

**7ème Conférence Africaine sur la Population
JOHANNESBURG – AFRIQUE DU SUD
30 NOV. – 4 DEC. 2015**

***DIVIDENDE DEMOGRAPHIQUE EN AFRIQUE : PERSPECTIVES,
OPPORTUNITES ET DEFIS***

**15. Questions méthodologiques et sources de données.
1501 Recensement et systèmes d'état civil**

**Evaluation des données de recensement au Sénégal en utilisant les
observatoires de population**

Cheikh Tidiane NDIAYE¹, Bruno MASQUELIER^{2,3}, Ndéye Binta DIEME¹, Gilles PISON³, Ibrahima DIOUF¹, Lucie LECOMTE⁴, Valérie DELAUNAY⁴, Samba NDIAYE¹, Ousmane NDIAYE⁴, Paul SENGHOR⁴, Pape Niokhor DIOUF⁴

¹Agence Nationale de la Statistique et de la Démographie (ANSD)

²Université Catholique de Louvain (UCL)

³Institut National d'Études Démographiques (INED)

⁴Institut de recherche pour le développement (IRD)

RESUME

L'évaluation de la qualité des données des recensements du Sénégal de 2002 et 2013 est réalisée en prenant comme référence les trois sites de suivi démographiques du pays (Niakhar, Mlomp, Bandafassi). Nos hypothèses sont que les données des observatoires de population sont plus précises que celles des RGPHs, et que la collecte numérique a amélioré la qualité des données. Les indicateurs de fécondité et ceux relatifs à mortalité des enfants issus des recensements sont relativement proches de ceux des observatoires. L'effet de l'usage du PDA n'est pas net puisque les niveaux de mortalité déduits du recensement de 2013 sont plus proches de ceux des observatoires, mais les effectifs de population et les niveaux de fécondité cadrent mieux avec le recensement de 2002. L'appariement au niveau individuel et les analyses complémentaires qui vont suivre permettront de se faire une meilleure idée sur la qualité des données en 2013.

I. Contexte de l'étude

L'état civil, en tant que système d'observation exhaustif et continu des phénomènes démographiques, devrait idéalement constituer une ressource de données régulières. Malheureusement, dans la plupart des pays africains, ce système d'observation des événements démographiques souffre de plusieurs insuffisances. Au Sénégal, malgré une volonté politique prononcée en faveur de son développement, le système d'enregistrement des faits d'état civil demeure encore très limité, aussi bien en termes de couverture que de qualité des données collectées. Le recensement de la population de 2013 indique que seuls 35% des décès survenus au cours des douze mois ayant précédé le recensement ont fait l'objet d'une déclaration auprès d'un officier de l'état civil (au niveau national). Dans le domaine de la nuptialité, seuls 28% des mariages ont été déclarés. Bien que plus fréquente, la déclaration des naissances demeure également faible par rapport à l'objectif d'une couverture complète (80%).

Dans ces circonstances, la réalisation de recensements généraux reste indispensable pour la connaissance de la population et la réalisation de projections démographiques. Cependant, étant donné la lourdeur de telles opérations, il est inévitable que les données issues des recensements soient entachées d'erreurs. Il peut s'agir tant d'erreurs de couverture (sous-dénombrement, doubles comptes, etc.) que d'erreurs de contenu (omissions d'événements, erreurs sur période de référence des 12 derniers mois, erreurs sur les âges ou les parités atteintes) (Ewbank, 1981). Les indicateurs démographiques déduits des recensements reflètent également pour partie Les définitions adoptées pour faciliter une énumération exhaustive de la population (telles que les règles utilisées pour définir la population résidente ou pour délimiter les contours des ménages).

L'objectif visé à travers cette étude consiste à évaluer la qualité des données collectées lors des deux derniers recensements du Sénégal, menés en 2002 et en 2013, en prenant comme référence les données de trois sites de suivi démographiques (HDSS) au Sénégal : Bandafassi, Niakhar et Mlomp. On fait ainsi l'hypothèse que les données HDSS sont plus précises que celles issues des RGPHs. On cherche également à évaluer si la collecte numérique organisée en 2013 a réduit les erreurs de déclaration et rapproché le RGPH des données des sites de suivi.

Cette communication présente les premiers résultats de comparaisons menées au niveau agrégé, et procure une première évaluation du niveau de couverture des recensements de même que de la qualité des données sur l'âge, la mortalité et la fécondité. Elle détaille également la procédure utilisée pour procéder aux appariements individuels. La communication complète présentée lors de la conférence APC fournira des résultats relatifs aux appariements, pour lesquels le travail de terrain est toujours en cours.

II. Sources de données

II.1. Sites de suivi démographique (SSD)

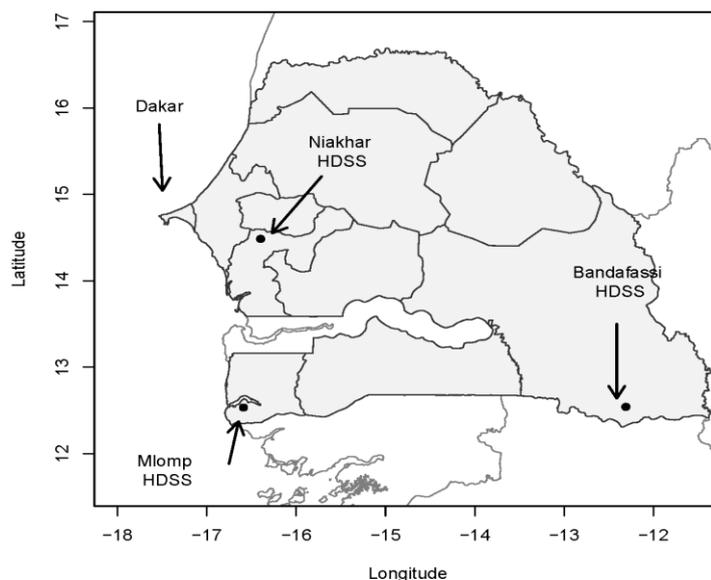
Les sites de suivi concernés par cette étude sont les observatoires de population de Niakhar, de Mlomp et de Bandafassi. L'observatoire de population constitue une approche relativement complexe et originale en matière de collecte de données démographiques. Ces observatoires permettent en effet d'observer de façon *continue et longue une petite région ou une société bien circonscrite, dont les objectifs sont à la fois la mesure et la compréhension des changements socio-démographiques et sanitaires* (Gilles Pison, 2006). Ces observatoires servent également de plateformes pour mener des projets de recherche ou d'interventions en santé.

De par la particularité de son approche, l'observatoire utilise une méthodologie de collecte spécifique pour saisir la dynamique des populations. On y effectue un recensement initial et des enquêtes à passages répétés, organisées à intervalles réguliers (par exemple, tous les 6 mois ou tous les ans) pour saisir l'évolution de la structure de la population et des événements démographiques tels que les naissances, les décès et les mouvements migratoires. La déclaration d'un décès est généralement suivie d'une autopsie verbale, afin d'établir la cause de décès.

Le suivi démographique de Niakhar est en place depuis 1962 pour les 8 villages de la communauté rurale de Ngayokheme et depuis 1983 dans les 30 villages (Delaunay et al., 2013). Situé au centre du Sénégal dans la région de Fatick, la population sous observation est estimée à 45 000 habitants en 2013. D'un point de vue culturel, Niakhar est essentiellement peuplé de Sérér avec une minorité de Wolof et de Laobé. Au cours du temps, la périodicité des passages a évolué (annuelle, hebdomadaire, trimestrielle, puis semestrielle) et les outils utilisés ont également changé (relevé sur fiches imprimées, Tablet PC).

La zone de Mlomp est un ensemble de 11 villages ou quartiers situés au sud-ouest du Sénégal, dans la région de Ziguinchor, en Casamance (Pison et al., 2002). Elle se situe dans le département d'Oussouye, à une cinquantaine de kilomètres de la frontière avec la Guinée-Bissau. Elle correspond à environ la moitié de l'arrondissement de Loudia-Ouolof et comptait environ 8700 personnes en 2013. Les villages sont regroupés dans un cercle de 15 km de diamètre et sont entourés de terres inondées durant la saison des pluies sur lesquelles est cultivé le riz. La plupart des habitants appartiennent à l'ethnie Joola. Les migrations saisonnières sont très fréquentes à Mlomp, près de 40% des habitants s'absentant une partie de l'année (situation au début des années 2000). La proportion de migrants saisonniers est maximale chez les jeunes adultes, atteignant près de 80% chez les femmes de 15-24 ans.

Graphique 1 : Localisation des observatoires de population au Sénégal



La zone d'étude de Bandafassi est située au sud-est du Sénégal dans la région de Kédougou (Pison et al. 2014). Elle comprend une population de 13 000 habitants en 2013, répartie en 42 villages de petite taille - 280 habitants en moyenne – et divisés en hameaux pour certains. La densité de population est de près de 20 habitants au km². La population est divisée en trois groupes ethniques vivant pour la plupart dans des villages distincts : il s'agit essentiellement de Bediks (25%), de Malinkés (17%) et de Peulhs (58%).

II.2. Recensements nationaux

Le recensement général de la population de 2002, le troisième du genre au Sénégal, avait comme mandat de renouveler le stock des données exhaustives existantes dans les domaines social, économique et démographique et de fournir des informations liées à des préoccupations nouvelles. Le questionnaire utilisé reprend l'essentiel des modules permettant de faire une estimation directe et indirecte de la mortalité et de la fécondité et recommandés par les Nations Unies (2008) : nombres d'enfants nés vivants et survivants, nombre de naissances des 12 derniers mois, décès des douze derniers mois, etc. La collecte des données a été organisée du 8 au 22 décembre 2002.

Le Sénégal a réalisé son 4^e recensement général de la population et de l'habitat en 2013. Le questionnaire utilisé en 2013 reprend la plupart des questions de 2002 et y ajoute la survie des parents. Ce recensement a été réalisé à l'aide des nouvelles technologies avec des PDA (Pocket Digital Assistant) comme outils de collecte des données. L'introduction de cette innovation technologique majeure visait principalement à raccourcir les délais de publications des résultats et à améliorer la qualité des données à travers des contrôles de cohérence intégrés et un suivi en temps réel de l'évolution de la collecte. Les résultats définitifs du recensement ont été publiés en septembre 2014, soit moins d'un an après la collecte des données (entre le 19 novembre et le 9 décembre 2013).

III. Méthodologie de l'évaluation

III.1. Harmonisation des codes des localités

Pour procéder à une comparaison avec les trois observatoires sénégalais, il est nécessaire d'identifier avec précision les districts de recensements couvrant les trois zones rurales. La délimitation des localités (villages/hameaux) a donc été harmonisée entre les deux bases de données. La mise en cohérence de l'identification des localités s'appuie sur le système de codification des localités (SYSCOL) du Sénégal. Compte tenu des différentes codifications des entités administratives en vigueur dans le pays, le projet SYSCOL a été développé pour disposer d'une codification harmonisée et unique, prenant en compte les découpages administratifs passés, en vigueur et à venir. Les localités RGPH et HDSS ont été affectées de ce code en vue de faciliter leur rapprochement. Ce premier effort d'harmonisation, basé sur les noms des localités, a été ensuite complété par des appariements au niveau des ménages et des individus (cf. infra), ce qui permettra à terme de circonscrire exactement dans les questionnaires du recensement les populations faisant l'objet du suivi démographique.

III.2. Inclusion et exclusion des populations observées

Dans le cadre d'un observatoire, il est important de définir de manière précise les critères d'inclusion et d'exclusion dans la population cible. Pour les trois observatoires retenus dans cette analyse, les individus peuvent entrer en observation à partir d'une naissance ou d'une immigration au sein de la zone d'étude. Pour les sorties, il s'agit des individus décédés ou ayant émigré. Les trois sites de suivi identifiés dans cette étude présentent une diversité en termes d'emplacement géographique et de caractéristiques socio-culturelles. La collecte des données est globalement similaire dans ces sites¹.

Entre recensements et observatoires, des différences dans la définition des concepts clés peuvent avoir une incidence importante sur les résultats obtenus. En particulier, selon la source de données utilisée, la notion de ménage ou de résidence prend des significations différentes. Le tableau 1 ci-dessous synthétise ces différences. Globalement, le recensement utilise une définition plus restrictive de la résidence, car son objectif est d'atteindre l'exhaustivité du dénombrement et d'éviter les doubles comptes. Les observatoires de population adoptent une définition bien plus large, car ils visent à saisir la dynamique démographique des zones suivies, en ce compris la contribution des migrants régulièrement de passage.

¹ Les visites sont organisées sur une base annuelle à Bandafassi et Mlomp et tous les 4 mois à Niakhar. Une autre différence réside dans le fait qu'à Bandafassi, l'information est recueillie auprès d'informateurs locaux plutôt qu'auprès de chaque chef de ménage.

Tableau 1 : Comparaisons des définitions des personnes résidentes entre le recensement et les sites de suivi démographique

Rubrique	Niakhar	Bandafassi	Mlomp	RGPHAE
Unité de résidence	Cuisine (localisée dans une concession)	Carré (ou concession) uniquement	Ménage (localisé dans une concession ou hank)	Ménage (localisé dans une concession)
Sont considérées comme personnes résidentes	<ul style="list-style-type: none"> Personnes résidant habituellement dans la cuisine Absents partis depuis moins de six mois sans intention de rester hors de la zone plus de six mois (y compris enfants confiés) Y compris certaines exceptions : les travailleurs saisonniers <i>Norane</i> qui passent au moins un mois dans la zone pendant l'hivernage, les saisonniers <i>Navétane</i>, qui passent au moins un mois pendant la saison sèche, les élèves et talibés dont les parents résident dans la zone Possibilité de distinguer les résidents absents 	<ul style="list-style-type: none"> Personnes résidant habituellement dans le carré et personnes absentes depuis moins de quatre ans Quelques exceptions : par exemple les personnes migrant suite à la formation ou la rupture d'une union ne sont plus enregistrés comme résident après leur date de sortie Peu de possibilités de distinguer les résidents absents. 	<ul style="list-style-type: none"> Personnes vivant habituellement Personnes absentes depuis moins de deux ans Possibilité de distinguer les résidents permanents et les migrants saisonniers. 	<ul style="list-style-type: none"> Personnes vivant habituellement avec le ménage pendant au moins 6 mois Personnes installées depuis moins de 6 mois (durée de présence dans le ménage inférieure à la norme retenue) mais avec l'intention d'y habiter. Personnes « visiteurs » de passage depuis 6 mois ou plus

III.3. Procédure de confrontation des données au niveau agrégé

Pour réaliser cette étude on procède d'abord à une comparaison, au niveau agrégé, des principaux indicateurs démographiques des deux sources. Les comparaisons portent sur les effectifs totaux des villages, les structures par âge et sexe, les tendances de la mortalité des enfants, les taux de fécondité par âge, les taux de mortalité par âge/sexe au cours des douze derniers mois ainsi que les décès pour cause maternelle. Les spécificités de chaque type de collecte sont mises en évidence afin de préciser leur impact sur la mesure des phénomènes démographiques. Cela permet d'apprécier la plausibilité des tendances et niveaux de la fécondité et de la mortalité déduits des RGPHs en comparaison à ceux dégagés des HDSS.

III.4. Appariements au niveau individuel

Une procédure d'appariement des données au niveau individuel a été établie en vue de faciliter l'identification des populations, mais aussi d'enrichir mutuellement les sources de données en vue des analyses approfondies. D'une part, les données du RGPH permettent d'affecter aux données HDSS des informations portant sur l'appartenance religieuse, la situation d'handicap, la fréquentation scolaire, le niveau d'instruction, les activités exercées, ou encore l'enregistrement des faits d'état civil, etc. D'autre part, les recherches portent

également sur l'évaluation des incohérences au niveau individuel. On tente en particulier de mettre en exergue la fréquence des omissions de ménages ou d'individus RGPHs et HDSS, les erreurs de déclaration d'âges, les incohérences en matière de survie du père ou de la mère, les omissions de décès survenus au cours des 12 derniers mois et les incohérences en matière d'âge au décès, date de décès et causes de décès (pour les décès maternels). Travailler au niveau individuel permet par ailleurs d'analyser les déterminants des différentes incohérences entre RGPHs et HDSS, en termes de niveaux d'instruction, de liens de parenté avec le chef du ménage, d'histoire migratoire, etc.

En raison des questions de protection de données personnelles, les bases de données disponibles pour les analyses dans les instituts nationaux de statistiques sont généralement dépourvues de noms et prénoms. Notre étude, qui s'apparente à une enquête post censitaire, a donc nécessité un travail de saisie de ces informations pour les données du RGPH de 2002 (les noms et prénoms étant disponibles pour les sites d'observation et le RGPHAE de 2013). A l'issue de ce traitement, les variables communes ont aussi été identifiées avant la production des fichiers comparables entre HDSS et RGPHs. Ces fichiers sont ensuite anonymisés.

Les appariements individuels sont réalisés en trois étapes.

- 1) Dans un premier temps, un appariement automatique est réalisé sur base d'un programme informatique. Chaque ménage énuméré au recensement (ou chaque concession dans le cas de Bandafassi) est comparé à tous les ménages identifiés dans l'observatoire de population concerné. Cette comparaison porte sur les noms et prénoms des membres du ménage. Pratiquement, chaque nom et prénom déclaré au recensement est confronté aux noms et prénoms dans la base, et la proximité entre ces noms est mesurée à l'aide de la distance de Jaro-Winkler. Cette distance tient compte de la longueur des noms, du nombre de caractères correspondants entre deux noms, ainsi que du nombre de caractères à transposer pour passer de l'un à l'autre. Sur base de cette distance est calculé un indice de similarité ; plus il est élevé, plus les noms sont proches. Par exemple, pour les deux orthographes d'un même prénom, BIRAM et BIRAME, l'indice vaut 0.94. Par contre, entre les prénoms ABLAYE et ABDOULAYE, l'indice vaut 0.89. Pour nos appariements, deux noms sont considérés comme identiques si cet indice est supérieur à 0.9. Ceci permet d'associer des noms en présence de variations orthographiques ainsi qu'erreurs de saisie. Dès que le seuil de 0.9 est dépassé pour plus de la moitié des membres du ménage, nous considérons provisoirement qu'il s'agit du même ménage. Chaque membre des observatoires dispose d'un numéro d'identifiant unique ; nous enregistrons alors ce numéro pour les personnes énumérées dans le recensement pour lesquels les noms et prénoms correspondent.
- 2) Dans un deuxième temps, nous reconstituons les liens de parenté entre les différents individus identifiés au recensement à partir des liens avec le chef de ménage. L'ordre d'énumération des individus permet aussi de reconstituer, même imparfaitement certains liens de parenté ; notamment, les enfants sont généralement placés à la suite de leur mère dans le questionnaire ménage. Sur cette base, nous dessinons pour chaque ménage un *diagramme de parenté*, localisant le chef de ménage, son épouse, ses parents, etc. Les individus non apparentés sont représentés sur le côté. Le même exercice est réalisé pour les données de l'observatoire, et les numéros d'identifiants unique sont visibles dans ces diagrammes. Les ménages déjà appariés par la procédure automatique décrite plus haut sont présentés deux à deux. Chaque diagramme de parenté est ensuite examiné visuellement par un membre de l'équipe.

L'objectif de ce travail (assez fastidieux) consiste à vérifier les appariements automatiques, et à chercher, pour les ménages pour lesquels aucune correspondance n'a pu être trouvée, un lien dans les données des observatoires.

- 3) Dans un troisième temps, les résultats des recherches menées sur les diagrammes imprimés sont saisis et permettent de reproduire de nouveaux diagrammes, plus complets et corrigés, pour mener des vérifications sur le terrain. Des membres de l'équipe de recherche se rendent directement dans les ménages pour poser des questions sur des individus qui n'ont pas pu être appariés. Une information relative aux objectifs de cette recherche est donnée et un consentement oral est recueilli à cette occasion.

A aucun moment durant cette procédure d'appariement ne sont utilisées directement les variables qui font l'objet d'analyses sur la qualité des données : la déclaration des âges, le nombre d'enfants nés vivants, etc. Les noms et prénoms, la structure des ménages et les relations de parenté sont les principales informations exploitées.

IV. Résultats au niveau agrégé

IV.1. Population par âge et sexe, concessions et ménages

L'analyse des résultats au niveau agrégé révèle d'importantes différences en termes d'effectifs de populations, du nombre de concessions et de ménages dénombrés par les observatoires et les deux recensements. Ces écarts sont présentés dans le Tableau 2 sous forme de différences relatives².

Les effectifs de population sont systématiquement plus élevés selon les données des sites de suivi démographiques. Ces différences d'effectifs varient par sexe ; à Bandafassi et Niakhar, elles sont plus larges chez les hommes, tandis qu'à Mlomp elles sont plus larges chez les femmes. Ces différences d'effectifs sont par ailleurs plus importantes en 2013 qu'en 2002, pour chaque sexe et dans tous les sites. A Mlomp comme à Niakhar, les données collectées dans les observatoires sur les absences et les migrations saisonnières ont été exploitées pour approcher au plus près la définition de la résidence qui prévaut dans les recensements. Il est toutefois possible que certains travailleurs saisonniers échappent à l'enregistrement des migrations de courte durée et restent donc comptabilisés comme résidents au moment du recensement. Un sous-enregistrement de certaines catégories de la population dans les recensements n'est toutefois pas à exclure, et cela pourra être évalué sur base des appariements individuels.

Des différences non négligeables sont également observées en comparant les nombres de concessions et de ménages recensés. Ces différences sont attendues, car les ménages et les concessions sont des réalités sociales multidimensionnelles qu'il est difficile d'appréhender statistiquement, surtout dans les zones rurales africaines caractérisées par une fréquence élevée de la polygamie et une forte mobilité entre unités d'habitation. Le recensement de 2002 définit les ménages comme « étant un groupe de personnes, apparentées ou non, qui

² En notant x les valeurs correspondantes aux observatoires de populations et y les valeurs correspondantes aux recensements, ces différences relatives sont calculées comme $100 * (x - y) / ((x + y) / 2)$. Des valeurs positives traduisent donc un effectif plus élevé selon les observatoires de population.

vivent ensemble sous le même toit et mettent en commun tout ou partie de leurs ressources pour subvenir à leurs besoins essentiels, notamment le logement et la nourriture. Ces personnes appelées membres du ménage, prennent généralement leurs repas en commun et reconnaissent l'autorité d'une seule et même personne, le chef de ménage ». Plusieurs dimensions sont donc à prendre en compte : le fait de mettre en commun des ressources, de disposer d'un budget commun, de dormir sous le même toit, de partager des repas, de reconnaître l'autorité d'une même personne de référence ? Ces différentes dimensions ne sont pas aisément objectivables, et il n'existe pas nécessairement d'unité sociale clairement définie qui puisse toutes les rassembler (Gruenais 1981, Hertrich 1996, Van de Walle 2006). Ces premières comparaisons entre observatoires et recensements (ainsi que l'appariement des données individuelles qui est en cours) permettent d'évaluer la sensibilité des estimations démographiques aux critères établis pour délimiter ce qu'est un ménage et qui le compose.

À Bandafassi, l'observatoire de population ne distingue pas les ménages mais regroupe uniquement les individus dans les concessions, ou carrés. Dans cette zone, le recensement comptabilise environ 50% de concessions de plus que l'observatoire, alors que les effectifs de population sont assez proches. On peut donc s'attendre à observer des concessions de plus petites tailles, où sont davantage représentées des familles nucléaires, dans les données du recensement, par rapport aux données des sites de suivi démographique. Peu de concessions enregistrées dans les deux recensements contiennent en fait plus d'un ménage. La façon de regrouper les résidents en concessions dans les observatoires peut également expliquer une partie de ces écarts ; il n'est pas toujours aisé d'évaluer si un déménagement de certains membres de la concession doit donner lieu à la création d'une nouvelle concession ou si ces personnes continuent à être rattachées à leur concession d'origine. Les comparaisons effectuées au niveau du site d'observation de Mlomp montrent également des écarts importants au niveau du nombre de concession dénombrés entre les deux sources. En effet, le nombre de concessions dénombrés au recensement est trois fois plus élevé que le nombre de concessions issu du HDSS, tandis que les effectifs de ménages cadrent davantage entre les deux sources. La tendance à conserver des individus dans une même concession (si les membres continuent de reconnaître l'autorité d'un même chef) est plus fréquente dans l'observatoire de Mlomp, même en cas de migrations dans d'autres parties de la zone, tandis que dans les recensements nationaux, les agents recenseurs vont davantage se référer à des ensembles d'habitations groupées. A Niakhar, par contre, l'observatoire comptabilise légèrement plus de concessions que les recensements, et également plus de ménages (appelées cuisines) en 2002.

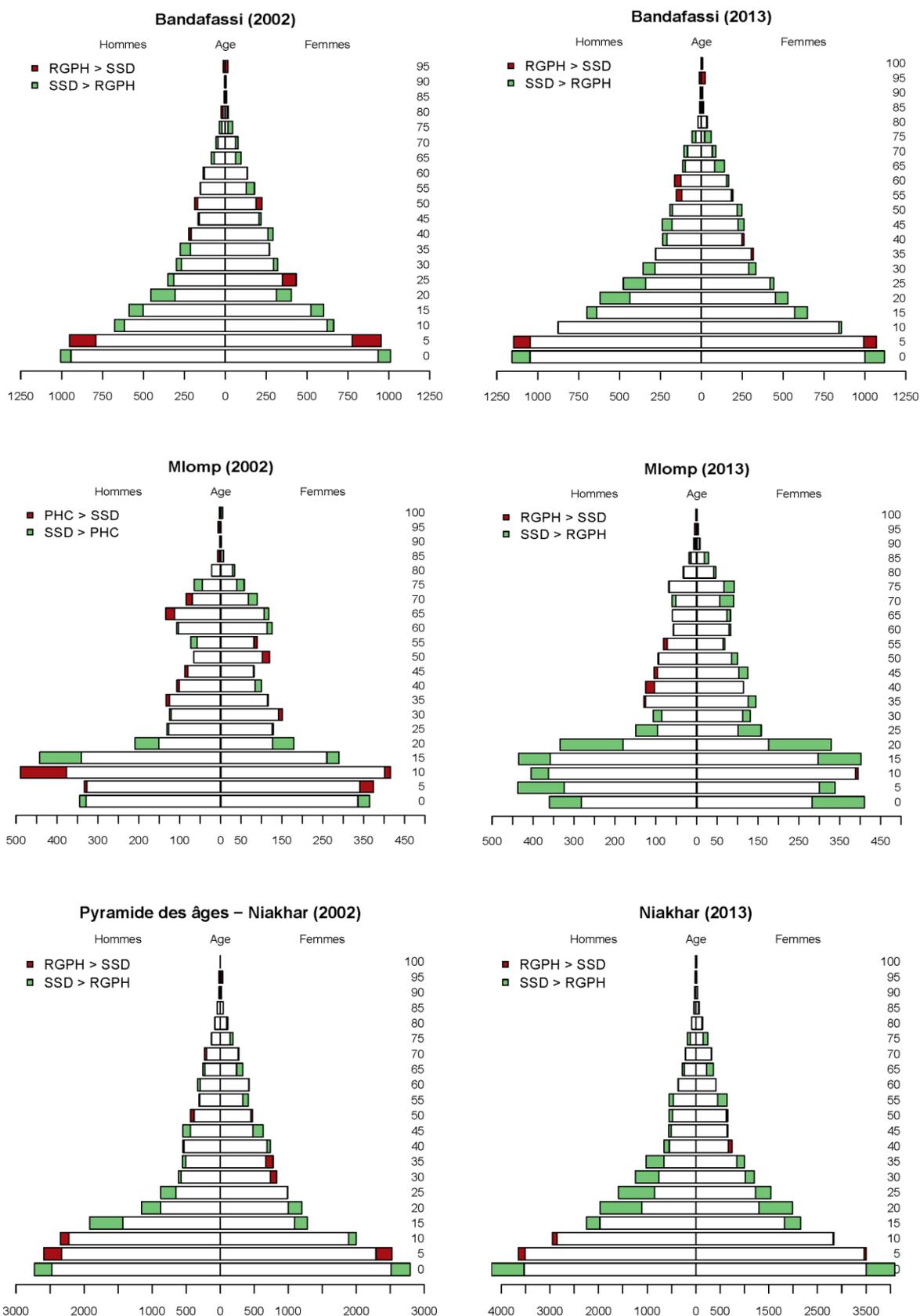
Tableau 2 : Comparaisons des effectifs par sexe, du nombre de concessions et du nombre de ménages selon les sources dans les trois observatoires de population (différences relatives en %)

		Effectifs		Unités d'habitation	
		Masculins	Féminins	Concessions	Ménages
Bandafassi	2002	6,3	3,4	-36,2	
	2013	8,6	7,3	-45,5	
Mlomp	2002	1,8	5	-102,9	10,5
	2013	19,1	23,6	-116,3	25,1
Niakhar	2002	7,1	4,6	12,6	28,6
	2013	18,3	12,5	2,3	-4,7

Ces écarts tiennent aux différences de définition du « ménage », de la « concession », et de la « résidence » d'une source à l'autre, ainsi qu'à leur application : même à supposer que la définition soit strictement la même, elle peut ne pas donner le même résultat selon le mode de collecte – transversal, comme dans le cas du recensement, ou longitudinal, comme dans le cas de l'observatoire. Par ailleurs, les écarts d'un site à l'autre varient en relation avec la diversité des contextes et des modes de collecte dans les trois observatoires de population sénégalais. Des analyses complémentaires sont nécessaires pour mieux comprendre ces écarts, car ils ont des implications importantes sur la mesure de la dynamique démographique, ainsi que sur l'analyse de la pauvreté ou des relations entre genres et générations.

Le graphique 2 présente les pyramides des âges observées dans les trois observatoires et selon les deux recensements. Les effectifs représentés en vert correspondent aux cas où les observatoires comptent davantage de résidents que le recensement (après avoir éliminé, autant que possible, les migrants saisonniers). Les barres horizontales représentées en rouge renvoient à des effectifs plus larges dans les recensements. Dans les trois sites, les structures par âge sont assez comparables entre recensements et observatoires. L'excédent de population dans les observatoires (par rapport au recensement) est plutôt concentré aux jeunes âges adultes, ce qui laisse penser qu'il s'agit de migrants comptabilisés dans l'observatoire comme résidents. Mais dans les trois sites, les recensements ont également enregistré moins de résidents entre 60 et 80 ans, ce qui pourrait plutôt traduire des omissions de personnes âgées. A l'inverse, il faut noter une sous-estimation de la population âgée de 5 à 10 ans, et parfois de 10 à 15 ans, au niveau des trois sites de suivi démographique en 2002, ainsi qu'en 2013 à Bandafassi.

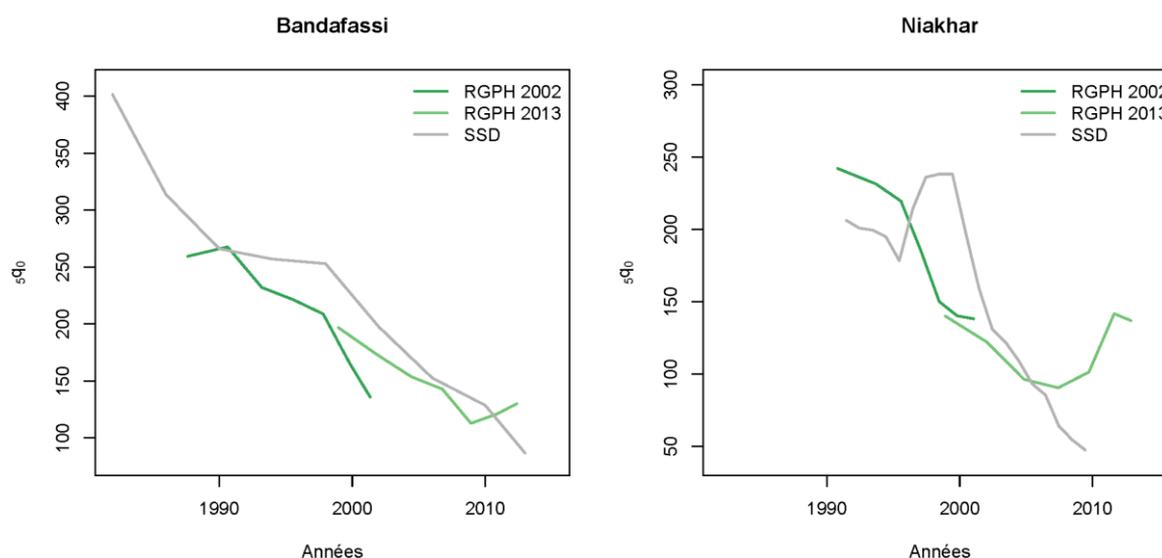
Graphique 2 : Structures par âge de la population des trois zones d'études selon la source de données en 2002 et 2013



IV.2. Analyse des principaux indicateurs démographiques

Les questions portant sur la survie des enfants permettent une estimation indirecte des niveaux et tendances de la mortalité infanto-juvénile (Hill, 2013). La méthode a été appliquée aux données des deux recensements. Seuls les résultats pour Bandafassi et Niakhar sont présentés dans le Graphique 3, car les estimations pour Mlomp sont très erratiques, compte tenu de la plus petite taille de cet observatoire et des niveaux de mortalité et de fécondité plus faibles. À Bandafassi, la tendance obtenue de façon indirecte est assez proche de celle réellement observés dans l'observatoire, bien que le niveau de $5q_0$ soit environ 10% trop bas. À Niakhar, les niveaux obtenus de façon indirecte cadrent plus ou moins avec la mortalité sous-jacente jusqu'au milieu des années 2000. Jusqu'en 1995, l'estimation du recensement est 20% plus élevée que celle de l'observatoire, après, elle est nettement plus basse. Par ailleurs, le pic de mortalité de la fin des années 1990 n'est pas représenté, car les estimations indirectes reposent sur une hypothèse de linéarité des tendances, et en conséquence, elles tendent à lisser les irrégularités. Ce pic de mortalité est imputable à une épidémie de méningite, ainsi qu'à une reprise des fièvres, maladies diarrhéiques et infections respiratoires aiguës (Etard et al., 2004). Les estimations pour les années les plus récentes, basées sur les femmes de moins de 30 ans en 2013, indiquent par contre que la mortalité a connu une hausse, tandis que les données de l'observatoire témoignent d'une baisse rapide et continue de la mortalité des enfants. Des analyses complémentaires sont nécessaires pour identifier la source de cet écart ; il pourrait s'agir d'erreurs de déclarations dans le recensement mais aussi d'un effet de sélection des répondantes les plus jeunes, puisque nous avons déjà relevé que le recensement comptabilisait moins de femmes à ces âges que celles enregistrées comme résidentes dans l'observatoire. Quoiqu'il en soit, les estimations indirectes de la mortalité des enfants nous semblent peu robustes.

Graphique 3 : Tendances de la mortalité des enfants à Bandafassi et Niakhar selon la source de données : comparaison des estimations indirectes du recensement avec les niveaux obtenus directement des données longitudinales



La comparaison des courbes des taux de fécondité obtenus à partir des deux sources donne des résultats assez proches à Bandafassi, surtout en 2002 (Graphique 4). A Mlomp, il est important de pouvoir isoler les migrantes saisonnières dans le calcul des taux de fécondité,

sans quoi la comparaison avec le recensement est biaisée. En effet, les femmes qui ne migrent pas de façon saisonnière (qui sont celles enregistrées comme résidentes au recensement) sont nettement plus fécondes que les autres. À partir de ces femmes uniquement, l'observatoire de Mlomp procure des taux de fécondité très proches de ceux obtenus à partir des deux recensements. Cette sélectivité de la migration explique certainement la plus forte fécondité observée au recensement à Niakhar en 2013, où il semble que certaines migrantes n'aient pas pu être repérées dans la base de données de l'observatoire et aient donc été exclues du calcul. Par contre, en 2002, le recensement procure des taux de fécondité plus faibles, et on est donc amené à penser à des omissions de naissances ou à des problèmes relatifs à la période de référence des douze mois.

L'introduction de questions sur les décès survenus dans les ménages au cours des 12 derniers mois dans les deux recensements permet de calculer des tables de mortalité. Les niveaux de mortalité infantile (${}_1q_0$), juvénile (${}_4q_1$), aux âges adultes (${}_{45}q_{15}$) et aux âges élevés (${}_{20}q_{60}$) sont présentés pour les deux sexes confondus dans le tableau pour les trois sites, ainsi que les espérances de vie à la naissance (e_0) et à 5 ans (e_5).

Les estimations issues des recensements sont comparées aux niveaux de mortalité obtenus à l'aide des données longitudinales des suivis démographiques. Les tables de mortalité ne peuvent être calculées pour Mlomp sur une seule année (en 2013), le nombre de décès annuel étant faible - la population est de plus petite taille et les niveaux de mortalité sont relativement plus faibles que dans les autres sites. En 2002 par contre, un calcul est possible, car la mortalité de cette année a été particulièrement élevée. Des dizaines de résidents de l'observatoire sont décédés le 26 Septembre, soit deux mois avant le recensement, dans le naufrage du Joola, un ferry reliant Ziguinchor à Dakar, qui a chaviré au large de la côte de la Gambie, coûtant la vie à 1863 personnes.

L'analyse des tables de mortalité obtenues à l'aide des deux sources de données met en évidence des schémas similaires dans les trois sites. Les informations collectées dans le recensement ont donné lieu à une sous-estimation sensible de la mortalité infantile dans le recensement de 2002. C'est moins le cas en 2013 à Bandafassi et à Niakhar. Par contre, la mortalité juvénile est partout assez bien estimée, sans doute en raison d'erreurs sur les âges aux décès qui prennent la forme d'une attirance sur 1 an, et qui conduisent à sous-estimer la mortalité infantile tout en rehaussant la mortalité juvénile. Entre 15 et 60 ans, par contre, les données du recensement tendent à procurer des niveaux de mortalité nettement plus proches que ceux enregistrés dans le suivi, voire même légèrement plus hauts en 2013. Entre 60 et 80 ans, compte-tenu des omissions plus fréquentes des personnes âgées (et de leurs décès) ainsi que des erreurs sur les âges, la concordance entre les estimations du recensement et celles du suivi est également remarquable. Dans l'ensemble, ces résultats confirment que les données portant sur les décès survenus dans les ménages ne sont pas adaptées à l'analyse de la mortalité des enfants, mais de façon moins attendue, ils indiquent que les taux obtenus de façon directe, sans aucun ajustement au-delà de 5 ans, procurent des niveaux de mortalité tout à fait plausibles. Les espérances de vie à 5 ans sont très proches, avec des écarts de moins de 2 ans dans tous les cas, sauf à Mlomp en 2002.

Graphique 4 : Taux comparés de fécondité par âge selon la source de données à Niakhar, Bandafassi et Mlomp en 2002.

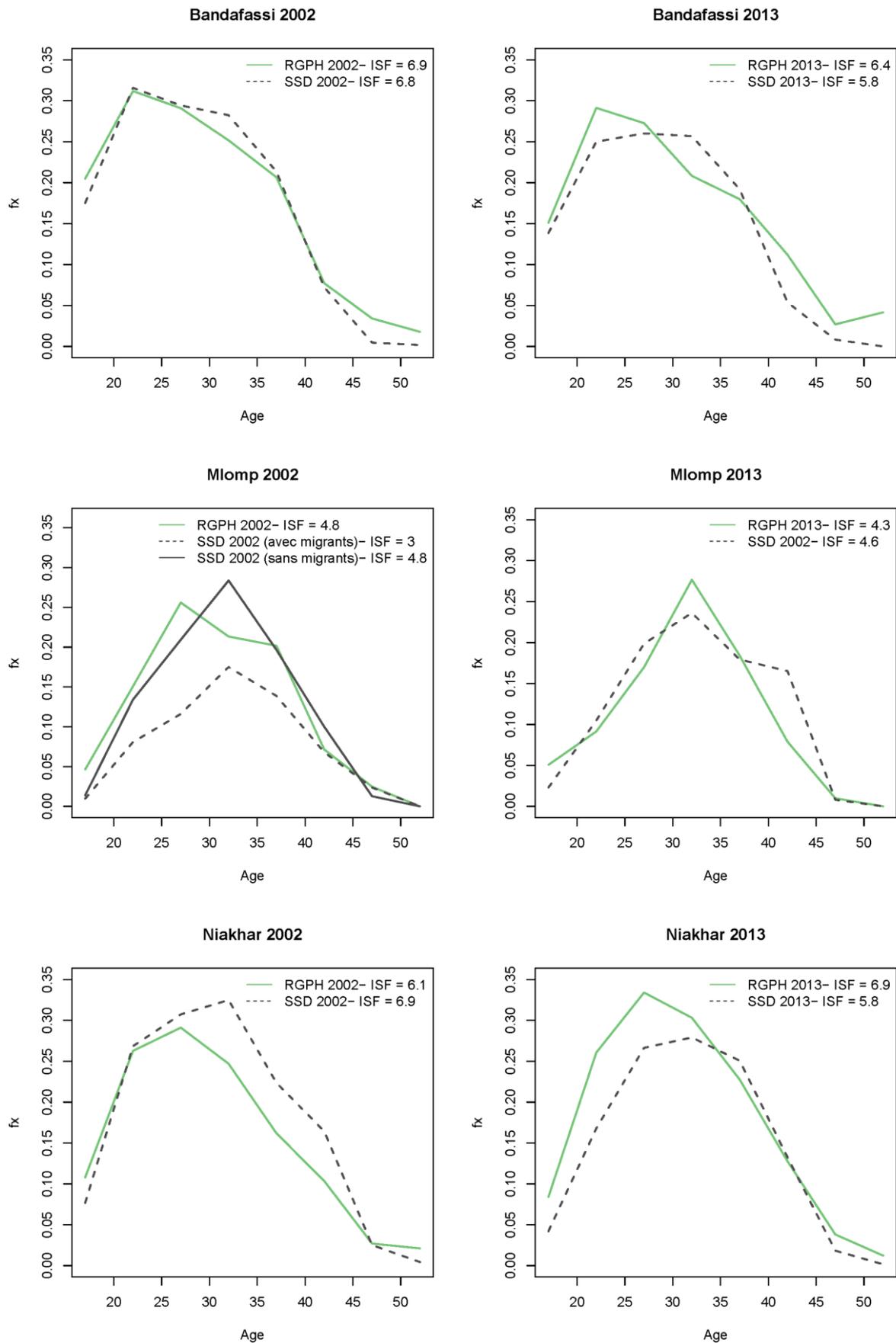


Tableau 3 : Indices de mortalité par âge obtenus à partir des déclarations sur les décès des 12 derniers mois dans les recensements et selon les données longitudinales des observatoires de population

		Bandafassi		Mlomp		Niakhar	
		DSS	RGPH	DSS	RGPH	DSS	RGPH
2002	${}_1q_0$	0.11	0.04	0.05	>0.01	0.06	0.02
	${}_4q_1$	0.09	0.11	0.04	0.05	0.08	0.08
	${}_{45}q_{15}$	0.25	0.21	0.54	0.59	0.26	0.24
	${}_{20}q_{60}$	0.68	0.71	0.64	0.43	0.52	0.49
	e_0	54.3	58.4	46.6	46.3	59.6	62.8
	e_5	61.9	62.9	45.7	43.5	64	64.5
2013	${}_1q_0$	0.06	0.08			0.02	0.02
	${}_4q_1$	0.04	0.04			0.02	0.03
	${}_{45}q_{15}$	0.29	0.38			0.13	0.19
	${}_{20}q_{60}$	0.59	0.48			0.52	0.52
	e_0	61.1	58.8			70.7	68.7
	e_5	62.4	61.3			68.7	66.6

V. Conclusion

L'intérêt de ce projet réside dans le fait qu'il adopte une approche innovante pour évaluer la qualité des données de recensement. En effet, même si des approches basées sur des appariements sont souvent utilisées dans le cadre des enquêtes post-censitaires, les observatoires de population fournissent souvent des données de meilleure qualité, du fait de leur caractère longitudinal. Cependant, les résultats de cette étude doivent être relativisés compte tenu des différences de définition de la résidence et des unités d'habitations d'une source à une autre. Des problèmes de qualité des données restent possibles également dans les observatoires, tels que des omissions d'événements, de concessions entières, des problèmes de double compte des migrants, etc. Par ailleurs, la définition de la population suivie est parfois problématique, surtout dans les cas où la collecte des migrations saisonnières semble incomplète, comme à Niakhar et à Bandafassi. Les sites de suivi peuvent donc également bénéficier d'un appariement avec les données des recensements. Indépendamment des enjeux relatifs à la qualité des données, les informations de type socio-économique restent assez sommaires dans les sites de suivi, et le couplage des informations du recensement permettra dans le futur de mieux éclairer certains comportements démographiques.

Ces premières comparaisons au niveau agrégé permettent de relativiser certaines idées assez fréquentes et relatives à la qualité des données des recensements africains. Elles montrent notamment que les taux de fécondité par âge calculés à partir des naissances récentes ne sont pas nécessairement sous-estimés, et que la prise en compte des migrantes saisonnières peut avoir un impact bien plus large que d'éventuelles omissions de naissances. Elles montrent également qu'au-delà des premiers âges de l'enfance, les décès déclarés dans les 12 mois

précédant le recensement peuvent procurer des niveaux de mortalité peu biaisés. Timæus (1991) estime que bien souvent, environ un tiers à la moitié des décès seulement sont déclarés en réponse aux questions sur les décès récents des ménages. Or dans les trois zones d'études, la très grande majorité des décès semblent avoir été déclarés au recensement. Par contre, les estimations indirectes de mortalité des moins de cinq ans sous-estiment la mortalité ou, dans le cas de Niakhar en 2013, procurent des tendances peu vraisemblables.

Nos résultats sont assez différents de ceux obtenus dans une analyse comparable menée sur le recensement sénégalais de 1988 par Pison et al. (1995). Ces auteurs ont comparé les niveaux de fécondité et de mortalité des trois observatoires avec ceux issus du recensement pour les zones rurales des départements dans lesquels sont localisés les observatoires. La comparaison était donc moins directe que ce n'est le cas ici. Les taux de fécondité pour Mlomp se sont d'ailleurs révélés très différents de ceux du département correspondant (Oussouye), tandis que les écarts étaient moins flagrants dans le cas de Niakhar et de Bandafassi, et laissaient penser à une sous-déclaration des naissances dans le recensement. Ce n'est pas le cas ici, sauf peut-être à Niakhar en 2002. En matière de mortalité, les décès de moins d'un an semblaient également sensiblement sous-déclarés en 1988, tandis que la mortalité entre 1 et 5 ans avait été considérée comme fiable. Il conviendrait aujourd'hui de reprendre ces analyses pour 1988 en limitant la comparaison aux villages couverts par les trois sites de suivi.

A ce stade, nos analyses ne permettent pas de conclure à une nette amélioration de la qualité des recensements entre 2002 et 2013, amélioration qui pourrait être imputée à l'usage de PDAs. Les niveaux de mortalité déduits du recensement de 2013 semblent plus proches des niveaux des observatoires, mais les effectifs de population et les niveaux de fécondité cadrent davantage dans le cas du recensement de 2002. L'appariement au niveau individuel et les analyses complémentaires qui vont suivre permettront de se faire une meilleure idée sur la qualité des données en 2013.

VI. Bibliographie

- Delaunay V., L. Douillot, A. Diallo, D. Dione, J-F. Trape, O. Medianikov, D. Raoult, and C. Sokhna, 2013, "Profile: The Niakhar health and demographic surveillance system", *International Journal of Epidemiology*, 42(4):1002_1011
- Etard J. F.; Le Hesran, J. Y.; Diallo, A.; Diallo, J. P.; Ndiaye, J. L. & Delaunay V., 2004, "Childhood mortality and probable causes of death using verbal autopsy in Niakhar, Senegal, 1989-2000", *Int J Epidemiol*, 33, 1286-1292
- Ewbank D., 1981, *Age misreporting and age-selective under-enumeration: Sources, patterns and consequences for demographic analysis*, Committee on Population and Demography, Report No. 4
- Gruenais M., 1981, 'Famille et démographie de la famille en Afrique', Document de travail 1, ORSTOM.
- Hertrich V., 1996, *Permanences et changements de l'Afrique rurale: dynamiques familiales chez les Bwa du Mali*, Paris: CEPED.
- Hill K., 2013, "Indirect estimation of child mortality", in Moultrie T. et al., *Tools for Demographic Estimation*, UNFPA/IUSSP

- Moultrie, TA; Dorrington, RE; Hill, AG; Hill, K; Timaeus, IM; Zaba, B; (2013) Tools for Demographic Estimation. International Union for the Scientific Study of Population, Paris.
- Pison G., 2006, "Les observatoires de population : un outil pour l'étude des changements démographiques et sanitaires dans les pays du sud", in Graziella et al., Démographie et analyse, VIII Observation, méthodes auxiliaires, enseignement et recherche, éditions de l'INED, pp. 125-150.
- Pison G., A. Gbadinho, A. Wade, and C. Enel, 2002, "Mlomp demographic surveillance system, Senegal". In IDRC, *Population and health in developing countries*, pages 271-278. International Development Research Centre
- Pison G., L. Douillot, M. Kante, O. Ndiaye, P. Diouf, P. Senghor, Sokhna C., and Delaunay V, 2014, "Health & Demographic Surveillance System Profile: Bandafassi Health and Demographic Surveillance System (Bandafassi HDSS), Senegal", *International Journal of Epidemiology*, 43(3):739_748
- Pison, G.; Hill, K.; Cohen, B. & Foote, K. 1995, *Population Dynamics of Senegal*, National Academy Press, Washington, 254 p.
- Timaeus, I., 1991, "Measurement of Adult Mortality in Less Developed Countries: A Comparative Review", *Population Index*, 57, pp. 552-568
- United Nations (2008), *Principles and Recommendations for Population and Housing Censuses*, UN, Dept. of International Economic and Social Affairs, United Nations, New York
- Van de Walle E. 2006, *African Households: Censuses and Surveys*, General demography of Africa, M.E. Sharpe, Incorporated.