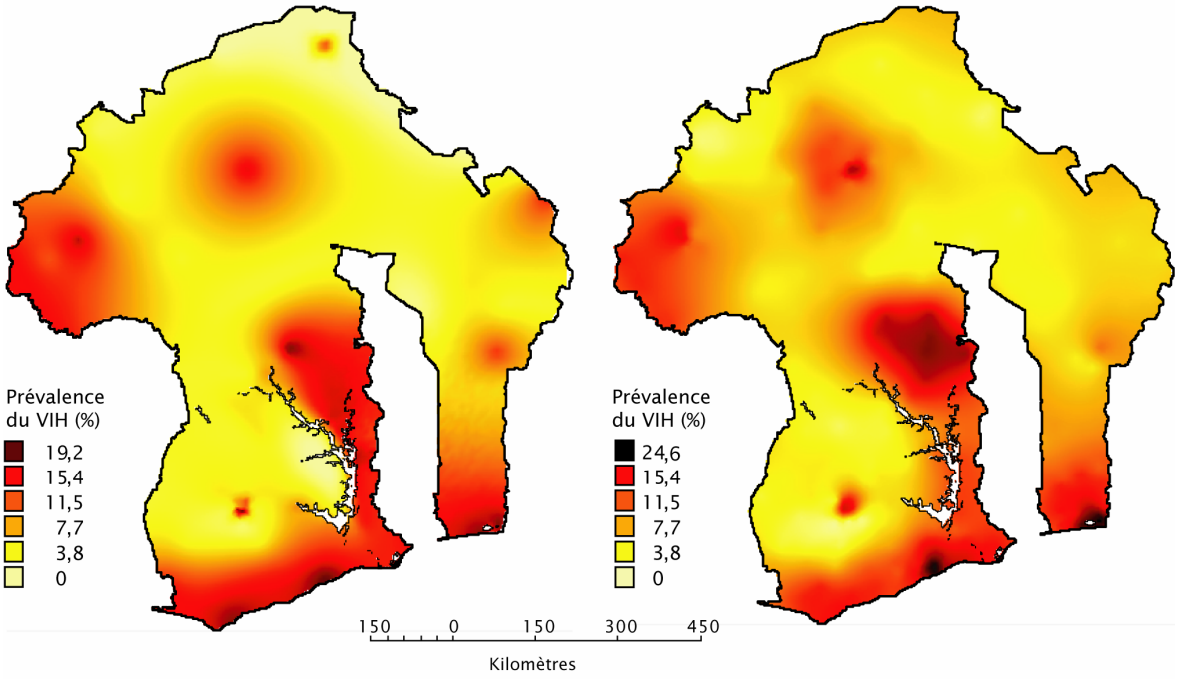
Après estimation de la prévalence de chaque grappe, la carte des prévalences est réalisée par interpolation spatiale (figure 6.b). La technique employée est celle du krigeage ordinaire [11] qui présente l'avantage de prendre en considération la structure de dépendance spatiale des données [12]. Elle est par ailleurs indiquée lorsque la répartition spatiale des points d'interpolation est irrégulière.

[7] Voir plus loin l'application aux données des EDS du Burkina Faso et du Cameroun pour une application concrète.

Figure 6. Prévalence du modèle et prévalence estimée à partir de la simulation de la figure 2

a. Prévalence du modèle (niveau national de 10%)

b. Prévalence estimée à partir d'une simulation d'EDS



Paramètres de la simulation : 8.000 personnes testées, réparties en 400 grappes avec une prévalence nationale de 10%.
Paramètres de l'estimation : N=250, R=128 km, U=6 (agglomérations A, I, C, F, E et H)

Cartes complémentaires

La qualité des estimations produites n'est pas constante à travers le pays. Dans les zones peuplées, l'information est précise du fait d'une bonne densité des grappes. Les variations infrarégionales du phénomène étudié sont alors visibles. A l'opposé, dans les zones où le nombre de personnes enquêtées est plus faible, seule une tendance régionale peut être mise en évidence. Enfin, pour les zones non enquêtées, les résultats doivent être interprétés avec prudence, les variations estimées résultant d'une interpolation à partir des zones voisines. Cette irrégularité de la qualité des estimations produites résulte, d'une part, de l'échantillonnage des EDS, conçues pour être représentatives des populations et non des territoires et, d'autre part, de l'approche méthodologique employée, à savoir un lissage adaptatif selon les zones (superficie des cercles variable).

Deux cartes complémentaires[8] sont produites afin de faciliter l'interprétation des résultats. La première représente les effectifs de personnes enquêtées par grappe par des cercles transparents de superficie proportionnelle aux effectifs. Les zones claires correspondent donc aux zones peu ou pas documentées (figure 7.b).

La seconde carte est obtenue par interpolation spatiale d'un indicateur de qualité. Il s'agit d'une aide à la lecture et à l'interprétation des résultats. La valeur de cet indicateur est calculée pour chaque grappe *g* de la manière suivante :

$$IQ = r_g^2 \Big/ \sqrt{n_g}$$

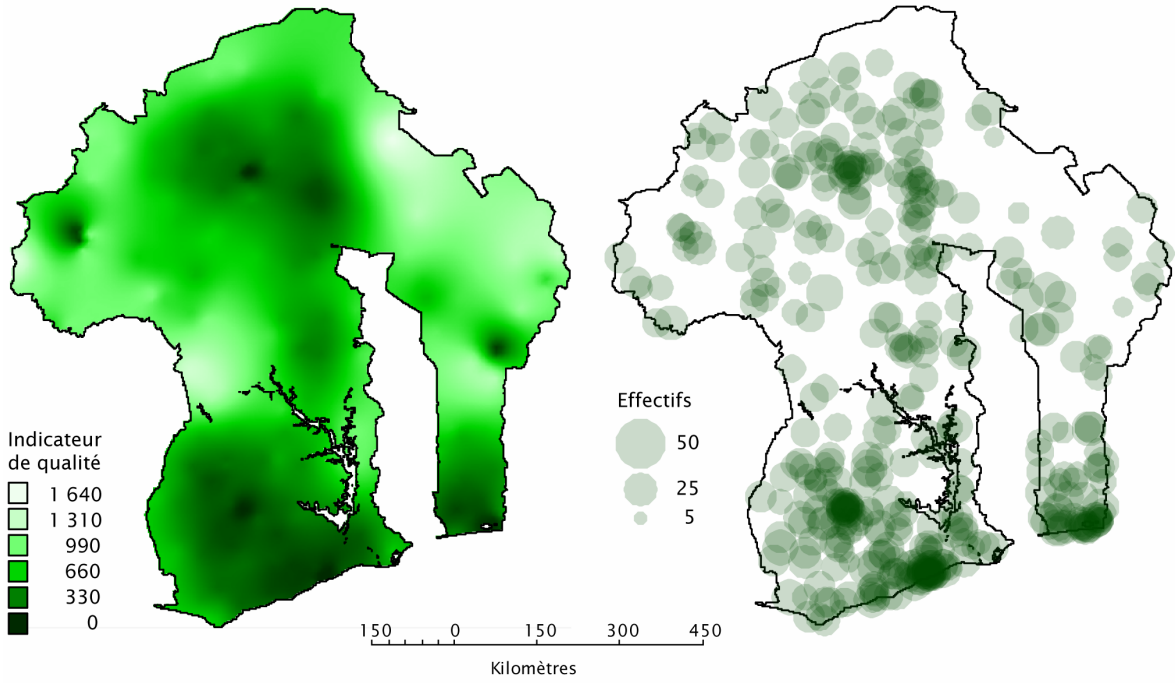
[8] Nous avons volontairement fait le choix de ne pas calculer d'intervalles de confiance de la prévalence estimée de chaque grappe. En effet, ces derniers ont un sens quand il s'agit de déterminer la précision d'un niveau. Mais comparer leur amplitude entre deux grappes est difficilement interprétable dans la mesure où celle-ci varie selon les effectifs mais aussi et surtout selon la prévalence estimée elle-même.

r_g représente le rayon du cercle de lissage pour cette grappe et n_g le nombre de personnes sur lequel la prévalence estimée de la grappe a été calculée. La qualité du lissage est inversement proportionnelle à la superficie du cercle de lissage (d'où r_g^2). Les intervalles de confiance, quant à eux, sont divisés par deux quand n_g est multiplié par quatre. Plus la valeur de cet indicateur est faible, meilleure est l'estimation. La position exacte des grappes est également représentée sur cette carte (figure 7.a).

Figure 7. Cartes complémentaires pour l'estimation de la figure 6.b

a. Indicateur de qualité

b. Nombre de personnes enquêtées par grappe



Paramètres de la simulation : 8.000 personnes testées réparties en 400 grappe avec une prévalence nationale de 10%.
Paramètre de l'estimation : N=250, R=128 km, U=6.
Note : l'indicateur de qualité est calculé pour chaque grappe selon la formule $r^2/\sqrt{n}$ où r est le rayon du cercle de lissage et n le nombre de personnes sur lesquelles la prévalence estimée de la grappe a été calculée.

Logiciel utilisé

Les analyses ont été réalisées, programmées et automatisées sous le logiciel de statistiques libre R [13], permettant ainsi une production aisée et facilitée des cartes. L'ensemble, nommé prevR et gratuit sera prochainement disponible sur internet sous la forme d'une libraire de fonctions, additionnelle à R[9]. Pour le calcul des krigeages, nous avons eu recours à la librairie gstat [14].

[9] Pour être informé de la mise en ligne de prevR, vous pouvez contacter le premier auteur par mail à joseph.larmarange@paris5.sorbonne.fr.

Résultats

Les cartes 6.a et 6.b sont présentées avec la même échelle colorimétrique. Cependant, le maximum de la prévalence estimée (figure 6.b) est supérieur à celui du modèle (figure 6.a), traduisant des variations plus marquées.

Globalement, les principales tendances du modèle ont été reconstituées. Apparaissent ainsi le gradient de la côte sud vers le nord et celui de la frontière ouest vers l'intérieur du pays. Les agglomérations A et I présentent un contraste plus marqué par rapport à l'épidémie de départ. Par contre, les deux petites agglomérations situées le long de la côte (B) ne se distinguent plus de leur voisinage, sur la carte des prévalences estimées. Les agglomérations C et F sont clairement visibles. C présente toujours une prévalence concentrée au niveau de la ville, bien que moins marquée, et le gradient autour de F est également rendu, bien que plus irrégulier. Il ressort ainsi que la prise en compte des principales agglomérations urbaines (paramètre U) permet, pour celles de taille importante, de mettre à jour la variation du phénomène étudié à leur voisinage.

De part et d'autre du grand lac, où une rupture nette avait été introduite dans le modèle, la prévalence a été surestimée à l'ouest et sous-estimée à l'est. L'estimation ne prenant pas en compte les frontières naturelles et considérant la surface géographique comme un continuum, les prévalences estimées sont uniformisées aux petites échelles.

La zone de fortes prévalences autour de D est également reconstituée, avec des écarts plus accentués par rapport au modèle. La ville D elle-même ne se distingue plus par rapport à son voisinage. Dans la mesure où il s'agit d'une petite agglomération, peu enquêtée dans cette simulation donnée, elle n'a pas été retenue pour le paramètre U, et les prévalences estimées sont celles de la tendance locale (agglomération et voisinage confondus).

Plusieurs éléments ne transparaissent plus sur la figure 6.b. Tout d'abord le pic local en milieu rural au nord du pays n'apparaît pas, du fait de l'absence de point de mesure dans cette zone (figure 7.b).

Les gradients autour de H et depuis la frontière sont guère visibles. Le nombre de grappes dans ces deux zones est faible (figure 7.b) et les valeurs de l'indicateur de qualité élevées. Le nombre de personnes enquêtées n'est pas suffisant pour reconstituer les variations locales et seule une tendance régionale peut être estimée (effet d'uniformisation).

Les cartes 7.a et 7.b présentent de par leur construction des similitudes. Il est alors possible de distinguer les zones non enquêtées où les prévalences estimées sont la poursuite des tendances des zones voisines ; les zones faiblement enquêtées où seule des variations régionales sont reconstituées, les variations locales étant alors uniformisées ; et les zones où l'information est suffisamment importante pour révéler des différentiels locaux.

Globalement, cette approche méthodologique permet bien une estimation des tendances nationales et régionales, avec un lissage plus ou moins important selon la qualité des données de chaque zone, bien que ce soit au détriment une estimation fine des niveaux en chaque point. L'analyse des différentiels est donc valable à des échelles régionales voir infrarégionales, mais perd son objectivité à une échelle locale. L'estimation réalisée est d'autant plus précise que le nombre de personnes enquêtées est élevé et que la prévalence du phénomène analysé est proche de 50 %. A l'opposé, un effet d'uniformisation important s'observe dans les zones faiblement peuplées.

Optimisation du paramètre N

Les estimations produites sont fortement dépendantes du choix du paramètre N. Il a été montré précédemment que, pour une simulation donnée, il existait une valeur de N pour laquelle les écarts entre prévalence du modèle et prévalence estimée se minimisaient.

En raison de légères variations de la valeur de N pour laquelle l'écart moyen se minimise d'une simulation à une autre, nous avons retenu la médiane[10] de cette valeur sur 100 simulations de même type[11] comme valeur optimale du paramètre N, pour ce type d'enquête.

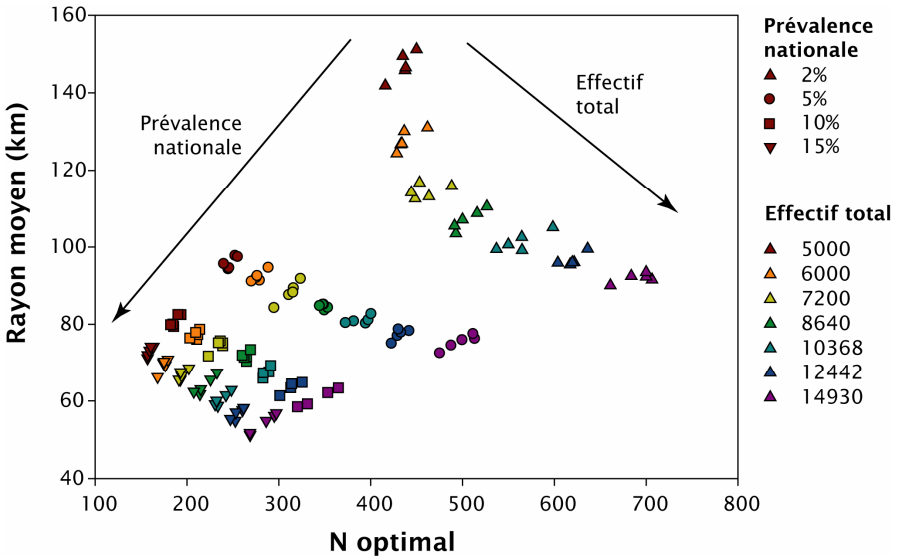
Pour comprendre comment la valeur de $N_{optimal}$ variait selon les paramètres de l'enquête, nous avons calculé les $N_{optimaux}$ pour des simulations présentant une prévalence globale de 2 %, 5 %, 10 % et 15 %, avec un effectif total de 5.000, 6.000, 7.200, 8.640, 10.368, 12.442 et 14.930, et un nombre de grappes de 300, 360, 432, 518 et 622[12]. Au total, 14.000 EDS ont été simulées, 100 par combinaison différente de ces trois paramètres.

Par ailleurs, pour chaque simulation, nous avons calculé le rayon moyen des cercles de lissage, puis le rayon moyen sur cent simulations. Ce dernier peut être considéré comme un indicateur du niveau du lissage effectué dans la mesure où, plus les cercles sont petits, plus les variations locales peuvent être mises à jour.

Sur la figure 8, chaque point correspond à 100 simulations de même type. La valeur optimale de N pour ces 100 simulations a été portée en abscisse et le rayon moyen en ordonnée. Les points de même forme correspondent à des simulations avec une même prévalence nationale et les points de même couleur à des simulations avec un même effectif total. Deux flèches, indiquant l'augmentation de ces deux paramètres, ont été ajoutées pour faciliter la lecture du graphique.

Les simulations correspondant aux points de même couleur et de même forme ne varient donc que par leur nombre de grappes. Il apparaît alors que ce dernier n'a que peu d'influence sur le rayon moyen et sur la valeur de $N_{optimal}$.

Figure 8. Rayon moyen des cercles de lissages selon $N_{optimal}$ pour différents types de simulation



[10] Plus précisément, il s'agit de la médiane de groupes, qui considère chaque valeur comme un centre de classes.
[11] Même prévalence globale, même effectif total et même nombre de grappes.
[12] Les valeurs des effectifs totaux et du nombre de grappes ont été choisies de telle manière que leur rapport soit constant. Ainsi, 5.000/300≈6.000/360≈7.200/432, etc.

Lorsque la prévalence nationale est plus élevée[13], N est optimisé pour des valeurs plus faibles, les variations aléatoires de l'échantillonnage étant moins élevées. Il en résulte que le lissage s'effectue sur des cercles de plus petite taille. Lorsque l'effectif total augmente, la valeur optimale du paramètre N augmente également. Dans le même temps, le rayon moyen des cercles de lissage diminue. Les prévalences estimées sont donc à la fois plus précises, car calculées sur un nombre de personnes plus important, et plus localisées, le lissage s'effectuant sur des zones plus petites.

En pratique, lorsque l'on applique cette méthodologie à des données d'enquêtes, il n'est pas possible de calculer les écarts entre prévalence réelle et prévalence estimée, dans la mesure où la prévalence réelle est alors inconnue et que c'est elle que l'on cherche à estimer. Une possibilité consiste à utiliser, par analogie, la valeur $N_{optimale}$ obtenue sur le pays modèle après y avoir simulé des enquêtes ayant le même échantillonnage que celui de nos données. Or, la simulation d'une centaine d'enquêtes et le calcul des écarts moyens correspondants nécessite plus d'une dizaine d'heures de calcul sur un ordinateur de bureau. Il serait alors pratique de pouvoir déterminer une valeur de $N_{optimal}$ à partir d'une équation simple, donc le calcul ne prendrait alors que quelques secondes.

Les relations entre $N_{optimal}$ et l'effectif total, pour une prévalence nationale donnée, et entre $N_{optimal}$ et la prévalence nationale, pour un effectif total donné, ne sont pas linéaires. Les courbes qui s'ajustent le mieux aux données sont les fonctions puissances. Il est donc possible de modéliser la relation entre $N_{optimal}$ et les paramètres des enquêtes selon la fonction :

Équation 2

$$N_{optimal} = b_0 \cdot effectif\ \ total^{b_1} \cdot prévalence\ nationale^{b_2} \cdot nombre\ de\ grappes^{b_3}$$

Une régression non linéaire sous SPSS a permis de déterminer les valeurs suivantes pour les différents coefficients : 4,812 pour b_0, 0,523 pour b_1, -0,434 pour b_2 et 0,041 pour b_3. Pour vérifier la robustesse de la modélisation effectuée, nous avons comparé, pour des échantillonnages correspondant à la prévalence du VIH des 15-49 ans dans cinq EDS récentes, la valeur de $N_{optimal}$ calculée à partir de cent simulations sur le pays modèle et celle obtenue à partir du modèle puissance (tableau 3).

Tableau 3. $N_{optimal}$ pour cinq enquêtes récentes, calculé sur 100 simulations et selon l'équation puissance

Enquête	Personnes testées (15-49 ans)	Nombre de grappes	Prévalence nationale du VIH	$N_{optimal}$ calculé avec l'équation 2	$N_{optimal}$ calculé sur 100 simulations
Burkina Faso 2003	7.244[a]	400	1,8 %	498	506
Cameroun 2004	9.900	466	5,5 %	363	352
Kenya 2003	6.001	400	6,7 %	255	246
Tanzanie 2003-2004	10.747	350	7,0 %	337	338
Ghana 2003	9.144	412	2,2 %	516	501

[a] Ce nombre est de 7.151 dans le tableau 1, conformément aux données du rapport final de l'enquête. Nos propres calculs à partir des fichiers de données aboutissent à 7.244 personnes testées, avec un résultat déterminé, âgées de 15 à 49 ans, et avec un poids non nul. Nous avons donc retenu cette valeur pour le présent tableau..

[13] C'est-à-dire proche de 50 %, puisque lorsque la prévalence nationale p est supérieure à 50 %, la situation est équivalente à estimer la prévalence complémentaire 1-p.

Les écarts entre les valeurs de $N_{optimal}$ calculées sur cent simulations et celles calculées à partir de l'équation 2 sont faibles (15 ou moins). Or, une précision à plus ou moins 25 est suffisante dans la mesure où cela ne modifie que très peu la carte produite.

Application à la prévalence du VIH du Burkina Faso et du Cameroun

Choix des paramètres

Nous avons appliqué cette approche aux données de prévalence du VIH (15-49 ans), fournies par les dernières EDS du Burkina Faso (2003) et du Cameroun (2004). La valeur seuil utilisée pour recoder les grappes urbaines a été 10 kilomètres. Les résultats restent néanmoins semblables pour des valeurs seuils de 15 ou 20 kilomètres. Le tableau 4 présente les données ayant présidées la sélection des agglomérations urbaines pour le paramètre U.

Nous avons sélectionné d'une part, les villes où le nombre de personnes testées est relativement important et réparties sur plusieurs grappes. Dans le cas présent, le choix a été de retenir les agglomérations où le nombre de personnes testées étaient supérieures à 150 au Burkina Faso et à 100 au Cameroun, réparties sur au moins 6 grappes. D'autre part, des agglomérations où, même si le nombre de personnes testées est relativement faible, la prévalence observée dans l'EDS est proche de celle observée en clinique prénatale. C'est le cas de Bamenda, Bertoua, Kumba, Edéa et Ebolowa au Cameroun.

Enfin, des agglomérations où la prévalence observée dans l'EDS, bien qu'inférieure à celle mesurée en clinique prénatale, reste supérieure à celle estimée pour une valeur nulle de U. C'est le cas de Ouahigouya, Garoua et Maroua[14].

Pour le paramètre N, nous avons utilisé 500 pour le Burkina Faso et 350 pour le Cameroun, conformément au tableau 3. La valeur de R a été déterminée en prenant le neuvième décile du rayon des cercles de lissage lorsque seul N est appliqué. R vaut ainsi 117 kilomètres pour les EDS du Burkina Faso et du Cameroun (la valeur inférieure de N au Cameroun est compensée par la superficie plus grande du pays).

Résultats

Les résultats préliminaires de l'application de cette méthodologie aux données du Burkina Faso et du Cameroun ont fait l'objet d'une présentation et d'une discussion avec les autorités nationales de ces deux pays, en mars 2006 à Yaoundé pour le Cameroun et en juin 2006 à Ouagadougou pour le Burkina Faso.

Au Burkina Faso (figure 9), les foyers d'infections se situent tout d'abord au niveau des agglomérations urbaines situées le long des axes routiers internationaux. Ouagadougou, la capitale, est concernée mais également Koudougou, Bobo-Dioulasso et Banfora sur l'axe reliant Ouagadougou à la Côte d'Ivoire, ou encore Ouahigouya sur un axe routier vers le Mali.

La région autour de Débougou et de Gaoua est particulièrement touchée. Outre sa proximité avec la Côte d'Ivoire et le Ghana, elle est connue comme une zone importante d'orpaillage, impliquant à la fois une présence d'hommes migrants venus seuls et une prostitution marchande non négligeable.

[14] Voir la discussion précédente sur le choix des agglomérations urbaines.

Le nord et l'ouest du pays, zones sahéliennes, présentent une prévalence faible. Étant peu habitées, et de ce fait peu enquêtées, seule une tendance régionale y est mise en évidence. Les variations locales, s'il y en a, ne transparaissent pas sur la carte. Par ailleurs, les pays frontaliers de ces régions ont des prévalences estimées du VIH faibles : 1,8 % pour le Mali, 1,1 % pour le Niger et 2,0 % pour le Bénin fin 2003 selon Onusida [15] tandis que celle de la Côte d'Ivoire est estimée à 7,0 %.

Tableau 4. Choix des agglomérations urbaines pour le paramètre U

Agglomération urbaine	Population†	EDS‡				Surveillance sentinelle§			Inclue pour U
		Prévalence du VIH (%)	IC 95 %	N	Nombre de grappes	Prévalence du VIH (%)	IC 95 %	N	
Cameroun									
Douala	1.494.700	4,4	3,2-6,0	931	43	8,0	5,6-11,2	400	Oui
Yaoundé	1.248.200	8,5	6,8-10,6	870	45	7,2	5,1-10,0	471	Oui
Garoua	356.900	5,3	1,9-12,6	93	6	8,0	5,6-11,2	402	Oui
Bamenda	316.100	10,6	5,0-20,4	75	4	10,1	7,0-14,4	286	Oui
Maroua	271.700	6,2	2,3-14,5	80	3	7,3	4,7-11,0	300	Oui
Bafoussam	242.000	9,6	5,5-15,9	146	6	5,9	3,1-10,6	186	Oui
Ngaoundéré	189.800	11,5	7,4-17,2	183	8	11,4	8,5-15,0	395	Oui
Bertoua	173.000	9,4	4,8-16,9	107	4	9,0*	5,3-14,7	166*	Oui
Loum	141.400	8,4	3,9-16,6	91	4	n. d.	n. d.	n. a.	Non
Kumba	125.600	7,2	3,2-14,9	96	5	9,8	3,7-22,2	51	Oui
Edéa	122.300	9,5	4,2-19,2	74	5	9,0*	4,5-16,8	100*	Oui
Kumbo	116.500	9,3	2,5-25,8	33	2	n. d.	n. d.	n. a.	Non
Foumban	113.100	5,1	1,3-15,1	59	3	7,3	3,0-15,8	82	Non
Nkongsamba	110.600	6,1	2,8-12,0	131	9	n. d.	n. d.	n. a.	Oui
Mbouda	101.100	5,3	0,3-28,1	19	1	n. d.	n. d.	n. a.	Non
Dschang	87.000	0,0	0,0-12,3	35	2	4,3*	1,9-8,9	164*	Non
Limbé	84.500	8,9	3,3-20,3	56	3	5,6	3,0-10,0	197	Non
Ebolowa	79.500	11,0	5,2-21,1	72	3	11,6	7,7-17,1	198	Oui
Burkina Faso									
Ouagadougou	709.736	3,8	2,1-6,5	346	18	4,0	2,3-7,0	321	Oui
Bobo-Dioulasso	309.711	3,3	1,8-5,7	402	19	4,3	2,9-6,3	630	Oui
Koudougou	72.490	4,3	2,0-8,5	190	7	n. d.	n. d.	n. a.	Oui
Ouahigouya	52.193	2,5	0,4-10,1	73	4	3,6	2,1-5,9	422	Oui
Banfora	49.724	4,5	2,0-9,2	167	10	n. d.	n. d.	n. a.	Oui
Pouytenga	35,720	0,0	0,0-16,6	25	2	n. d.	n. d.	n. a.	Non
Kaya	33.958	0,4	0,0-9,5	50	3	3,6	2,1-5,8	473	Non
Dédougou	33.815	0,0	0,0-19,2	21	1	1,2	0,5-2,7	506	Non
Tenkodogo	31.466	2,9	0,1-18,0	31	2	2,6	1,4-4,7	430	Non
Fada N'Gourma	29.254	0,0	0,0-10,9	40	3	1,3	0,5-3,0	455	Non
Dori	23.768	0,0	0,0-17,2	24	2	1,6	0,6-3,5	384	Non
Gaoua	16.424	0,0	0,0-10,7	41	2	3,0	1,7-5,2	437	Non

n. d. : non disponible – IC : intervalle de confiance

Test exact de Fisher de comparaison de deux proportions : pour l'ensemble des agglomérations, aucune différence significative à 20 % n'est observée en prévalence observée dans l'EDS et celle provenant de la surveillance sentinelle des femmes enceintes, à l'exception de Douala où la différence est significative (p-value = 0,01203). † Cameroun : population en 2001, résultats préliminaires, Cameroon Statistical Yearbook 2004 [16] ; Burkina Faso : population en 1996, Recensement de 1996 cité par City Population[17].

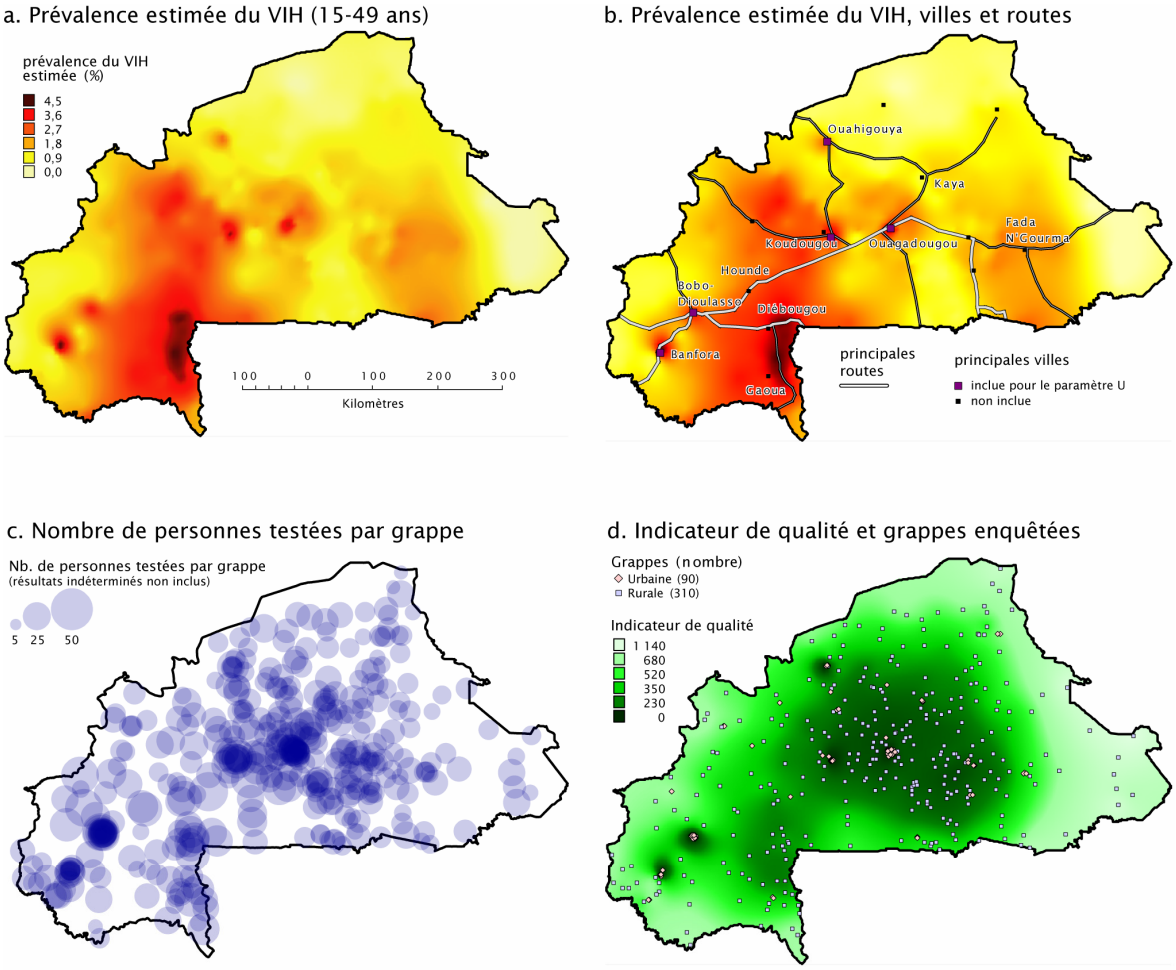
‡ Cameroun: EDS 2004; Burkina Faso : EDS 2003. Une grappe urbaine est considérée comme appartenant à une agglomération si elle est située dans un rayon de dix kilomètres. La pondération des individus est prise en compte pour les calculs (prévalence, intervalle de confiance, test de comparaison).

§ Cameroun : prévalence du VIH1 en 2002 [18] ; Burkina Faso : prévalence du VIH en 2003 [19].

* Résultats non disponibles en 2002; les données présentées datent de 2000[20].

Note : le taux d'urbanisation était de 52,8 % en 2003 au Cameroun[16] et de 15,5 % en 1996 au Burkina Faso[21].

Figure 9. Prévalence du VIH estimée au Burkina Faso et cartes complémentaires



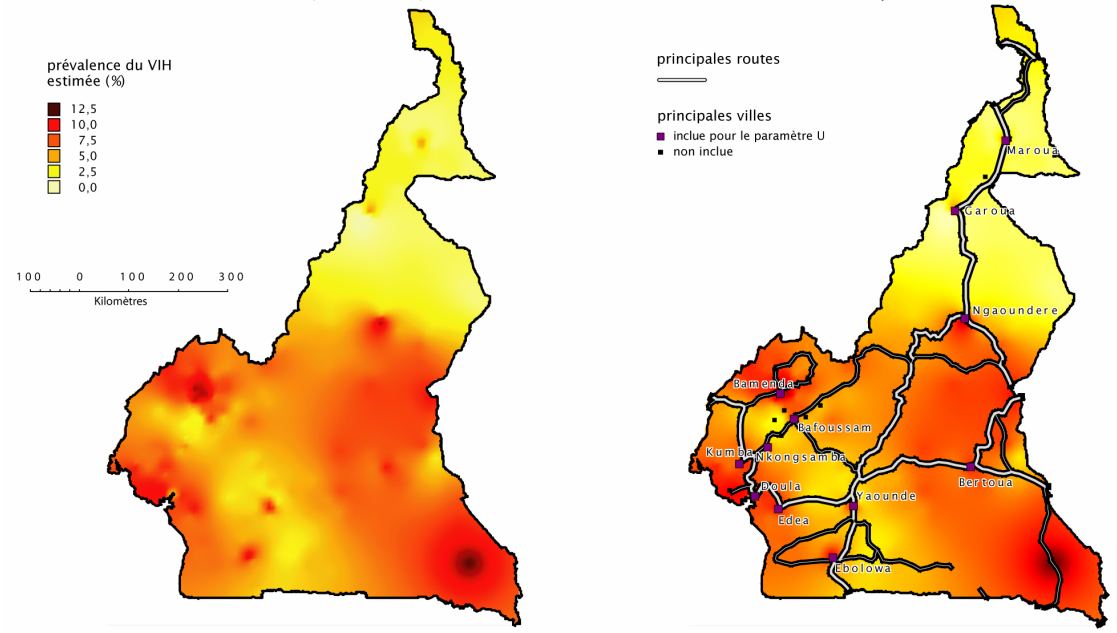
Note 1 : La prévalence nationale du VIH (15-49 ans) est de 1,8% dans l'EDS 2003 du Burkina Faso. 7.244 personnes ont été testées (résultats indéterminés exclus), réparties en 400 grappes.
Note 2 : Les paramètres utilisés pour l'estimation de la prévalence du VIH sont : N=500, R=117 km et U=5 (voir carte b pour les villes retenues).
Note 3 : L'indicateur de qualité est calculé pour chaque grappe selon l'expression $r^2/\sqrt{n}$ où r est le rayon du cercle de lissage et n le nombre de personnes testées dans ce cercle.
Sources : EDS 2003 du Burkina Faso pour les données VIH, DCW pour les frontières nationales, GRUMP pour les principales villes et ArcAtlas (ESRI) pour les principales routes.

Enfin, cette carte présente des similitudes avec celle des zones migratoires sur la période 1998-1992 [22] et celle des retours de rapatriés burkinabés depuis la Côte d'Ivoire fin 2002 et début 2003 [23]. Si ces similitudes doivent être interprétées avec prudence, ne constituant en rien une preuve de cause à effet, elles plaident néanmoins en faveur de la vraisemblance de la carte produite dans la mesure où plusieurs enquêtes ont montré que les migrations pouvaient être associées avec une prévalence du VIH plus élevée [24-27].

Figure 10. Prévalence du VIH estimée au Cameroun et cartes complémentaires

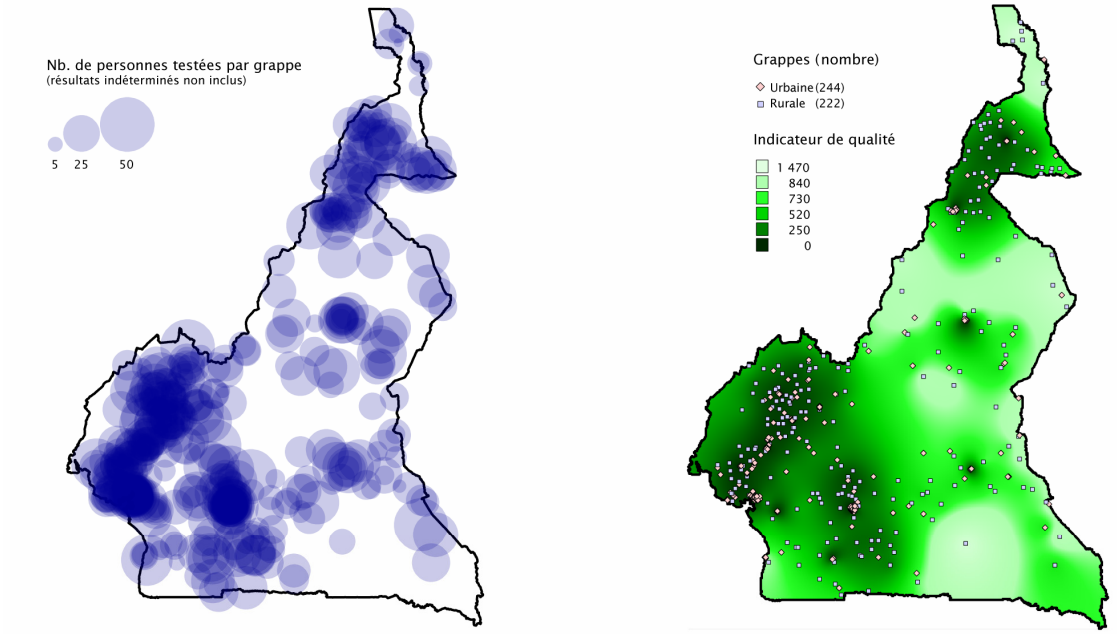
a. Prévalence estimée du VIH (15-49 ans)

b. Prévalence estiée du VIH, villes et routes



c. Nombre de personnes testées par grappe

d. Indicateur de qualité et grappes enquêtées



Note 1 : La prévalence nationale du VIH (15-49 ans) est de 5,5% dans l'EDS 2004 du Cameroun. 9.900 personnes ont été testées (résultats indéterminés exclus), réparties en 466 grappes.
Note 2 : Les paramètres utilisés pour l'estimation de la prévalence du VIH sont : N=350, R=117 km et U=12 (voir carte b pour les villes retenues).
Note 3 : L'indicateur de qualité est calculé pour chaque grappe selon l'expression $r^2/\sqrt{n}$ où r est le rayon du cercle de lissage et n le nombre de personnes testées dans ce cercle.
Sources : EDS 2004 du Cameroun pour les données VIH, DCW pour les frontières nationales, GRUMP pour les principales villes et ArcAtlas (ESRI) pour les principales routes.

Au Cameroun (figure 10), l'épidémie se concentre dans le sud-ouest du pays, particulièrement dense et urbanisé, Douala, Yaoundé et Bamenda/Bafoussam constituant un triangle d'importants mouvements migratoires internes[28].

En dehors de cette zone, on retrouve des foyers épidémiques dans les villes situées le long des axes routiers : Maroua, Garoua, Bertoua, Ebolawa et surtout N'Gaoundéré qui présente les

niveaux de prévalence les plus élevés du pays. Une rupture nette est observée entre le nord et le sud de N'Gaoundéré, le nord sahélien présentant une épidémie plus faible. Cependant, la ligne de démarcation ne peut être dessinée précisément du fait du faible nombre de grappes dans cette zone et de la valeur élevée de l'indicateur de qualité (figure 10.d).

Le Nigéria et le Tchad qui bordent le nord du pays présentent une prévalence du VIH relativement faible (3,9 % et 3,5 % fin 2005 selon Onusida [15]) par rapport aux pays limitrophes du sud-est (10,7 % en Afrique Centrale, 5,5 % au Congo et 7,9 % au Gabon). Cette zone frontalière semble présenter des prévalences relativement élevées. Cependant, bien que ce résultat soit plausible, il doit être interprété avec prudence vu le faible nombre de personnes testées et des valeurs élevées de l'indicateur de qualité dans cette zone (figures 10.c et 10.d).

Conclusion

Les EDS sont conçues pour être représentatives de la population nationale et non des territoires. De ce fait, la répartition spatiale des personnes enquêtées n'est pas aléatoire mais reflète les densités de population. La méthodologie employée ici repose sur deux éléments. D'une part une estimation de la prévalence de chaque grappe à partir de cercles de même effectif, afin de compenser les aléas du tirage au second degré (sélection des ménages éligibles de chaque grappe). D'autre part, une interpolation spatiale pour estimer la surface de variation spatiale de la prévalence et compenser ainsi le tirage au premier degré (sélection des grappes).

Nous avons fait le choix d'une méthodologie reposant sur des cercles de même effectif[15] plutôt que des cercles de même rayon. Cette seconde alternative aurait été appropriée dans le cas d'enquêtes échantillonnées sur les territoires (sélection des zones d'enquêtes avec une probabilité égale à leur superficie). Cependant, pour les EDS qui sont échantillonnées avec une probabilité de sondage des zones d'enquête égale à leur population[16], nous aurions alors estimée la prévalence sur un nombre de personnes très élevés dans les zones fortement peuplée et procéder ainsi à une perte d'informations. Quant aux zones peu denses, les effectifs auraient été faibles et les estimations auraient reflétés essentiellement la variance de l'échantillonnage.

Le choix d'une méthodologie reposant sur des cercles de même effectif permet donc de rendre une interpolation spatiale possible, mais également de procéder à un lissage différentiel suivant que le territoire soit plus ou moins peuplé et de mettre en évidence un maximum de variations spatiales de la prévalence à partir des données parcellaires des EDS. Cependant, cette méthode se heurte aux limites de ce type d'enquête, d'où l'impossibilité de mettre à jour des variations très localisées. La simulation d'EDS sur des données réelles de recensement constitue une voie pour mieux comprendre les limites de cette méthode et nécessite d'être explorée.

Les cartes obtenues sur les données de prévalence du VIH au Burkina Faso et au Cameroun sont plausibles au regard des informations que nous avons sur ces épidémies. Cette approche méthodologique n'est pas spécifique au VIH et peut être appliquée plus généralement à tout type d'indicateur correspondant à une proportion, du moment que l'enquête utilisée ait un échantillonnage comparable à celui d'une EDS. Grace à la libraire logicielle prevR conçue, elle peut être facilement mise en œuvre sur un jeu de données.

[15] L'ajout des paramètres R et U permet d'affiner les estimations produites. Cependant, le paramètre fondamental reste le paramètre N.
[16] Plus précisément le nombre de ménages au dernier recensement de la population.

Références

1. Tanzania Commission for AIDS (TACAIDS), National Bureau of Statistics (NBS), National AIDS Control Programme (NACP), ORC Macro. **Tanzania Atlas of HIv/AIDS Indicators 2003-2004**. Dar es Salaam: TACAIDS, NBS, NACP, ORC Macro; 2006: 37 pages (http://www.measuredhs.com/pubs/pdf/GS5/GS5.pdf).

2. Measure DHS. **Measure DHS Brochure**; 2006 (http://www.measuredhs.com/pdfs/DHS%20brochure%20with%20branding-updated.pdf).

3. Measure DHS. **Methodology - Collecting Geographic Data**; 2006 (http://www.measuredhs.com/topics/gis/methodology.cfm).

4. Center for International Earth Science Information Network (CIESIN) of Columbia University, International Food Policy Research Institute (IFPRI), The World Bank, Centro Internacional de Agricultura Tropical (CIAT). **Global Rural-Urban Mapping Project (GRUMP), Alpha Version: Urban Extents**. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC) of Columbia University; 2004 (http://sedac.ciesin.columbia.edu/gpw).

5. Center for International Earth Science Information Network (CIESIN) of Columbia University, International Food Policy Research Institute (IFPRI), The World Bank, Centro Internacional de Agricultura Tropical (CIAT). **Global Rural-Urban Mapping Project (GRUMP), Alpha Version: Population Density Grids**. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC) of Columbia University; 2004 (http://sedac.ciesin.columbia.edu/gpw).

6. Chorley RJ, Haggett P. **Trend-Surface Mapping in Geographical Research**. *Transactions of the Institute of British Geographers* 1965:47-67.

7. Griffin WR. **Residual Gravity in Theory and Practice**. *Geophysics* 1949,14:39-56.

8. Krumbein WC. **Regional and local components in facies maps**. *AAPG Bulletin* 1956,40:2163-2194.

9. Nettleton LL. **Regionals, Residuals and Structures**. *Geophysics* 1954,19:1-22.

10. Center for International Earth Science Information Network (CIESIN) of Columbia University, International Food Policy Research Institute (IFPRI), The World Bank, Centro Internacional de Agricultura Tropical (CIAT). **Global Rural-Urban Mapping Project (GRUMP), Alpha Version: Settlement Points**. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC) of Columbia University; 2004 (http://sedac.ciesin.columbia.edu/gpw).

11. Krige D. **A statiscal approcah to some basic mine valuation problemens on the witwatersrand**. *Journal of the Chemical, Metallurgical and Mining Society* 1951,52:119-139.

12. Baillargeon S. **Le krigeage : revue de la théorie et application à l'interpolation spatiale de données de précipitations** [Maître ès sciences (M.Sc.)]. laval: Université de Laval; 2005:137.

13. R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing; 2006 (http://www.R-project.org).

14. Pebesma EJ. **Multivariable geostatistics in S: the gstat package**. *Computers & Geosciences* 2004,30:683-691.

15. UNAIDS. **Report on the global AIDS epidemic**. Geneva; 2006: 629 pages (http://www.unaids.org/en/HIV_data/2006GlobalReport/default.asp).

16. National Institute of Statistics. **Cameroon Statistical Yearbook 2004**. Yaoundé: NIS; 2004: 635 pages (http://www.statistics-cameroon.org/ins/annuaire.htm).

17. Population C. **Burkina Faso - City Population - Cities, Towns & Provinces - Statistics and Maps**; 2006 (http://www.citypopulation.de/BurkinaFaso.html).

18. National AIDS Control Committee. **National HIV Sentinel surveillance Report (2002)**. Yaoundé: Ministry of Public Health; 2003: 42 pages.

19. CNSL Burkina Faso. **Récapitulatif des taux de prévalence du VIH chez les femmes enceintes dans les sites sentinnelles de 1997 à 2003**: CNLS; 2003.

20. National AIDS Control Committee. **Technical Report National Serosurvey on VIH/Syphilis**. Yaoundé: Ministry of Public Health; 2001: 37 pages.

21. Institut National de la Statistique et de la Démographie. **Annuaire Statistique du Burkina faso 1999**. Ouagadougou: Ministère de l'Économie et des Finances; 2000: 219 pages.

22. Réseau Migrations et Urbanisation en Afrique de l'Ouest (REMUAO). **Enquête sur les Migrations et l'Urbanisation au Burkina Faso (EMUBF) 1992-1993 Rapport national descriptif**. Bamako: CERPOD; 1997: 140 pages.

23. SP/CONASUR, UNICEF, PAM. **Analyse des données sur les rapatriés de Côte d'Ivoire**. Ouagadougou: Comité National de Secours d'Urgence et de Réhabilitation; 2004: 65 pages.

24. Lurie MN, Williams BG, Zuma K, Mkaya-Mwamburi D, Garnett GP, Sweat MD, *et al.* **Who infects whom? HIV-1 concordance and discordance among migrant and non-migrant couples in South Africa**. *AIDS* 2003,17:2245-2252.

25. Decosas J. **Special report: West Africa. Migration factor makes regional approach essential**. *AIDS Anal Afr* 1995,5:8-9.

26. Quinn TC. **Population migration and the spread of types 1 and 2 human immunodeficiency viruses**. *Proceedings of the National Academy of Sciences of the United States of America* 1994,91:2407-2414.

27. Lalou R, Piche V. **Migration et sida en Afrique de l'Ouest: un état des connaissances**. *Les Dossiers du Ceped* 1994,28:1-48.

28. Gubry P, Negadi G, Tayo J. **[The population of Cameroon according to the 1976 census] La Population du Cameroun au recensement de 1976**. *Revue science et technique. Série Sciences humaines / Science and technology review. Social sciences series* 1983:7-38.