
Préparation des Enquêtes annuelles de recensement à Mayotte : étude du plan de sondage et redéfinition des zones de collecte

Martin CHEVALIER(), Laurent COSTA(**)*

() Drees, Sous direction Observation de la solidarité, France,
martin.chevalier@sante.gouv.fr*

*(**) Insee, Direction des statistiques démographiques et sociales, France,
laurent.costa@insee.fr*

Mots-clés : Enquête, Sondage, Échantillonnage, Simulations, Recensement, Collecte.

Domaines : Théorie des sondages amont (échantillonnages particuliers).

Résumé

La première enquête annuelle de recensement (EAR) a eu lieu à Mayotte en 2021. La méthodologie habituellement employée pour le recensement de la population en « grande » commune (commune de 10 000 habitants ou plus) dans les DOM repose sur l'interrogation d'environ 8 % des logements de la commune chaque année, sélectionnés aléatoirement par sondage au sein de zones de collecte appelées « îlots ». Préalablement à ce tirage, une enquête cartographique est réalisée entre cinq et dix mois avant l'EAR.

L'importance de l'habitat de fortune à Mayotte – qui représente 39 % des logements recensés au RP 2017 – est susceptible d'affecter sensiblement les performances de cette méthodologie. En effet, cet habitat évolue très rapidement (constructions, destructions, etc.) et pose des problèmes de réparabilité sur le terrain. Sur la base de simulations réalisées à partir du RP 2017, ce document propose une évaluation de l'impact de l'habitat de fortune sur les estimateurs du recensement ainsi que des performances de scénarios de tirage alternatifs.

Par ailleurs, les îlots mobilisés dans le cadre du RP 2017 sont très hétérogènes en taille, ce qui pose des difficultés aussi bien méthodologiques que de terrain. En étroite collaboration avec le Service régional de l'Insee à Mayotte, ces îlots ont donc été redéfinis de sorte que leurs frontières soient durablement réparables sur le terrain et qu'ils soient raisonnablement homogènes en taille. Ce document expose dans un second temps le processus mis en œuvre pour ce faire.

Abstract

The first annual census survey was held in Mayotte in 2021. The methodology usually used for the census of the population in a « large » municipality (municipality of 10,000 inhabitants or more) in the French overseas departments is based on the questioning of approximately 8 % of the dwellings in the municipality each year, randomly selected by survey within areas called « islets ». Prior to this draw, a cartographic survey is carried out between five and ten months before the census survey.

The importance of makeshift housing in Mayotte – which represents 39 % of the dwellings in the exhaustive census of 2017 – is likely to significantly affect the performance of this methodology. Indeed, this type of dwellings evolves very quickly (construction, destruction, etc.) and poses problems to find the real ones selected. Based on simulations carried out from the census of 2017, this document provides an assessment of the impact of makeshift housing on census estimators as well as the performance of alternative drawing scenarios.

In addition, the islets mobilized for the census of 2017 are very large and heterogeneous in size, which implies methodological and field difficulties. In close collaboration with the Insee regional service in Mayotte, these islets have been redefined so that their borders are durably identifiable on the ground and that they are reasonably homogeneous in size. This document then describes the process implemented to do this.

Table des matières

Introduction	1
1 Les EAR dans les grandes communes des DOM et leurs applications au contexte mahorais	2
1.1 L'échantillonnage des EAR dans les grandes communes des DOM	2
1.2 Le cas particulier de Mayotte	4
1.3 Le problème du biais de couverture de la base de sondage à Mayotte	6
1.4 Le déroulement des travaux pour préparer le lancement des EAR à Mayotte . . .	9
2 Les scénarios de tirage alternatifs	11
2.1 Scénario 1 : Méthodologie DOM classique	11
2.2 Scénario 2 : Tirage d'îlots à recenser exhaustivement	12
2.3 Scénario 3 : Adaptation du traitement des habitations de fortune sur celui des habitations mobiles et des sans-abris (HMSA)	13
2.4 Scénario 4 : Tirage en îlots uniquement pour les habitations de fortune	15
2.5 Bilan des scénarios envisagés	17
3 Le redécoupage des îlots de Mayotte	19
3.1 Découpage du territoire en zones élémentaires	20
3.2 Regroupement algorithmique des zones élémentaires en îlots	21
3.3 Validation des îlots	23
3.4 Impact du redécoupage d'îlot sur la précision des estimateurs	24
Bilan et perspectives	27
Annexe A : Le tirage équilibré et la méthode du cube	29
Annexe B : Les différents plans de sondages envisagés	30
Annexe C : Le détail des estimateurs obtenus dans les simulations sur les grandes communes de Mayotte	31

Remerciements

Pour leurs grandes participations à ce travail, nous voudrions chaleureusement remercier :

- le SR Mayotte : Muriel Granjon, Jamel Mekkaoui, Cédric Mureau, Marc-Antoine Georget ;
- les autres membres du groupe de travail : Cyril Favre-Martinoz (Criem), Pascal Ardilly (SeRN), Thomas Merly-Alpa (Sondages), Lionel Delta (Sondages), Christine Raoul (Dorre)
- Sébastien Hallépée (Dares) pour avoir animé la discussion lors du séminaire DSDS sur ce sujet.

Introduction

Depuis 2004, le recensement est réalisé par sondage en France par les enquêtes annuelles de recensement (EAR). La méthodologie habituellement employée pour le recensement de la population se différencie selon la taille de la commune : en « petite » commune (commune de moins de 10 000 habitants) un recensement exhaustif de la population est effectué tous les ans 5 ans alors qu'en « grande » commune (commune de 10 000 habitants ou plus) une interrogation d'environ 8 % des logements de la commune chaque année est réalisée. Préalablement à ce tirage, dans les DOM, une enquête cartographique est réalisée entre cinq et dix mois avant l'EAR et permet de constituer la base de sondage d'adresses (BSA) dans laquelle seront tirés les logements à recenser au sein de zones de collecte appelées « îlots ».

Jusqu'en 2017, à Mayotte, comme dans la plupart des COM, on procédait (selon l'article 157 de la loi n° 2002-276 relative à la démocratie de proximité et par dérogation de l'article 156) à des recensements généraux de la population tous les cinq ans : le dernier recensement exhaustif effectué à Mayotte a eu lieu en 2017. Depuis que Mayotte est devenu un département français, la volonté des acteurs locaux est d'avoir un recensement similaire à celui des autres départements. Cela a entraîné une modification de la loi de 2002 par la loi n° 2017-256 relative à l'égalité réelle outre-mer (article 147).

La première EAR a eu lieu à Mayotte en 2021. Cependant, les spécificités du terrain mahorais font que le plan de sondage habituellement employé dans les DOM doit s'adapter. En effet, 39 % des logements recensés au RP 2017 sont des habitations de fortune et font partie d'un type d'habitat qui évolue très rapidement et pose des problèmes de repérabilité sur le terrain ce qui risque d'affecter les estimateurs produits par la méthode classiquement appliquée dans les grandes communes des DOM, principalement à cause du défaut de couverture engendré. En ce sens, plusieurs scénarios alternatifs sont évalués à partir de simulations réalisées sur le RP 2017 afin de déterminer une adaptation du protocole classique qui répondrait aux spécificités du terrain mahorais.

Par ailleurs, les îlots mobilisés dans le cadre du RP 2017 sont de grande taille (105 logements par îlots en moyenne à Mayotte dans les grandes communes, contre 58 dans les autres DOM) et hétérogènes. Ainsi, une redéfinition du territoire mahorais en îlots est réalisée en collaboration avec le Service Régional de Mayotte de sorte que leurs frontières soient durablement repérables sur le terrain et qu'ils soient raisonnablement homogènes en taille. Pour ce faire, l'idée est de découper l'ensemble des communes en les plus petites zones (dites « zones élémentaires ») dont on pense que les contours seront durablement repérables sur le terrain et d'appliquer un algorithme de regroupement de ces zones de façon à s'approcher le plus possible d'une taille-cible. Le résultat attendu est un fort gain d'homogénéité en taille des îlots obtenus et une taille moyenne globale contrôlée.

Parmi les 17 communes de Mayotte, dix communes ont plus de 10 000 habitants au dernier RP exhaustif de 2017 : Bandraboua, Bandrele, Dembeni, Dzaoudzi, Koungou, Mamoudzou, Ouangani, Pamandzi, Sada et Tsingoni. Ce document se concentre sur la mise en place des EAR et la redéfinition des îlots dans ces dix communes. Dans les communes de moins de 10 000 habitants, il n'y a pas de sujets d'évaluations de méthodes car le recensement y est exhaustif tous les 5 ans.

1 Les EAR dans les grandes communes des DOM et leurs applications au contexte mahorais

Nous nous intéressons ici au plan de sondage mis en place dans les « grandes » communes des DOM, c'est-à-dire toutes les communes de 10 000 habitants ou plus des départements de Guadeloupe (971), Martinique (972), Guyane (973) et La Réunion (974) ainsi qu'à Saint-Martin (97801) avant de mettre en évidence les problèmes que cette méthode engendrerait si elle était classiquement appliquée à Mayotte.

1.1 L'échantillonnage des EAR dans les grandes communes des DOM

L'échantillon de l'EAR pour ces communes est tiré dans la Base de Sondage d'Adresses (BSA) qui est créée à partir du millésime annuel du Répertoire d'Immeubles Localisés (RIL). Le RIL contient l'ensemble des adresses de logements (les habitations, les établissements touristiques et les communautés) nécessaires au recensement de la population et au calcul des populations légales dans les communes de plus de 10 000 habitants. Il intègre notamment le nombre de logements, le type d'habitation, le caractère habitable des adresses et leur géolocalisation. La BSA est au final un sous-ensemble du RIL dans la mesure où elle ne contient que les adresses d'habitation habitables. Dans les DOM, le RIL est mis à jour annuellement par des enquêtes cartographiques réalisées par l'Insee entre cinq et dix mois avant l'EAR et par les permis (de construire, de démolir ou de réhabilitation) qui sont délivrés. Le RIL est géré par une application appelée « Rorcal »¹ et sa gestion est partagée entre l'Insee et les mairies des communes. Il est validé tous les ans par les communes.

L'échantillon de l'EAR pour les grandes communes des DOM suit un plan de sondage en deux phases² :

1. Première phase de tirage : la constitution des groupes de rotation

En amont de la première EAR en 2004, chaque grande commune (GC) des DOM a été partitionnée en plusieurs ensembles d'adresses contiguës appelés îlots. Dans chaque commune, les îlots ont été répartis en cinq groupes de rotation (GR) par tirage aléatoire équilibré³ avec pour variables d'équilibrage (calculées selon le RP99) :

- ▶ le nombre total de logements ;
- ▶ le nombre de logements en immeuble collectif ;
- ▶ la population selon cinq classes d'âge ;
- ▶ la population par sexe.

Ainsi, toutes les adresses d'un même îlot appartiennent au même groupe de rotation⁴. Le tirage de l'échantillon d'adresses d'une EAR donnée ne mobilise qu'un seul des cinq groupes de rotation, par roulement : par exemple, en 2020, il s'agit du groupe 2. La séquence des groupes de rotation a été déterminée une fois pour toute à l'initialisation du recensement rotatif de la population ; la constitution des groupes de rotation constitue dans ce cadre la première phase de tirage de l'EAR.

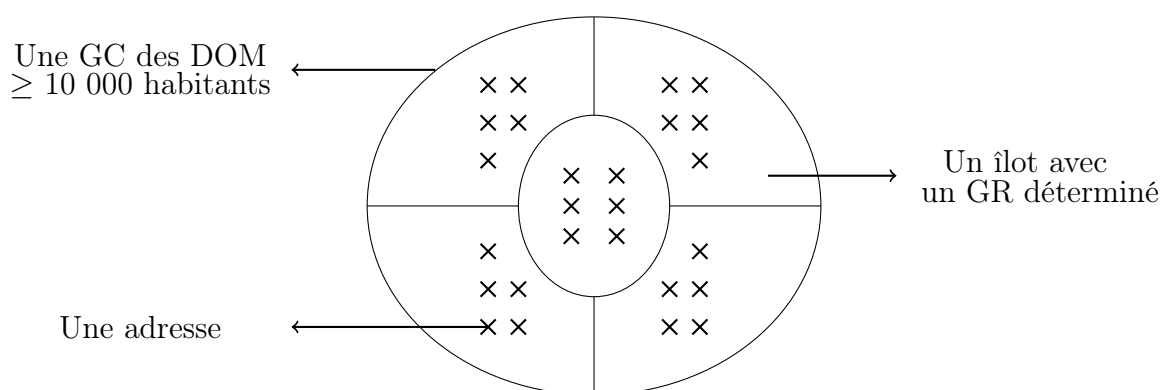
1. RORCAL signifie la Réorganisation du Répertoire Commun d'Adresses Localisées.

2. Pour plus de détails sur la méthode du recensement, le lecteur pourra se référer au n°483-485 d'*Economie et statistique* (2016) portant sur « Le recensement rénové : avancées méthodologiques et apports à la connaissance ».

3. Voir l'annexe A pour plus de détails sur l'échantillonnage équilibré.

4. Quand une petite commune franchit à la hausse le seuil de 10 000 habitants, on réalise également les mêmes opérations avec les données de son dernier recensement exhaustif.

FIGURE 1 – Schématisation d'un découpage en îlot d'une grande commune des DOM



Note : Une GC des DOM a été présentée ici avec seulement 5 îlots pour rester simple. Chaque GR aurait donc un seul et unique îlot dans cette commune. Pour une EAR donnée, l'îlot correspondant constituera la première phase de tirage de l'EAR et les adresses incluses à l'intérieur formeront la base de sondage d'adresses sélectionnables pour cette EAR.

2. Deuxième phase de tirage : stratification et sélection des adresses

On restreint à présent le champ de l'EAR aux adresses incluses dans les îlots dont le groupe de rotation est celui de l'année en cours. Le tirage des adresses est stratifié par type d'adresses avec les strates exhaustives des grandes adresses et des adresses nouvelles : elles sont alors automatiquement ajoutées dans l'échantillon de l'EAR.

On a donc trois strates d'adresses :

- ▶ Si le nombre de logements de l'adresse est supérieur au seuil communal⁵ (en général 60 logements), alors elle est affectée à la strate des grandes adresses.
- ▶ Sinon :
 - ▶ Si l'adresse est considérée comme nouvelle dans la BSA (car jamais enquêtée), alors elle est affectée à la strate des adresses nouvelles.
 - ▶ Sinon l'adresse est affectée à la strate des petites adresses connues.

Nous ajustons ensuite la probabilité d'inclusion des adresses restantes de telle sorte à chercher à recenser 40 % des logements du groupe de rotation annuel (soit 8 % des logements de la commune totale), sachant que l'on recense exhaustivement les logements des grandes adresses et des adresses nouvelles, ce qui constitue déjà une part de ces 40 %. Ce tirage est réalisé de manière équilibrée sur le nombre de logements.

Enfin, les établissements touristiques font également partie de l'échantillon des EAR et sont recensés exhaustivement en fonction du groupe de rotation auxquels ils appartiennent (défini selon leur îlot comme pour logements ordinaires). Leur volume recensé s'ajoute à l'enveloppe des 40 % de logements ordinaires du GR. Leur cas ne sera pas traité au cours de cette étude.

5. Ce seuil communal (en nombre de logements à l'adresse) est déterminé à la livraison annuelle du nouveau RIL de la commune de sorte à ce que la strate des grandes adresses ne dépasse pas un seuil fixé à 10 % des logements de la commune totale.

Ces 2 étapes constituent le tirage d'une enquête annuelle de recensement dans les DOM. Pour produire les estimateurs de population du RP de l'année N , 5 EAR sont requises (celles de l'année $N-2$ à l'année $N+2$) pour obtenir ce qu'on appelle un « cycle RP » : tous les îlots de la grande commune sont donc une seule fois dans le champ d'un cycle RP et on obtient *in fine* un échantillon d'environ 40 % des logements ordinaires de la commune totale.

Le calcul des estimations de population dans les DOM en quelques lignes

Le recensement de la population dans les grandes communes des DOM reposant sur une collecte par sondage, les unités enquêtées (logements, individus) sont associées à un poids qui dépend de la strate de tirage associée à l'adresse enquêtée :

- ▶ pour un logement issu de la strate d'adresse échantillonnée (les petites adresses connues), le poids est défini dans les DOM par une méthode dite « du ratio ». Il correspond alors au nombre total de logements du GR appartenant à cette strate divisé par le nombre total de logements tirés dans cette strate.
- ▶ pour un logement issu d'une strate d'adresse exhaustive, le poids est de 1.

Le poids obtenu pour les logements est le même pour les individus de ces logements. On obtient l'estimateur de population de la grande commune de l'année N , en sommant le nombre d'individus enquêtés et pondérés entre les EAR de l'année $N-2$ à l'année $N+2$ (définissant le cycle RP de l'année N).

1.2 Le cas particulier de Mayotte

Le dernier recensement exhaustif à Mayotte a eu lieu en 2017. Mayotte est constituée de 17 communes dont 10 grandes communes. Ces dernières disposent aujourd'hui d'un faible nombre d'îlots qui sont volumineux et hétérogènes en nombre de logements (25 % des îlots comptent moins de 58 logements, 25 % plus de 129).

	Nombre de grandes communes	Nombre d'îlots	Taille moyenne des îlots (en nb de logements)	Écart-type des tailles d'îlots (en nb de logements)
Guadeloupe	14	2 215	77,7	65
Guyane	12	2 632	57,7	50
Martinique	6	1 502	47,1	62
La Réunion	17	7 144	50,7	62
Mayotte	10	548	104,7	63

Lecture : Les 10 grandes communes de Mayotte contiennent 548 îlots qui ont 104,7 logements en moyenne et une dispersion en taille de logements de 63.

Source : RP 2017.

TABLE 1 – Statistiques descriptives des îlots des GC des DOM et de Mayotte au RP 2017.

Cette situation appelle une réflexion sur les gains à attendre d'un redécoupage des îlots des GC de Mayotte afin de les rendre plus fins et plus homogènes en taille⁶. Ce redécoupage doit cependant être robuste et stable dans le temps pour assurer la qualité des résultats, ce qui réduit significativement la marge de manœuvre et oblige en général à accepter un certain effet de grappe. Du point de vue d'une EAR prise indépendamment⁷, la précision sera d'autant meilleure que les îlots seront plus hétérogènes (en « intra », afin de limiter l'effet de grappe) et de taille moins dispersée.

Pour les unités secondaires d'échantillonnage (les adresses ou autrement dit, les entités adressées), une attention particulière est portée aux habitations en tôle qui constituent 39 % des logements recensés en 2017. Ces habitations, que l'on qualifiera également de logements « non-durs »⁸ (ou de habitations de fortune plus simplement) dans le reste de ce document, correspondent le plus souvent aux constructions précaires faites de panneaux de bois divers et de tôles mais aussi des maisons en ruine ou des bâtiments non destinés à l'habitation mais qui sont habités. Ces habitations, couramment désignées par le terme mahorais de « banga », sont très souvent regroupées en zones de taille très variables selon les communes et rendent le repérage d'un logement tiré difficile, voire impossible.



FIGURE 2 – Photographies de zones à Mayotte principalement constituées d'habitations « non-dures ».

6. On se reportera à la partie 3 du document sur le redécoupage des îlots pour un développement plus explicite des gains attendus.

7. Ou toute autre exploitation ne mobilisant pas l'ensemble des groupes de rotation, par exemple le tirage des enquêtes ménages.

8. À opposer donc aux logements durs qui correspondent aux logements dont l'entité adressée a été récemment construite et dont les fondations et les principaux murs sont réalisés avec des matériaux tels le béton, les parpaings ou les briques.

Rien ne garantit donc que les logements *in fine* recensés sont bel et bien ceux qui ont été tirés au sort, ce qui est susceptible d'induire davantage de variance (échantillon moins bien équilibré) voire du biais si le mécanisme de substitution n'est pas aléatoire (choix des logements dont l'entrée est la plus facilement identifiable, etc.). En effet, pour des problèmes évidents de repérage, l'agent recenseur pourrait ne pas recenser les bons logements, ce qui pourrait engendrer un biais de sélection. De plus ces zones sont très dynamiques du point de vue démographique, ce qui peut rendre les évolutions de l'habitat très importantes entre l'enquête cartographique et l'EAR : le décalage de quelques mois entre l'enquête cartographique servant au sondage et l'arrivée sur le terrain de l'enquête de recensement peut induire un biais de couverture. Tout ceci va donc dans le sens de la difficulté de repérage dans ces zones dans la mesure où l'on ne pourrait pas identifier un logement repéré lors de l'enquête cartographique pour la collecte de l'EAR. Il apparaît ainsi en première approche qu'il est impossible d'échantillonner au sein de ces zones et donc incontournable d'envisager d'autres façons de recenser les adresses dans ces zones.

En ce sens, une grande commune « pilote » de Mayotte a été sélectionnée afin de donner des idées sur les premières méthodes applicables au contexte mahorais. La commune choisie est Bandraboua en raison de son caractère représentatif du contexte (taille moyenne, diversité de la nature de l'habitat selon les villages, présence standard d'habitations en tôle). Un redécoupage en zones géographiquement cohérentes et les plus petites possibles (respectant des frontières identifiables) a pu être fourni rapidement avec en plus, une identification de la qualité de repérage d'un logement dans chaque zone découpée. Ces premiers travaux ont motivé l'élaboration de plans de sondage alternatifs à partir des deux conclusions tirées :

- ▶ les zones où le repérage est difficile (voire impossible) représentent souvent plus de 30 % des logements d'un groupe de rotation hypothétique et donc une part souvent très importante (voire la totalité) de l'enveloppe globale de 40 % de logements à enquêter.
- ▶ la qualification de repérabilité des logements au niveau « zones » apparaît trop complexe sur le terrain : il est donc préférable de raisonner pour la suite directement au niveau « adresses » et donc d'opposer les habitations « durs » et « non-durs ».

1.3 Le problème du biais de couverture de la base de sondage à Mayotte

Lorsque l'on veut réaliser une enquête par sondage, il faut commencer par définir la population sur laquelle on cherche à obtenir de l'information et à estimer des paramètres d'intérêt. La base de sondage est une liste d'unités permettant d'identifier les éléments de cette population, à partir de laquelle on va pouvoir sélectionner un échantillon. Une base de sondage parfaite doit donc couvrir plusieurs caractéristiques pour fournir des estimateurs non-biaisés :

- ▶ La couverture de la population ciblée : on veut que la population présente dans la base de sondage coïncide parfaitement avec la population d'intérêt à la date de l'enquête ;
- ▶ Des informations de localisation : les unités échantillonnées dans la base de sondage doivent se trouver au même endroit lors de l'enquête et être retrouvées sans ambiguïté ;
- ▶ De la richesse d'information : des variables auxiliaires de la population ciblée permettant d'avoir une bonne méthodologie de tirage (stratification, équilibrage,...).

Dans les cas d'une imperfection de la base de sondage, on s'attend à plusieurs conséquences sur les estimateurs produits :

- ▶ Défaut de couverture : on s'attend à la présence de biais et tout particulièrement pour les estimateurs en niveau (pour notre cas par exemple, on s'attend forcément à la présence de biais pour nos estimateurs de nombre de logements et de population) ;
- ▶ Défaut de localisation : on s'attend à la présence de biais dans le cas où l'on substituerait

l'unité sélectionnée par une autre sur le terrain de manière non-aléatoire. On s'attend également à la présence de variance par dégradation de l'équilibre obtenu lors du tirage ;

- Imperfection des variables auxiliaires : on s'attend à la présence de variance car on aurait alors une moins bonne corrélation avec les variables de l'enquête.

Dans notre cas ici, nous allons nous intéresser au défaut de couverture de la base de sondage à Mayotte. En règle générale, le défaut de couverture d'une base de sondage est à relier à un problème d'actualité : la population d'intérêt a beaucoup évolué entre la confection de la base de sondage et la date de validité de l'enquête. Dans ce type de situation, on distingue trois types d'unités.

- Les « présents-présents » : figurent dans la base de sondage et dans la population d'intérêt. Aucun problème ici donc. Dans notre étude, cela représente 37 140 adresses (50 779 logements).
- Les « sortants » : figurent dans la base de sondage mais pas dans la population d'intérêt (destructions, cessations, décès, etc.). Ils vont alors faire augmenter la variance des estimateurs par dégradation de l'équilibre. Dans notre étude, cela représente 5 467 adresses (6 044 logements).
- Les « entrants » : figurent dans la population d'intérêt mais pas dans la base de sondage (constructions, créations, naissances, etc.). Ils vont alors introduire du biais par défaut de couverture dans la base de sondage. Dans notre étude, cela représente 5 642 adresses (6 640 logements).

Un biais lié à un défaut de couverture de la base de sondage s'évalue donc sur la base des adresses « entrantes ». Plus le nombre d'adresses qui apparaissent entre l'enquête cartographique et le RP 2017 est important, plus la sous-estimation de la population sera forte. Ces adresses « entrantes » n'ont en effet aucune chance d'être recensées avec la méthode DOM classique, dans la mesure où elles sont absentes de la base de sondage servant au tirage. Le tableau 2 permet de qualifier ce biais sur deux de nos variables d'intérêt : la population et le pourcentage d'étrangers ⁹.

Ensemble des grandes communes	Présents	Entrants	Présents + Entrants	Biais attendu du fait du défaut de couverture
Population	184 152	23 146	207 298	-11,2 %
Pourcentage d'étrangers	52,3 %	66,8 %	53,9 %	-3,0 %

Lecture : 184 152 individus sont présent dans la BSA et au moment de l'enquête alors que 23 146 individus apparaissent au moment de l'enquête ce qui induit un biais attendu de -11,2 %.

TABLE 2 – Première évaluation du défaut de couverture par comparaison de l'enquête cartographique 2016 et du RP 2017 sur l'ensemble des grandes communes de Mayotte.

Deux pistes ont été poursuivies pour limiter le biais de couverture ou affiner son évaluation. Premièrement, les entités adressées en travaux au moment de l'enquête cartographique ont été ajoutés à la base de sondage d'adresses. Cette piste reprend la démarche déjà mise en place dans les DOM. Cette extension de la base de sondage permet donc une diminution du nombre

9. Ces deux variables seront par ailleurs réutilisées tout au long de cette étude afin d'avoir des évaluations de nos simulations en niveaux et en taux.

d'entrants (on prend en compte ici les constructions entre l'enquête cartographique et l'EAR en les plaçant dans la BSA) mais également une augmentation du nombre de sortants (on a donc *a fortiori* plus d'adresses présentes dans la BSA mais toujours non habitables au moment de l'enquête, et donc hors champ).

La deuxième piste est basée sur la notion de « faux entrants ». Il s'agit d'adresses qui au moment de la collecte du recensement n'ont pas pu être rattachées à leur correspondance dans l'enquête cartographique, alors que ce lien existe bel et bien. Ce problème est dû aux difficultés de repérage lors de la collecte du RP 2017. Ces « faux entrants » augmentent ainsi artificiellement la part des adresses « apparues » (et « disparues ») entre l'enquête cartographique et le RP 2017 et conduisent de fait à une évaluation pessimiste du biais de couverture (évaluation plus forte en valeur absolue que la vraie valeur du biais). Faute de pouvoir corriger précisément ce phénomène, une hypothèse est formulée pour diminuer drastiquement la part des « faux entrants » et aboutir à une évaluation plus optimiste du biais de couverture (évaluation plus faible en valeur absolue que la vraie valeur du biais).

Cette hypothèse se traduit dans la suite de cette étude par deux variantes des scénarios de plans de sondage alternatifs proposés dans la partie 2. Ces deux variantes, basée sur la prise en compte des « faux entrants » permettent d'encadrer le biais de couverture. La première variante, dite « optimiste », est élaborée en réalisant un appariement imparfait entre les entités adressées sortantes et les adresses entrantes au niveau d'un même îlot et d'un même type d'habitat (dur/non-dur). On considère ainsi que les sortants de la BSA dans un îlot sont en fait les entrants présents au moment de l'EAR (en affinant au niveau du type d'habitat). Une seconde variante dite « très optimiste » est également envisagé en appariant les entités adressées sortantes et les adresses entrantes simplement au sein d'un même îlot (même quand le type d'habitat diffère).

<i>Ensemble des grandes communes</i>	Biais attendu du fait du défaut de couverture	
	Population	Pourcentage d'étrangers
En excluant les EA en travaux	-11,2 %	-3,0 %
En intégrant les EA en travaux	-9,1 %	-3,0 %
Hypothèse optimiste sur les « faux entrants »	-3,8 %	-1,5 %
Hypothèse très optimiste sur les « faux entrants »	-3,1 %	-1,2 %

Lecture : Avec l'hypothèse optimiste effectuée sur les « faux entrants », on s'attend à un biais en population de -3,8 % du défaut de couverture sur l'ensemble des grandes communes de Mayotte avec la méthode DOM.

TABLE 3 – Évaluation du défaut de couverture de la méthode DOM sous hypothèses par comparaison de l'enquête cartographique 2016 et du RP 2017 sur l'ensemble des grandes communes de Mayotte.

De ces hypothèses et des résultats obtenus, on peut en déduire que le biais de couverture de la base de sondage du RP 2017 construite à partir de l'enquête cartographique de 2016 est situé entre -9,1 % et -3,8 % en population pour l'ensemble des grandes communes de Mayotte. L'impact du défaut de couverture de la base de sondage avec la méthode classique employée dans les DOM pour les EAR nous pousse alors à rechercher des scénarios de tirages alternatifs qui répondraient au contexte mahorais et permettraient de limiter ce biais observé dans nos simulations de tirage.

1.4 Le déroulement des travaux pour préparer le lancement des EAR à Mayotte

Un groupe de travail méthodologique a été formé afin d'échanger sur ce qui peut être réalisé théoriquement et pratiquement pour lancer les EAR à Mayotte. Concernant les travaux méthodologiques en grande commune, les motivations principales au sein de ce groupe de travail étaient :

- ▶ d'anticiper l'impact de l'importance de l'habitat de fortune sur les performances de la méthodologie habituellement mise en œuvre dans les DOM et de proposer, le cas échéant, une stratégie d'échantillonnage alternative adaptée aux particularités mahoraises.
- ▶ de redéfinir le territoire en îlots de taille plus faible et la plus homogène possible.

Pour l'évaluation des performances des différents plans de sondages (incluant celui classiquement employé dans les DOM), les données utilisées sont basées sur le dernier RP exhaustif de 2017 du territoire mahorais. L'utilisation de ce recensement est importante pour cette étude dans la mesure où l'on dispose ainsi d'une base de données exhaustive avec une information raisonnablement fraîche. De plus, une enquête cartographique (elle-même exhaustive) a précédé le RP 2017 et nous permet de disposer d'une géolocalisation fine des entités adressées : il est ainsi possible de localiser les adresses du RP dans n'importe quelle partition du territoire. On peut dès lors observer le décalage (de plusieurs mois) entre enquête cartographique et l'enquête du recensement exhaustif qui sera analogue à celui qui surviendra pour les EAR.

L'appariement entre les données du RP et les données de RORCAL

Dans le cadre de cette étude, il est essentiel de pouvoir géolocaliser finement l'intégralité des logements enquêtés lors du dernier RP par le biais de leurs coordonnées XY. Or cette géolocalisation n'est pas collectée en même temps que les autres caractéristiques des logements, les agents recenseurs ne disposant pas d'instruments de géolocalisation. C'est le cas en revanche lors de l'enquête cartographique, dont les données géolocalisées sont stockées dans l'application de gestion RORCAL.

Un appariement entre les données de l'enquête cartographique et de la collecte du recensement est possible dans la mesure où elles disposent d'un identifiant commun, mais toutes les adresses collectées lors du RP ne sont pas retrouvées dans RORCAL (construction, destruction entre l'enquête cartographique et le RP, problèmes de repérage sur le terrain). Des coordonnées XY ont donc dû être imputées aux adresses du RP qu'il n'a pas été possible d'apparier avec l'enquête cartographique. Pour ce faire, on exploite l'îlot de collecte (toujours renseigné) : en pratique, les adresses sans XY sont positionnées aléatoirement à proximité (10m) des autres adresses de l'îlot ou des adresses des îlots voisins situés proches de la limite d'îlot. De la sorte, la géolocalisation imputée est crédible dans la mesure où elle évite de positionner des adresses dans des endroits reculés (au milieu de la forêt par exemple).

Nombre total d'adresses dans le RP 2017	Proportion de l'appariement avec l'id commun	Proportion de l'appariement par imputation
42 782	95,75 %	4,25 %

Lecture : 95,75 % des adresses collectées lors du RP de 2017 ont été retrouvées dans RORCAL afin d'obtenir leur géolocalisation.

TABLE 4 – Statistiques sur l'appariement entre données du RP 2017 des grandes communes de Mayotte et celles de RORCAL.

Pour préparer le lancement des EAR à Mayotte, un découpage de l'ensemble du territoire en zones élémentaires a été réalisé par le Service Régional de Mayotte. Ce découpage, à partir de critères uniquement géographiques (pérennité des limites), définit une zone élémentaire comme la plus petite zone dont on pense que ses limites seront durablement repérables sur le terrain. Plusieurs stratégies d'échantillonnage ont été élaborées ensuite et évaluées par simulations. Leurs performances ont été comparées en effectuant, pour chaque scénario envisagé, 2 000 tirages de cycles RP (comprenant donc 5 tirages d'EAR chacun) à partir de la base du RP 2017 constituée. Les scénarios les plus pertinents sont détaillés en partie 2 de ce document. Ces scénarios sont évalués sur le découpage en zones élémentaires car le regroupement en îlot (voir partie 3) n'était pas encore réalisé.

Chacune des simulations produit un échantillon d'adresses et un jeu de pondérations obtenu de manière similaire à celle de la méthode classique des DOM¹⁰. Ces poids permettent de calculer pour chaque variable d'intérêt de notre étude (population, pourcentage d'étrangers, classes d'âge, ...) une estimation obtenue sur chaque simulation à partir de l'échantillon sélectionnée pour chaque scénario élaboré. La distribution des 2 000 estimations de chacune de ces variables d'intérêt estimées est finalement comparée à la vraie valeur connue dans le RP 2017 pour déterminer le biais et la variance de l'estimateur.

Autrement dit, soit Y une de nos variables d'intérêt pour ce tirage, on peut noter t_y son total dans le RP 2017 (qui est considéré ici comme sa vraie valeur) et $\hat{t}_{y,(i)}$ son estimateur du total dans le i ème tirage de cycle RP simulé. Alors on a sur un nombre de N simulations pour chaque scénario :

$\bar{\hat{t}}_y = \sum_{i=1}^N \frac{\hat{t}_{y,(i)}}{N}$ qui converge fortement vers $\mathbb{E}(\hat{t}_y)$ et qui correspond à la moyenne de l'estimateur de notre variable d'intérêt sur l'ensemble des simulations de cycles RP sur ce scénario.

Alors on obtient ceci pour le biais relatif et la variance de l'estimateur :

$$\hat{B}_r(\hat{t}_y) = \frac{(\bar{\hat{t}}_y - t_y)}{t_y} \quad \text{et} \quad \text{Var}(\hat{t}_y) = \frac{1}{N-1} \sum_{i=1}^N (\hat{t}_{y,(i)} - \bar{\hat{t}}_y)^2$$

Et enfin ceci pour le coefficient de variation de l'estimateur : $CV(\hat{t}_y) = \frac{\sqrt{\text{Var}(\hat{t}_y)}}{t_y}$

Au final, 62 scénarios ont été comparés (23 plans de sondage avec à chaque fois plusieurs variantes : hypothèses ou estimateurs différents), ce qui représente de l'ordre de 460 000 simulations de tirage.

10. Méthode du « ratio » pour la strate des petites adresses connues et une valeur de 1 pour les strates exhaustives, voir encadré sur le calcul des population en partie 1.1.

2 Les scénarios de tirage alternatifs

L'annexe B présente une liste des différents plans de sondage qui ont été envisagés pour la préparation des EAR à Mayotte. Dans cette partie, nous ne présenterons que les 4 scénarios de tirage qui ont montré les meilleurs résultats avec une réelle possibilité d'application sur le terrain. Dans cette partie, nous commençons tout d'abord par détailler les résultats obtenus à Mayotte avec la méthode classique des DOM. Puis, nous présentons un à un les scénarios proposés, en estimant systématiquement le biais et le coefficient de variation des estimateurs pour chaque situation et en dressant la liste des principaux avantages et inconvénients de chaque méthode.

Les différents tableaux de cette partie présente la précision obtenue sur deux de nos variables d'intérêt (la population et le pourcentage d'étrangers) sur l'ensemble des grandes communes de Mayotte ainsi que leurs *extremums* au niveau communal. On peut se reporter à l'annexe C pour avoir l'évaluation complète de toutes les grandes communes de Mayotte.

2.1 Scénario 1 : Méthodologie DOM classique

Le scénario 1 de tirage pour les EAR à Mayotte correspond à la méthode décrite en partie 1.1. Ce scénario calque donc la méthode en place dans les DOM. Elle se résume en trois étapes :

1. Réalisation entre avril et août $N-1$ d'une enquête cartographique exhaustive sur le groupe de rotation dans lequel sera réalisé l'EAR N .
2. Tirage dans l'enquête cartographique d'un échantillon d'adresses de sorte qu'il représente environ 40 % des logements de l'enquête cartographique.
3. Recensement en février-mars N des seules adresses tirées.

Avec 2 000 simulations de cycles RP selon ce scénario, nous pouvons évaluer, la méthode DOM soumise à l'hypothèse optimiste sur les « faux entrants » permet d'avoir un encadrement des estimateurs. Dans ces évaluations, la strate des adresses « neuves » n'apparaît pas dans les tirages car elle ne peut être simulée avec les éléments dont nous disposons.

<i>Ensemble des grandes communes Minimum parmi les GC / Maximum parmi les GC</i>	Population		Pourcentage d'étrangers	
	Biais	CV	Biais	CV
Méthode DOM	-9,1 % -18,3 % -2,6 %	0,5 % 0,9 % 2,5 %	-2,9 % -10,6 % -0,6 %	0,5 % 0,8 % 4,2 %
Méthode DOM hypothèse optimiste sur les « faux entrants »	-3,8 % -7,7 % -0,5 %	0,5 % 0,9 % 2,4 %	-1,5 % -5,6 % -0,3 %	0,5 % 0,8 % 4,0 %
Méthode DOM avec base de sondage parfaite	0,0 % 0,0 % 0,0 %	0,4 % 0,7 % 1,8 %	0,0 % -0,1 % 0,0 %	0,5 % 0,8 % 3,9 %

Lecture : Avec l'hypothèse optimiste effectuée sur les « faux entrants », on s'attend à un biais en population de -3,8 % du défaut de couverture sur l'ensemble des grandes communes de Mayotte. Avec une base de sondage parfaite (sans entrants ni sortants), on observe comme attendu qu'il n'y a aucun biais.

TABLE 5 – Caractéristiques des estimateurs de la population et du pourcentage d'étrangers avec le scénario 1.

Le principal avantage de cette méthode est d'avoir évidemment une méthodologie identique à celle mise en œuvre dans les autres DOM, ce qui permettrait d'avoir une forme d'égalité sur la façon d'enquêter les DOM pour les EAR. De plus, cela permettrait aussi d'éviter des coûts de développements spécifiques pour Mayotte au niveau informatique ou protocolaire. Autre point positif de cette méthode : elle produit des variances d'estimateurs plutôt faibles.

Mais cette méthode génère un biais de couverture important : entre -9,1 % et -3,8 % de la population et entre -2,9 % et -1,5 % sur le taux d'étrangers sur l'ensemble des grandes communes de Mayotte. Le niveau de la population (et des autres variables) est donc mal estimé. Cela pose notamment un problème important lors de la diffusion des premiers résultats du nouveau recensement à Mayotte, ce qui pourrait discréditer la méthode employée dans le recensement en général. Enfin, cette méthode n'apparaît pas envisageable du fait du problème de repérabilité des adresses tirées sur le terrain : l'agent recenseur pourrait ne pas recenser les bons logements, ce qui pourrait engendrer un biais supplémentaire et aussi une variance.

2.2 Scénario 2 : Tirage d'îlots à recenser exhaustivement

Le scénario 2 propose de tirer non pas des adresses mais des îlots et d'y enquêter l'ensemble des adresses à l'intérieur de ceux-ci. L'idée est de pallier à la difficulté de repérage des logements à l'intérieur des îlots en les recensant exhaustivement. En reprenant le calendrier d'une EAR des DOM, le scénario 2 consiste à :

1. Réaliser entre avril et août $N-1$ d'une enquête cartographique sur le groupe de rotation dans lequel sera réalisé l'EAR N .
2. Tirer dans l'enquête cartographique d'un échantillon d'îlots de sorte que les logements de ces îlots représentent environ 40 % des logements de l'enquête cartographique.
3. Recenser en février-mars N de l'ensemble des adresses figurant dans les limites des îlots tirés au moment de l'enquête (recensement exhaustif des îlots tirés).

Nous mettons ici les estimateurs obtenus avec 2 000 simulations de cycles RP selon ce scénario en comparaison avec le scénario 1 évoqué plus haut.

<i>Ensemble des grandes communes Minimum parmi les GC / Maximum parmi les GC</i>	Population		Pourcentage d'étrangers	
	Biais	CV	Biais	CV
Scénario 1 (méthode classique DOM)	-9,1 % -18,3 % -2,6 %	0,5 % 0,9 % 2,5 %	-2,9 % -10,6 % -0,6 %	0,5 % 0,8 % 4,2 %
Scénario 2 (tirage d'îlots)	0,0 % -0,5 % 0,3 %	1,2 % 2,2 % 4,9 %	-0,2 % -2,0 % -0,1 %	1,4 % 2,4 % 7,4 %

Lecture : Dans le scénario 2, l'estimateur de la population sur l'ensemble des grandes communes de Mayotte est sans biais. Son coefficient de variation est de l'ordre de 1,2 %. Le coefficient de variation de l'estimateur de la population est compris entre 2,2 % et 4,9 % selon la commune.

TABLE 6 – Caractéristiques des estimateurs de la population et du pourcentage d'étrangers avec le scénario 2.

Ce scénario protège complètement du défaut de couverture de la base de sondage. Pour certaines communes, il peut y avoir un biais résiduel qui est dû à la méthode d'estimation par le ratio asymptotiquement sans biais (on n'est pas ici précisément dans un contexte asymptotique, le nombre d'îlots tirés par commune pouvant être relativement faible). De plus, en anticipant une mise en place de cette méthode avant une éventuelle bascule vers la méthode classique DOM (scénario 1), on pourrait mesurer ici périodiquement le biais de couverture ce qui sera très utile pour nos estimations lorsque la méthode classique DOM est mise en place.

Néanmoins avec cette méthode, les estimateurs présentent une très forte variance à cause de l'effet de grappe engendré par ce type de tirage. Ceci peut rendre délicat l'interprétation des évolutions et remettre en cause toute ou partie de la diffusion des statistiques infra-communales (au village ou à l'IRIS). On peut également déjà anticiper que certains îlots contiendront à terme beaucoup de logements, ce qui réduira la précision par amplification de l'effet de grappe. Des scissions d'îlots pourront cependant être opérés si des frontières repérables le permettent.

Autre inconvénient de cette méthode, le tirage en grappe entraîne ici une forte instabilité de la taille des échantillons : il arrive qu'il faille réitérer le tirage pour que le taux de sondage soit compris entre nos bornes classiques fixées pour le tirage équilibré des adresses. Ces bornes correspondent à obtenir un échantillon représentant entre 37 % et 45 % (pour une cible de 40 %) des logements du groupe de rotation : avoir un encadrement du tirage de 2e phase dans la méthode classique des DOM est essentiel dans la mesure où on tire des adresses en ciblant un nombre de logements ce qui explique une instabilité autour de notre cible de 40 % des logements. Pour ce scénario 2, cette instabilité est amplifiée dans la mesure où on tire des îlots (qui sont plus hétérogènes et plus volumineux que des adresses) en ciblant un nombre de logements.

Par ailleurs, ce scénario (et comme les 2 autres alternatifs qui sont présentés ci-après) présente l'inconvénient de forcer une évolution du protocole mis en place dans les DOM et des différents traitements engendrant la modification de certaines applications informatiques (tirage, collecte, ...).

Enfin, pour cette méthode, nous avons testé d'accroître le taux de sondage pour les grandes communes à Mayotte afin d'améliorer la précision des estimations. Cette variante présente les mêmes inconvénients que le scénario 3 en termes de charges de collecte et de coûts.

2.3 Scénario 3 : Adaptation du traitement des habitations de fortune sur celui des habitations mobiles et des sans-abris (HMSA)

Le troisième plan de sondage proposé consiste à adapter le traitement actuellement mis en place dans le RP pour les habitations mobiles et des sans-abris (HMSA) aux habitations de fortune à Mayotte. Ces habitations sont recensés une fois tous les 5 ans en grande commune sur l'ensemble du territoire. L'idée ici est toujours de pallier au problème de défaut de couverture en ciblant la difficulté de repérage des habitations de fortune. Dans ce scénario, on les recense exhaustivement selon le groupe de rotation de l'îlot auxquels ils appartiennent.

La méthode du scénario 3 proposé peut être résumée ainsi en suivant le calendrier classique d'une EAR des DOM :

1. Réalisation entre avril et août $N-1$ d'une enquête cartographique sur le groupe de rotation dans lequel sera réalisé l'EAR N .
2. Tirage parmi les logements « en dur » de l'enquête cartographique d'un échantillon d'adresses de sorte qu'il représente environ 40 % des logements « en dur » de l'enquête cartographique (selon la méthode classique utilisée dans les autres DOM, mais restreinte aux logements « en dur »).

3. Recensement en février-mars N :

- ▶ parmi les logements « en dur » : des seules adresses tirées ;
- ▶ de tous les habitations de fortune localisés au début de la collecte du recensement dans les îlots du groupe de rotation enquêté. Afin de lisser la charge de collecte et de réduire les ruptures dans les estimations de population, ce recensement exhaustif a lieu tous les ans sur l'ensemble des îlots du groupe de rotation enquêté : cela constitue une variante par rapport au protocole HMSA classique (habituellement recensés une fois tous les 5 ans en grande commune sur l'ensemble du territoire).

Les 2 000 simulations de cycles RP selon ce scénario permettent de comparer les estimateurs avec ceux des 2 scénarios évoqués plus haut. De plus, ce scénario peut être évalué avec l'hypothèse optimiste sur les « faux entrants » afin d'obtenir un encadrement des résultats observés.

<i>Ensemble des grandes communes</i> <i>Minimum parmi les GC / Maximum</i> <i>parmi les GC</i>	Population		Pourcentage d'étrangers	
	Biais	CV	Biais	CV
Scénario 1 (méthode classique DOM)	-9,1 % -18,3 % -2,6 %	0,5 % 0,9 % 2,5 %	-2,9 % -10,6 % -0,6 %	0,5 % 0,8 % 4,2 %
Scénario 2 (tirage d'îlots)	0,0 % -0,5 % 0,3 %	1,2 % 2,2 % 4,9 %	-0,2 % -2,0 % -0,1 %	1,4 % 2,4 % 7,4 %
Scénario 3 (recensement exhaustif des habitations de fortune)	-2,0 % -5,3 % -0,4 %	0,4 % 0,8 % 2,4 %	0,6 % 0,1 % 1,4 %	0,4 % 0,6 % 3,5 %
Scénario 3 avec hypothèse optimiste sur les « faux entrants »	-0,3 % -0,9 % 0,0 %	0,4 % 0,7 % 2,2 %	0,1 % -0,1 % 0,6 %	0,4 % 0,6 % 3,6 %

Lecture : Dans le scénario 3, l'estimateur de la population sur l'ensemble des grandes communes de Mayotte présente un biais compris entre -2,0 % et -0,3 % selon les hypothèses formulées sur l'importance des nouvelles adresses entre enquête cartographique et RP 2017. Dans l'hypothèse la plus optimiste, le biais en population est compris entre -0,9 % et 0,0 % selon les communes.

TABLE 7 – Caractéristiques des estimateurs de la population et du pourcentage d'étrangers avec le scénario 3.

Les résultats de ce scénario sont très avantageux au niveau de la précision : elle est globalement comparable à la méthode classique DOM et très satisfaisante sur les variables relatives aux habitations de fortune (voir l'annexe C). De plus, on réduit ici sensiblement le biais de couverture, qui n'est pas de nature à être supérieur à celui présent dans les autres DOM : le biais résiduel est dû aux adresses entrantes (entre l'enquête cartographique et l'EAR) qui ont un type d'habitat « dur » (les habitations de fortune étant recensés exhaustivement au moment de l'EAR, ils n'ont pas de problème de couverture dans la base de sondage). Ce scénario fait donc globalement aussi bien que la méthode DOM sur la précision et gomme en plus drastiquement notre problème de biais de couverture.

Le principal problème de ce scénario est sur la charge que celui-ci impose aux communes. En effet, du fait du recensement exhaustif des habitations de fortune, l'effort de collecte pour les

communes excède largement 8 % de leurs logements chaque année. Il s'établit entre 12 % et 14 % selon la commune¹¹, soit un recensement de 60 à 70 % des logements selon la commune. Un autre inconvénient de ce scénario est de suivre une application stricte du protocole HMSA pour les habitations de fortune de Mayotte : ils ne sont alors pas comptabilisés parmi les logements ordinaires dans le cadre du recensement. Ce non-comptage a plusieurs conséquences :

- ▶ baisse du nombre de logements à Mayotte par rapport au RP 2017 ;
- ▶ pas de résultats sur les caractéristiques de ces logements ;
- ▶ impact sur les dotations forfaitaires de recensement (DFR).

Pour pallier cela, il est possible de faire remplir une feuille de logement¹² pour ces habitations de fortune, mais cela implique de faire une exception par rapport aux autres territoires.

Enfin, dernier point négatif : les consignes pour les agents recenseurs sont plus complexes que dans les deux premiers scénarios. En effet, avec cette méthode, on aurait alors deux protocoles distincts selon le type de bâti et un besoin d'identification claire sur le terrain du bon type de bâti.

Une variante de ce scénario a été envisagée pour alléger l'effort de collecte pour les communes en tirant au sort sur le terrain les habitations de fortune à recenser (à hauteur de 40 %) : elle a été finalement écartée car jugée trop complexe, encore coûteuse et susceptible d'être affectée par des biais de la part des agents recenseurs dans le choix des logements à recenser.

2.4 Scénario 4 : Tirage en îlots uniquement pour les habitations de fortune

Le scénario 4 est le dernier plan de sondage alternatif proposé ici : il s'agit en fait d'une combinaison des scénarios 1 et 2 dans la mesure où on va différencier le tirage selon le type d'habitat (tirage classique pour les logements en dur et tirage en îlots pour les habitations de fortune). On stratifie donc le tirage tout en visant toujours au global 40 % des logements du groupe de rotation. L'idée derrière cette stratification est à nouveau de pallier la difficulté de repérage des habitations de fortune en les recensant exhaustivement dans l'îlot auquel ils appartiennent et le défaut de couverture pour les logements non durs. Cependant, alors que le scénario 2 envisageait cela pour tous les types d'adresses, on limite l'effet grappe ici en échantillonnant les entités adressées « en dur » directement. La méthode de ce scénario 4 peut être résumée ainsi (toujours selon le calendrier classique d'une EAR des DOM) :

1. Réalisation entre avril et août $N-1$ d'une enquête cartographique sur le groupe de rotation dans lequel sera réalisé l'EAR N .
2. Tirage parmi les logements « en dur » de l'enquête cartographique d'un échantillon d'adresses de sorte qu'il représente environ 40 % des logements « en dur » de l'enquête cartographique (selon la méthode classique utilisée dans les autres DOM, mais restreinte aux logements « en dur »).
3. Tirage dans l'enquête cartographique d'un échantillon d'îlots de sorte que les habitations de fortune de ces îlots représentent environ 40 % des habitations de fortune de l'enquête cartographique.
4. Recensement en février-mars N :
 - ▶ parmi les logements « en dur » : des seules adresses tirées ;

11. Seule la commune de Sada aurait un taux de sondage inférieure : 10 % pour une enquête et 50 % sur 5 ans.

12. La feuille de logement est un formulaire à remplir lors de l'EAR contenant toutes les questions relatives aux caractéristiques et confort du logement, puis aux personnes vivant habituellement dans le logement.

- parmi les habitations de fortune : de tous les habitations de fortune des seuls îlots tirés.

Les 2 000 simulations de cycles RP selon ce scénario permettent de comparer les estimateurs avec ceux des 3 scénarios évoqués plus haut. Ce scénario peut également être évalué avec l'hypothèse optimiste sur les « faux entrants » afin d'obtenir un encadrement des résultats observés.

<i>Ensemble des grandes communes</i> <i>Minimum parmi les GC / Maximum</i> <i>parmi les GC</i>	Population		Pourcentage d'étrangers	
	Biais	CV	Biais	CV
Scénario 1 (méthode classique DOM)	-9,1 % -18,3 % -2,6 %	0,5 % 0,9 % 2,5 %	-2,9 % -10,6 % -0,6 %	0,5 % 0,8 % 4,2 %
Scénario 2 (tirage d'îlots)	0,0 % -0,5 % 0,3 %	1,2 % 2,2 % 4,9 %	-0,2 % -2,0 % -0,1 %	1,4 % 2,4 % 7,4 %
Scénario 3 (recensement exhaustif des habitations de fortune)	-2,0 % -5,3 % -0,4 %	0,4 % 0,8 % 2,4 %	0,6 % 0,1 % 1,4 %	0,4 % 0,6 % 3,5 %
Scénario 4 (tirage mixte)	-2,2 % -5,6 % -0,9 %	0,9 % 1,4 % 3,9 %	0,4 % -0,6 % 1,4 %	0,8 % 1,3 % 4,1 %
Scénario 4 avec hypothèse optimiste sur les « faux entrants »	-0,4 % -1,1 % 0,1 %	0,8 % 1,4 % 3,7 %	-0,1 % -1,0 % 0,5 %	0,8 % 1,3 % 4,0 %

Lecture : Dans le scénario 4, l'estimateur de la population sur l'ensemble des grandes communes de Mayotte présente un biais compris entre -2,2 % et -0,4 % selon les hypothèses formulées sur l'importance des nouvelles adresses entre enquête cartographique et RP 2017. Dans l'hypothèse la plus optimiste, le biais en population est compris entre -1,1 % et 0,1 % selon les communes.

TABLE 8 – Caractéristiques des estimateurs de la population et du pourcentage d'étrangers avec le scénario 4.

Les résultats de ce scénario sont plus avantageux que ceux du scénario 2 (tirage complet d'îlots) au niveau de la précision. De plus, on réduit à nouveau ici sensiblement le biais de couverture par rapport au scénario classique DOM (scénario 1). Il n'est pas de nature à être supérieur à celui actuellement présent dans les autres DOM : le biais résiduel est dû aux adresses entrantes (entre l'enquête cartographique et l'EAR) qui ont un type d'habitat « dur » (les habitations de fortune étant recensés exhaustivement au moment de l'EAR, ils n'ont pas de problème de couverture dans la base de sondage). Enfin, cette méthode mixte a une précision plus forte que celle du scénario précédent mais permet de maintenir à 8 % de leurs logements l'effort de collecte des communes.

Outre la modification des applications de pré-collecte et de collecte pour produire ici une collecte HMSA annuelle en grande commune, l'inconvénient majeur de cette méthode est les consignes complexes pour les agents recenseurs. On aurait toujours ici deux protocoles distincts selon le type de bâti et un besoin d'identification claire sur le terrain du bon type de bâti. Cette complexité est amplifiée dans ce scénario dans la mesure où la collecte diffère selon les îlots : certains où il y a seulement des adresses « dures » à recenser, d'autres où uniquement toutes les adresses « non-dures » sont à recenser, et enfin certains avec les deux cas.

2.5 Bilan des scénarios envisagés

Avant de dresser le bilan des scénarios envisagés, nous pouvons également enfin présenter les résultats bruts des simulations sur ces différents plans de sondages. L'idée est de les mettre directement en opposition avec la vraie valeur du RP 2017 sur l'ensemble des grandes communes : **207 298 habitants en logement ordinaire et 53,9 % d'étrangers**. On peut ainsi ici mieux se rendre compte de l'éloignement à la réalité, et de l'ampleur de l'imprécision. Dans le tableau 9, nous présentons ainsi sur l'ensemble des GC de Mayotte les bornes inférieures et supérieures obtenues lors des simulations pour les estimateurs de population et de pourcentage d'étrangers.

<i>Ensemble des grandes communes</i>	Population		Pourcentage d'étrangers	
	Borne inf	Borne sup	Borne inf	Borne sup
Scénario 1 (méthode classique DOM)	197 473	201 565	52,5 %	53,6 %
Scénario 2 (tirage d'îlots)	202 662	212 160	52,3 %	55,2 %
Scénario 3 (recensement exhaustif des habitations de fortune)	205 054	208 441	53,5 %	54,4 %
Scénario 4 (tirage mixte)	203 093	209 717	53,1 %	54,7 %

Lecture : En 2017, l'ensemble des GC de Mayotte comptait environ 207 300 habitants en logement ordinaire. 95 % des estimateurs obtenus lors de l'évaluation du scénario 3 sont compris entre 205 054 et 208 441 habitants.

TABLE 9 – Comparaisons brutes des simulations sur les variables de population et de pourcentage d'étrangers avec les différents scénarios.

L'ensemble des scénarios envisagés a montré tout d'abord que la méthodologie mise en œuvre habituellement dans les DOM (scénario 1) conduit à des biais, potentiellement forts : en niveau, la population de l'ensemble des grandes communes de Mayotte serait sous-estimée entre 3,8 % et 9,1 % (selon les hypothèses formulées sur l'importance des adresses apparues entre l'enquête cartographique et le RP 2017). Un biais, plus faible, est également observé sur les variables en structure.

La seconde méthodologie de tirage proposé ici consiste à tirer non pas des adresses mais directement des îlots (scénario 2). Les adresses de ces îlots sont ensuite recensés exhaustivement. Cette méthodologie corrige complètement le biais de couverture au prix d'une forte dégradation de la précision : les coefficients de variation des estimateurs sont multipliés par 2 ou 3 selon les variables d'intérêt et les communes.

Le troisième scénario évalue l'impact d'un traitement des habitations de fortune à Mayotte similaire à celui mis en place pour les habitations mobiles et les sans-abris (HMSA), qui sont recensés exhaustivement (scénario 3). Cette méthode consiste en un recensement exhaustif des habitations de fortune sur le cycle de 5 ans. Ce protocole permet de réduire fortement le biais de couverture pour le rendre comparable à celui rencontré dans les autres DOM, sans augmenter l'imprécision des résultats par rapport à la méthode DOM habituelle. En revanche, elle induit une charge beaucoup plus importante pour les communes : entre 60 % et 70 % des logements des grandes communes sont alors recensés sur un cycle de 5 ans (contre 40 % avec la méthode habituelle). Le taux de sondage sur un cycle de 5 ans est de 65 % en moyenne sur les grandes communes de Mayotte.

Enfin, la dernière méthode proposée pour les EAR à Mayotte est une combinaison des deux précédentes : tirage au niveau des adresses pour les logements en dur et tirage au niveau des îlots pour un recensement exhaustif des habitations de fortune. Ce dernier scénario permet d'obtenir des résultats aussi satisfaisants que le scénario 3 en terme de biais de couverture et une meilleure variance des estimateurs que ceux obtenus dans le scénario 2 où l'on tirait seulement au niveau des îlots. De plus, contrairement au scénario 3, cette méthode permet de maintenir la charge de collecte annuelle des commune au seuil de 8 % de ses logements. Cependant, sa forte complexité est un frein à sa mise en place immédiate.

	Biais	Variance	Charge pour les communes	Évolutions spécifiques des EAR pour Mayotte
Scénario 1 (méthode classique DOM)	Fort	Faible	Standard	Non
Scénario 2 (tirage d'îlots)	Nul	Forte	Standard	Oui
Scénario 3 (recensement exhaustif des habitations de fortune)	Comparable aux autres DOM	Faible	Beaucoup plus importante	Oui
Scénario 4 (tirage mixte)	Comparable aux autres DOM	Moyenne	Standard	Oui

TABLE 10 – Tableau de synthèse des différents plans de sondage envisagés pour le lancement des EAR à Mayotte.

Le tableau 10 montre qu'aucun scénario remplit tous les critères de manière optimale. Le groupe de travail méthodologique est davantage favorable au scénario 3 (recensement exhaustif des logements de fortune par adaptation HMSA) qui présente les meilleures propriétés statistiques. Le directeur général de l'Insee a arbitré en faveur de ce scénario sur les recommandations du groupe de travail. Le plan de sondage des EAR à Mayotte consiste donc, pour chaque commune, à tirer parmi les adresses « en dur » un échantillon d'adresses de sorte qu'il représente environ 40 % des logements « en dur » du groupe de rotation de l'année. Lors du recensement, on enquête alors tous les logements des adresses « en dur » tirées ainsi que tous les habitations de fortune localisés au début de la collecte du recensement dans les îlots du groupe de rotation de l'année.

3 Le redécoupage des îlots de Mayotte

Comme dans les autres DOM, l'organisation du recensement à Mayotte s'appuie sur un découpage fin du territoire en îlots, dans lesquels l'enquête cartographique est réalisée (chaque année, 20 % des îlots sont repérés exhaustivement au moment de l'enquête cartographique). De ce fait, les caractéristiques de ces îlots (taille, plus ou moins grande hétérogénéité, etc.) ont une influence sur les performances du plan de sondage du RP. Or les îlots utilisés lors du recensement exhaustif de la population de 2017 sont sensiblement plus gros que ceux des autres DOM (Tableau 11) : les réutiliser pour le plan de sondage des EAR à Mayotte conduirait mécaniquement à un effet de grappe supérieur à celui rencontré dans les autres DOM, ce qui induirait une augmentation de la variance des estimateurs.

Statistiques (en nb de logements)	Îlots 2017 GC Mayotte	Îlots GC DOM
Taille moyenne	105	58
Écart-type	63	61
Minimum	1	0
Premier décile	15	6
Premier quartile	63	18
Médiane	103	40
Troisième quartile	143	77
Neuvième décile	183	128
Maximum	513	1 055

TABLE 11 – Caractéristiques des îlots du RP 2017 à Mayotte en GC

Par ailleurs, il est ressorti des échanges avec le service régional de Mayotte que les limites des îlots du RP 2017 sont susceptibles de ne pas être durablement repérables sur le terrain, du fait des fortes évolutions du bâti à Mayotte dues à la très forte dynamique démographique. Ce point est particulièrement problématique, dans la mesure où il est de nature à rendre la mise en œuvre des enquêtes cartographiques et de la collecte de l'EAR difficile voire impossible sur le terrain.

Ces éléments ont conduit à engager des travaux sur le redécoupage des îlots de Mayotte en grande commune, avec un triple objectif : constituer de nouveaux îlots (i) dont les limites soient autant que faire se peut durablement repérables sur le terrain, (ii) dont la taille moyenne soit plus faible que les îlots du RP 2017 et (iii) de taille plus homogène. En pratique, le processus mis en œuvre pour obtenir les nouveaux îlots des grandes communes de Mayotte a été le suivant :

1. Découpage de l'ensemble du territoire en zones élémentaires selon des critères uniquement géographiques (pérennité des limites) : une zone élémentaire est la plus petite zone dont on pense que ses limites seront durablement repérables sur le terrain ;
2. Application d'un algorithme de regroupement de zones élémentaires en îlots de façon à s'approcher le plus possible d'une taille-cible ;
3. Validation des îlots produits et propositions d'ajustements à la marge.

Ces travaux ont considérablement profité de l'étroite collaboration entre statisticiens experts du territoire mahorais et méthodologues.

Remarque : le même processus a été mis en œuvre entre pour les îlots des petites communes. Les algorithmes utilisés pour le regroupement de zones élémentaires diffèrent (*cf. infra*) mais fournissent des résultats qualitativement proches.

3.1 Découpage du territoire en zones élémentaires

L'objectif de cette première étape du travail est d'obtenir une partition du territoire en zones telles que :

- leurs limites respectent les contours administratifs préexistants (communes, villages au sein des communes) et soient autant que faire se peut durablement repérables sur le terrain ;
- leur taille soit la plus petite possible, de façon à pouvoir être combinées efficacement par l'algorithme de constitution des îlots (*cf.* 3.2).

Lors de ces travaux menés en 2019, le service régional de Mayotte a pu mobiliser l'imagerie satellitaire ainsi que des photographies aériennes récentes, mais aussi sa connaissance du terrain afin d'évaluer la pérennité des limites. Ce redécoupage a permis d'aboutir à une partition en zones élémentaires de taille sensiblement plus faible que celle des îlots du RP 2017 (Tableau 12).

Statistiques (en nb de logements)	Îlots 2017 GC Mayotte	Zones 2019 GC Mayotte
Taille moyenne	105	25
Écart-type	63	63
Minimum	1	0
Premier décile	15	3
Premier quartile	63	8
Médiane	103	16
Troisième quartile	143	31
Neuvième décile	183	58
Maximum	513	413

TABLE 12 – Caractéristiques des zones élémentaires redécoupées

L'analyse de la distribution de la taille des zones élémentaires révèle néanmoins que celle-ci reste extrêmement dispersée, les considérations statistiques (îlots de taille homogène) ayant été explicitement exclues de cette première étape du travail. En conséquence, si ces zones étaient utilisées telles quelles dans le plan de sondage des EAR, cette dispersion serait de nature à nuire à la qualité des estimateurs obtenus (augmentation de la variance des estimateurs du RP en évolution). Par ailleurs, des îlots de taille trop faible compliquent l'organisation de la collecte (les agents recenseurs ont un carnet de tournée par îlot). Ces raisons ont poussé au regroupement de ces zones élémentaires en îlots de taille homogène par le biais d'un algorithme spécifique.

Rappel : C'est sur le découpage en zones élémentaires que ce sont appuyés les évaluations de plans de sondage présentés dans les parties précédentes : le regroupement en îlot n'était pas disponible au moment de l'étude sur les plans de sondage alternatifs. Ce point n'est néanmoins pas de nature à affecter qualitativement les résultats obtenus précédemment : le découpage n'a pas d'impact sur le biais et uniquement indirectement sur la variance (*cf. infra*, dernière sous-partie).

3.2 Regroupement algorithmique des zones élémentaires en îlots

Une fois la partition du territoire en zones élémentaires aux limites durablement repérables sur le terrain constituée, « il ne reste plus » qu'à regrouper ces zones élémentaires le plus efficacement possible pour obtenir *in fine* des îlots d'une taille la plus proche possible de la cible de nombre de logements.

Cette cible a été déterminée à l'issue d'un arbitrage entre limitation de l'effet de grappe (limiter les îlots de taille trop élevée) et facilité d'organisation de la collecte (limiter les îlots de taille trop faible), en comparaison avec la taille des îlots dans les autres DOM. Cette cible a ainsi finalement été fixée à 80 logements par îlot (du fait des impératifs de collecte : une tournée d'agent recenseur par îlot) : c'est sensiblement plus que la moyenne des îlots des grandes communes des autres DOM (de l'ordre de 60 logements par îlot), mais cela représente déjà un gain sensible par rapport aux îlots du RP 2017 (plus de 100 logements en moyenne). Par ailleurs, des travaux d'évaluation par simulation ont montré que, plus que de la réduction de la taille moyenne des îlots, c'est de la réduction de la dispersion de leur taille que provient l'essentiel du gain de variance par rapport à l'utilisation des îlots du RP 2017 (*cf.* également tableau 15).

D'autre part, outre la taille-cible plusieurs contraintes interviennent dans cette opération de regroupement :

- îlots formés au sein du même village : les communes de Mayotte ont été constituées en regroupant les villages préexistants. Il apparaît de ce fait souhaitable d'être en mesure de produire des résultats du recensement par village. Cela suppose qu'aucun îlot ne soit à cheval sur deux villages ;
- îlots formés exclusivement de zones contiguës, afin de faciliter le travail sur le terrain (tout particulièrement le repérage des limites d'îlot).

D'un point de vue algorithmique, la recherche du meilleur regroupement de zones élémentaires de sorte que leur taille s'approche le plus possible d'une taille cible est un problème difficile : le nombre de partitions possibles (nombre de Bell) augmente en effet considérablement avec la taille de l'ensemble considéré, pour devenir en pratique rédhibitoire à partir de 15 éléments (Tableau 13).

Nombre d'éléments d'un ensemble	Nombre de partitions possibles
1	1
2	2
3	5
4	15
5	52
10	115 975
15	1 382 958 545
20	51 724 158 235 372

TABLE 13 – Nombre de partitions possibles selon la taille de l'ensemble (nombre de Bell)

Les villages de Mayotte étant en général constitués de plusieurs dizaines de zones élémentaires, une recherche exhaustive de la meilleure partition apparaît en pratique impossible avec les moyens informatiques actuels¹³. Des algorithmes moins coûteux doivent donc être imaginés pour s'approcher de la partition optimale, par exemple :

- algorithme purement aléatoire : tirer au sort un très grand nombre de partitions respectant les contraintes (contiguïté notamment), calculer pour chaque la valeur de la fonction de perte (somme des écarts quadratique à la taille-cible) et sélectionner la meilleure partition selon ce critère.
 - Du fait du très grand nombre de partitions possibles, il est peu probable d'obtenir une partition de qualité en un temps raisonnable.
- algorithme quasi-déterministe par accréation successive : choisir aléatoirement une zone de départ puis lui agréger des zones contiguës jusqu'à s'approcher le plus possible de la taille-cible ; recommencer jusqu'à épuisement des zones à traiter. Cet algorithme est quasi-déterministe dans la mesure où, une fois la séquence des points de départ connus, le résultat de la partition est connu. Il pourrait gagner à être lancé un certain nombre de fois pour ne conserver que la meilleure partition.
 - Ce type d'algorithme fait courir le risque de conduire à des solutions structurellement sous-optimales : en cherchant à maximiser la proximité à la taille cible à chaque étape, certaines fusions de zones élémentaires judicieuses en termes d'optimalité globale peuvent ne jamais se produire.
- algorithme déterministe inspiré de la résolution du problème du voyageur de commerce : en assimilant chaque zone à son barycentre, chercher le plus court chemin pour parcourir l'ensemble des zones (et revenir au point de départ). A partir de ce circuit, il est assez simple de produire un regroupement en îlots de taille proche de la taille-cible : pour un point de départ et un sens donnés, regrouper les zones dans l'ordre du circuit jusqu'à s'approcher suffisamment (par exemple, à 10 % près) ou dépasser la taille-cible. Du fait que le circuit soit solution du problème du voyageur de commerce, les zones qui se suivent dans le circuit ont de grandes chances d'être contiguës. En faisant varier le point de départ et le sens, il est possible d'obtenir la meilleure partition pour un circuit donné. Cet algorithme est déterministe dès lors que l'algorithme de résolution du problème du voyageur du commerce l'est et que le mécanisme d'agrégation des zones à partir du circuit l'est également.
 - Cette méthode est éprouvée et a déjà été utilisée au sein de l'Insee (pour constituer les unités primaires du nouvel échantillon-maître Nautile notamment¹⁴) ; c'est par ailleurs cette méthode qu'a mise en œuvre pour constituer les îlots de Mayotte en petites communes. Elle présente néanmoins l'inconvénient de chercher à résoudre un problème formulé en deux dimensions (fusionner des zones contiguës) en l'« aplatissant » en une dimension (le circuit du voyageur de commerce) : ce faisant, il court le risque de ne pas atteindre certaines solutions intéressantes.

L'algorithme finalement mis en œuvre procède en deux étapes :

1. Lancement un très grand nombre de fois (5 millions pour chaque village) d'un algorithme de regroupement de zones élémentaires en partie aléatoire :
 - 1.1. Sélection aléatoire d'une zone parmi les zones non-achevées (*cf.* 1.3) ;
 - 1.2. Examen de toutes les paires de zones susceptibles d'être constituées avec cette zone (zones contiguës) ;

13. Ce type de problème pourrait en revanche constituer un cas d'application de l'informatique quantique actuellement en pleine croissance. D'ici quelques années, il pourrait devenir envisageable de rechercher exhaustivement des solutions à ce problème à l'aide de cette technologie (au prix d'un travail conséquent d'adaptation de l'implémentation).

14. http://jms-insee.fr/jms2018s09_1/

- 1.3. Pour chaque paire potentielle, comparaison de l'écart à la taille-cible avant et après fusion ;
→ Si aucune paire ne permet de se rapprocher de la taille-cible : arrêt et marquage de la zone comme achevée.
 - 1.4. Sélection aléatoire de la fusion à opérer proportionnellement à la proximité à la taille-cible ;
 - 1.5. Mise en œuvre de la fusion et retour à l'étape 1.1. L'algorithme s'arrête quand toutes les zones sont considérées comme achevées.
2. Parmi l'ensemble des 5 millions de simulations, sélection en trois étapes :
 - conservation uniquement des 10 000 partitions qui minimisent l'écart à la taille-cible ;
 - parmi ces 10 000 partitions, sélection des 200 partitions les plus compactes ;
 - parmi ces 200 partitions, sélection de celle qui minimise l'écart à la taille-cible.

Cet algorithme reprend ainsi la logique de l'algorithme quasi-déterministe par accrétion successive, avec deux variations :

- d'une part, davantage d'aléa est introduit à chaque étape (*via* la sélection aléatoire de la fusion à opérer) pour limiter le risque de sous-optimalité globale. Sous cet angle, cet algorithme apparaît comme un intermédiaire entre un algorithme purement déterministe (une seule solution dont l'optimalité est difficile à garantir) et un algorithme purement aléatoire (tirage complètement aléatoire de partitions dont la plupart sont très éloignées de la taille-cible) : il s'agit d'un algorithme garantissant un certain degré de proximité à la taille-cible tout en laissant une grande place à l'aléatoire ;
- d'autre part, le critère de validation n'est pas uniquement la somme des écarts quadratiques à la taille-cible : ce critère intervient certes dans un premier temps (pour ne retenir que les 10 000 partitions les plus proches de la taille-cible), mais il est complété lors de la seconde étape par un critère de compacité des îlots obtenus (plus les zones formant un îlot partagent de limites communes, plus cet îlot est considéré comme compact).

3.3 Validation des îlots

Dernière étape du processus, la partition obtenue par regroupement algorithmique a été analysée et validée par le service régional de Mayotte. Le principal objectif de cet exercice était de vérifier que les regroupements proposés ne conduisent pas à des aberrations sur le terrain.

En pratique, les demandes de modifications ont été peu nombreuses (quelques dizaines pour de l'ordre de 800 îlots) :

- pour un village, la partition initialement retenue a conduit à former un îlot en « couronne » autour du reste du village, le rendant en pratique très difficile à enquêter. À la demande du SR Mayotte, c'est la deuxième partition proposée par l'algorithme qui a finalement été retenue ;
- dans la plupart des autres cas, les modifications ont consisté en une permutation de zones élémentaires entre îlots, notamment quand le relief rend l'accès à certaines zones difficiles depuis le reste de leur îlot d'origine¹⁵ ;
- quelques redécoupages de zones élémentaires ont également été effectués *ex post* afin de mieux garantir la pérennité des limites d'îlot et le respect des limites de villages.

15. Il a également été envisagé de spécifier les frontières de zones réputées infranchissables en amont de l'algorithme de regroupement de zones afin que ces regroupements impossibles sur le terrain n'aient jamais lieu. Cette stratégie aurait cependant nécessité d'identifier *ex ante* toutes ces frontières infranchissables, ce qui aurait été extrêmement long et coûteux.

Statistiques (en nb de logements)	Îlots 2017 GC Mayotte	Zones 2019 GC Mayotte	Îlots 2021 GC Mayotte
Taille moyenne	105	25	87
Écart-type	63	63	32
Minimum	1	0	28
Premier décile	15	3	60
Premier quartile	63	8	70
Médiane	103	16	82
Troisième quartile	143	31	96
Neuvième décile	183	58	119
Maximum	513	413	413

TABLE 14 – Caractéristiques des îlots redécoupés pour l’EAR 2021 (et les suivantes)

Le redécoupage des îlots à Mayotte a permis d’atteindre les objectifs fixés initialement : les îlots obtenus sont plus petits que ceux du RP 2017 (87 logements en moyenne contre 105, 82 en médiane contre 103) et de taille sensiblement moins dispersée (écart-type de 32 contre 63). Leur taille moyenne ne correspond donc pas à la taille-cible, essentiellement en raison de très grandes zones élémentaires qui forment à elles seules un îlot et font sensiblement augmenter leur taille moyenne. Inversement, il n’y a désormais plus aucun îlot de très petite taille : le plus petit îlot compte 28 logements, 90 % des îlots comptent plus de 60 logements. Par ailleurs, du fait du travail approfondi du service régional de Mayotte, les limites de ces îlots ont vocation à être durablement repérables sur le terrain.

3.4 Impact du redécoupage d’îlot sur la précision des estimateurs

À côté des considérations pratiques de pérennité des limites d’îlots, la principale motivation du redécoupage des îlots du RP 2017 est méthodologique : en mobilisant des îlots en moyenne plus petits et de taille moins dispersée, le plan de sondage est de nature à produire des estimateurs plus précis. Cette intuition s’appuie sur des résultats classiques de la théorie des sondages sur l’intensité de l’effet de grappe en fonction de la taille moyenne et de la dispersion de la taille des grappes, établis dans le cas du sondage aléatoire simple.

Force est de constater cependant que le plan de sondage des EAR à Mayotte s’éloigne très sensiblement d’un sondage aléatoire simple par grappe : il s’agit d’un sondage à deux degrés avec tirage équilibré aux différents degrés, avec une strate exhaustive pour les habitations de fortune (cf. Partie 2). Par ailleurs, le recensement mobilise conjointement les cinq groupes de rotation d’adresses, ce qui vient *a priori* diminuer très sensiblement l’impact du premier degré de tirage (et donc de la méthode de constitution des groupes de rotation) sur les estimateurs en niveau du recensement.

Afin d’avoir une idée plus claire des gains réels associés au redécoupage des îlots du RP 2017 dans les grandes communes de Mayotte, une étude par simulation a été menée en reproduisant un grand nombre de fois (10 000) l’ensemble du processus du recensement (tirage des groupes de rotation d’adresses, tirage de l’échantillon dans le bâti en dur, récupération des données collectées dans le bâti en dur et l’habitat de fortune lors du RP 2017) en mobilisant trois partitions différentes du territoire : la partition en îlots du RP 2017, la partition en zones élémentaires redécoupées par le SR Mayotte en 2019, la partition finale en îlots utilisée pour l’EAR 2021 et les suivantes.

Pour chaque scénario sont présentés les propriétés des estimateurs de la population en logement ordinaire calculés à partir d'une à cinq EAR : cinq EAR correspondent au cycle RP, une seule EAR permet d'avoir une idée de la précision des estimations en évolution d'un RP au suivant (ces évolutions n'étant portées que par l'EAR entrante dans le cycle RP). Plus le nombre d'EAR diminue, plus on s'attend à ce que l'imprécision augmente : le choix d'une partition efficace (îlots de taille faible et homogène) doit cependant permettre de limiter cette augmentation.

L'indicateur présenté est dérivé de l'erreur quadratique moyenne (EQM, *mean square error* ou MSE en anglais) :

$$MSE^{(S)} = \frac{1}{B} \sum_{b=1}^B \left(\hat{Y}_b^{(S)} - Y \right)^2$$

où b désigne une simulation, B le nombre total de simulations (10 000), Y la variable d'intérêt et $\hat{Y}_b^{(S)}$ l'estimation pour la variable Y , la simulation b et le scénario S . Plus précisément sont comparées des erreurs quadratiques moyennes en racine relatives (*relative root mean square error* ou RRMSE) : étant donné un scénario de référence (ici, îlots du RP 2017 pour un cycle RP complet, soit 5 EAR), la RRMSE est le rapport de la racine carré des MSE du scénario comparé avec le scénario de référence. Autrement dit :

$$RRMSE^{(S)} = \frac{\sqrt{MSE^{(S)}}}{\sqrt{MSE^{(S_{ref})}}}$$

La RRMSE permet de facilement comparer des scénarios entre eux : une RRMSE supérieure à 1 indique une perte de précision par rapport au scénario de référence, une RRMSE inférieure à 1 indique un gain de précision par rapport au scénario de référence.

Ensemble des GC Min GC Max GC	1 EAR	2 EAR	3 EAR	4 EAR	RP (cinq EAR)
Îlots 2017 GC Mayotte	1,6 1,8 10,3	1,2 1,3 6,4	1,1 1,2 4,3	1,0 1,1 2,8	1,0 (ref) 1,0 1,0
Zones 2019 GC Mayotte	1,3 1,7 7,4	1,1 1,3 4,6	1,1 1,1 3,2	1,0 1,1 2,1	1,0 1,0 1,0
Îlots 2021 GC Mayotte	1,3 1,7 7,9	1,1 1,3 4,9	1,1 1,1 4,3	1,0 1,1 2,2	1,0 1,0 1,0

Lecture : Avec la partition en îlots du RP 2017, l'estimateur de la population de l'ensemble des GC de Mayotte est 1,6 fois moins précis avec une seule EAR qu'avec cinq EAR. Au niveau communal, ce rapport est compris entre 1,8 et 10,3 selon les communes. En revanche, quelle que soit la partition, la précision est rigoureusement identique sur cycle RP complet (RRMSE de 1,0).

TABLE 15 – Impact du redécoupage des îlots de Mayotte sur l'efficacité du plan de sondage (RRMSE, référence Îlots 2017 x RP)

Cette étude (Tableau 15) permet tout d'abord de confirmer l'évidence : quand les cinq groupes de rotation sont mobilisés conjointement, la manière de découper le territoire en îlots (qui servent à former les groupes de rotation) importe peu. Les RRMSE relatifs au scénario « Îlots 2017 GC Mayotte » sont de ce fait toujours égaux à 1,0 quand les 5 EAR sont mobilisées. En revanche, on constate que la partition retenue affecte les résultats quand moins d'EAR sont mobilisées. En particulier, l'imprécision semble 1,6 fois plus élevée quand une seule EAR est mobilisée avec les îlots de 2017 contre 1,3 seulement avec les deux autres partitions, les écarts pouvant être sensibles pour certaines communes. En pratique, cela signifie que les évolutions de population

d'un millésime du RP au suivant sont sensiblement mieux estimées avec la partition en zones élémentaires 2019 ou en îlots 2021 qu'avec les îlots du RP 2017.

Il faut par ailleurs noter la proximité entre les résultats obtenus avec les zones élémentaires 2019 et les îlots 2021 : alors même que les îlots 2021 sont sensiblement plus grands que les zones élémentaires (87 logements en moyenne contre 25), l'effet de grappe n'augmente pas, vraisemblablement du fait des gains en précision permis par l'homogénéisation de la taille des îlots (forte diminution du CV de la taille des îlots). Ce faisant, le regroupement des zones élémentaires en îlots de taille proche de la taille-cible permet de respecter une contrainte de collecte (îlots pas trop petits) sans détériorer la précision des estimations par rapport à la partition la plus fine. *In fine*, les gains de précisions permis par le redécoupage en zones élémentaires sont donc conservés dans la partition en îlot finalement utilisée pour le tirage de l'EAR 2021 et des suivantes.

Bilan

L'ensemble des travaux menés au cours de cette étude méthodologique préparatoire au lancement des EAR à Mayotte ont permis d'aboutir à une méthode à mettre en place dès 2021. Ces travaux ont été essentiels dans la préparation car ils ont premièrement mis en évidence le biais de couverture de la base de sondage à Mayotte et les risques associés à la méthodologie habituelle du recensement. À la suite de ce constat, nous avons pu définir différents scénarios de tirage adaptés aux spécificités du terrain mahorais. Ces scénarios ont été évalués à partir d'un grand nombre de simulations qui ont permis de dresser des estimateurs robustes et cohérents afin de prendre la décision sur la méthode à appliquer.

Cette opération a été rendue possible par trois choses :

- ▶ La collaboration étroite avec le Service Régional de Mayotte pour le redécoupage des îlots du RP 2017 de façon à faciliter la collecte sur le terrain et à améliorer la qualité des résultats.
- ▶ L'ensemble des échanges avec les acteurs impliqués lors des groupes de travail.
- ▶ La disponibilité d'une base de données exhaustive finement géolocalisée (RP 2017), ainsi que par la mise en oeuvre de méthodes d'évaluation innovantes, déjà expérimentées dans le cadre du projet Nautile.

Bibliographie

Deville, J.-C. & Tillé, Y. (2004). Efficient Balanced Sampling : The Cube Method. *Biometrika*, Vol 91, No 4, pp 893-912.

Le Recensement rénové : avancées méthodologiques et apports à la connaissance. *Économie et statistique*, n°483-485, 2016.

Godinot, A. (2005). Pour comprendre le recensement de la population. *Insee Méthodes*, Hors série.

Rousseau, S. & Tardieu, F. (2004). La macro SAS CUBE d'échantillonnage équilibré, Documentation de l'utilisateur. *Rapport technique*, Insee, Paris.

Ardilly, P. (2006). *Les techniques de sondages*. Éditions Technip, Paris.

Annexe A : Le tirage équilibré et la méthode du cube

Un échantillon est dit équilibré sur une ou plusieurs variables disponibles dans la base de sondage, lorsque pour chacune d'entre elles, l'estimateur Horvitz-Thompson du total coïncide exactement avec le vrai total issu de la base de sondage.

Un échantillon S d'une population U équilibré sur la variable de contrôle x respecte donc la contrainte suivante :

$$\sum_{i \in S} \frac{x_i}{\pi_i} = \sum_{i \in U} x_i \quad \text{soit} \quad \hat{t}_{x\pi} = t_x$$

où π_i est la probabilité d'inclusion de l'individu i dans l'échantillon S .

L'algorithme du Cube proposé par Deville et Tillé (2004)¹⁶ a un cadre général et permet la sélection d'échantillons équilibrés sur un nombre quelconque de variables, avec un jeu de probabilités d'inclusion $\boldsymbol{\pi}=(\pi_1, \dots, \pi_N)$ quelconque. Un échantillon s est vu comme un sommet $(s_1, \dots, s_N) \in \{0,1\}^N$ du N -cube $C=[0,1]^N$. L'algorithme consiste en une marche aléatoire pour passer du vecteur des probabilités d'inclusion $\boldsymbol{\pi}$ au vecteur des indicatrices de sélections \mathbf{I} en arrondissant aléatoirement les π_i à 0 ou 1.

On a :

$$\hat{t}_{x\pi} = \sum_{i \in U} \frac{x_i}{\pi_i} I_i = t_x = \sum_{i \in U} x_i = \sum_{i \in U} x_i \frac{\pi_i}{\pi_i}$$

$$\text{soit} \quad \sum_{i \in U} \frac{x_i}{\pi_i} (I_i - \pi_i) = 0 \quad \text{ou} \quad \mathbf{A}(\mathbf{I} - \boldsymbol{\pi}) = 0$$

avec :

$$\mathbf{A} = \left(\frac{x_1}{\pi_1}, \dots, \frac{x_N}{\pi_N} \right);$$

$\mathbf{I} = (I_1, \dots, I_N)^T$ le vecteur des indicatrices de sélections ;

et $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)^T$ le vecteur des probabilités d'inclusion.

On voit alors que I doit se situer dans l'espace des contraintes $\boldsymbol{\pi} + \text{Ker}(\mathbf{A})$ qui représente l'espace où les conditions d'équilibrage sont respectées.

On peut représenter facilement cette méthode dans un espace de dimension 3 pour une population de 3 unités : il s'agit d'un cube. En se plaçant dans le cas d'un sondage aléatoire simple sans remise de taille 2 et en affectant les même probabilités d'inclusion à chacune des unités ($\pi_i=2/3$), on peut remarquer que l'équilibrage est toujours exact en équilibrant sur la variable constante égale à 1. On sait alors qu'il existe 3 échantillons équilibrés composés de 2 unités distinctes : il s'agit ici des sommets $(0,1,1)$; $(1,1,0)$ et $(1,0,1)$.

16. La macro CUBE est disponible sur le site de l'Insee avec sa documentation ici : <https://www.insee.fr/fr/information/2021904>

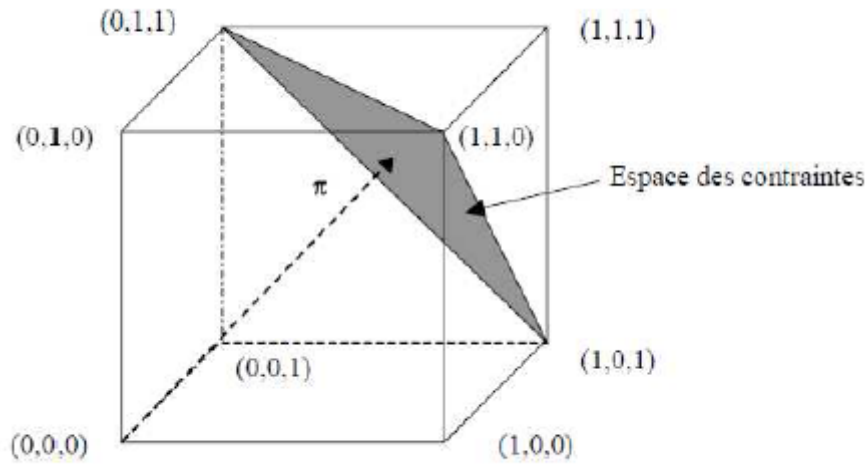


FIGURE 3 – Représentation graphique pour une population de 3 unités pour un sondage aléatoire simple sans remise équilibré sur la variable "1".

L'algorithme se déroule en deux phases : la phase de vol et la phase d'atterrissage.

- ▶ La phase de vol : marche aléatoire dans l'espace des contraintes où, à chaque étape, une composante du vecteur est changée en 1 ou 0 ;
- ▶ La phase d'atterrissage : elle permet de choisir un échantillon aussi équilibré que possible en relâchant les contraintes.

Enfin, en remarques, nous pouvons rappeler qu'il est nécessaire d'équilibrer sur la probabilité d'inclusion pour obtenir un échantillon de taille fixe et qu'il est astucieux de retirer une modalité de notre variable dans l'équilibrage car si celle-ci est équilibrée sur ces $(n-1)$ modalités, elle le sera également sur sa dernière par complémentarité¹⁷.

17. Une fiche méthodologique sur ce type de tirage est disponible sur le site de l'Insee ici : <https://www.insee.fr/fr/information/2838097>

Annexe B : Les différents plans de sondages envisagés

Pour la préparation des EAR à Mayotte, plusieurs scénarios de tirage ont été élaborés pour répondre au contexte mahorais. Tous ces scénarios sont des dérivés du plan de sondage habituellement employé dans les DOM. Ils ont été ensuite simulés à partir des jeux de données dont nous disposons et présentés en partie 1.4. Au final, 23 plans de sondage ont été testés avec à chaque fois plusieurs variantes sur les hypothèses de base de sondage ou sur les calculs d'estimateurs.

Nous listons ici l'esprit de ces différents scénarios :

- ▶ Le tirage est réalisé avec une base de sondage parfaite (sans défaut de couverture).
- ▶ Le tirage est mené intégralement au niveau des adresses, sans inclure dans le tirage les EA en travaux. Aucune tentative d'appariement entre sortants et entrants n'est menée.
- ▶ Le tirage est mené intégralement au niveau des adresses (y compris travaux). Aucune tentative d'appariement entre sortants et entrants n'est menée.
- ▶ Le tirage est mené intégralement au niveau des adresses (y compris travaux). Un appariement est mené entre adresses sortantes et adresses entrantes au sein d'un même îlot et en respectant la distinction habitations dures et habitations de fortune.
- ▶ Le tirage est mené intégralement au niveau des zones élémentaires.
- ▶ Le tirage est mené intégralement au niveau des zones élémentaires avec une dimension réjective pour garantir un certain taux de sondage des logements.
- ▶ Le tirage est mené intégralement au niveau des zones élémentaires à probabilité proportionnelle à leur taille dans le RIL.
- ▶ Le tirage est mené intégralement au niveau des zones élémentaires à probabilité proportionnelle à leur taille dans le RP 2017.
- ▶ Le tirage est mené au niveau des zones non-repérables puis au niveau des adresses dans les zones repérables avec ajustement des probabilités de tirage.
- ▶ Le tirage est mené intégralement au niveau des îlots avec une taille cible de 80.
- ▶ Protocole HMSA pour toutes les habitations de fortune.
- ▶ L'indicatrice d'habitat en dur de la base de sondage est corrigée de façon à correspondre à celle de l'enquête (*ie* on table sur une amélioration de la variable dans la base de sondage).
- ▶ Le tirage d'adresses est effectué dans toute la base de sondage (en dur ou pas). Sur le terrain, les adresses en dur tirées sont enquêtées avec le protocole standard. Les adresses non en dur, tirées ou non, sont toutes enquêtées avec le protocole HMSA.
- ▶ Protocole HMSA pour toutes les habitations de fortune avec indicatrice corrigée. Un appariement entre entrants et sortants est effectué par croisement de l'îlot et de la nature du bâti (dur / non-dur).
- ▶ Tirage classique pour les adresses en dur dans la base de sondage, dénombrement de la population des habitations en non-dur au moment de la collecte et tirage sur le terrain (via les documents de collecte) de 40 %.
- ▶ Tirage mixte.

Annexe C : Le détail des estimateurs obtenus dans les simulations sur les grandes communes de Mayotte

<i>Grandes communes de Mayotte</i> <i>Scénario 1</i> <i>Méthode DOM classique</i>	Population		Pourcentage d'étrangers		Nombre de logements		Pourcentage des habitations de fortune	
	Biais	CV	Biais	CV	Biais	CV	Biais	CV
Bandraboua	-9,0 %	1,7 %	-5,3 %	2,2 %	-9,5 %	1,1 %	-10,8 %	1,6 %
Bandrele	-8,6 %	1,9 %	-5,5 %	2,9 %	-8,4 %	1,1 %	-9,9 %	2,0 %
Dembeni	-11,6 %	1,6 %	-3,5 %	1,8 %	-12,0 %	1,2 %	-6,2 %	1,8 %
Dzaoudzi	-11,9 %	1,7 %	-2,4 %	2,0 %	-11,3 %	1,3 %	-7,6 %	2,1 %
Koungou	-12,6 %	1,4 %	-3,6 %	1,2 %	-12,2 %	0,9 %	-8,1 %	1,5 %
Mamoudzou	-6,5 %	0,9 %	-1,4 %	0,8 %	-6,5 %	0,8 %	-3,6 %	1,0 %
Ouangani	-2,6 %	2,1 %	-0,6 %	2,2 %	-2,5 %	1,6 %	-3,0 %	2,4 %
Pamandzi	-18,3 %	2,1 %	-10,6 %	3,1 %	-16,0 %	1,7 %	-16,9 %	3,2 %
Sada	-5,6 %	2,5 %	-5,6 %	4,2 %	-4,6 %	1,6 %	-16,6 %	4,2 %
Tsingoni	-7,8 %	1,9 %	-3,3 %	2,1 %	-7,6 %	1,5 %	-6,3 %	2,2 %
Ensemble des GC	-9,1 %	0,5 %	-2,9 %	0,5 %	-8,7 %	0,4 %	-6,9 %	0,6 %

Lecture : Avec le plan de sondage de la méthode DOM classique, on s'attend à un biais en population de -6,5 % du défaut de couverture pour Mamoudzou.

<i>Grandes communes de Mayotte</i> <i>Scénario 1 « optimiste »</i> <i>Méthode DOM classique</i>	Population		Pourcentage d'étrangers		Nombre de logements		Pourcentage des habitations de fortune	
	Biais	CV	Biais	CV	Biais	CV	Biais	CV
Bandraboua	-3,6 %	1,8 %	-2,6 %	2,1 %	-3,9 %	1,0 %	-5,3 %	1,5 %
Bandrele	-3,8 %	1,9 %	-3,6 %	2,8 %	-3,6 %	1,0 %	-5,9 %	1,8 %
Dembeni	-7,7 %	1,6 %	-2,4 %	1,7 %	-7,6 %	1,1 %	-5,4 %	1,7 %
Dzaoudzi	-5,7 %	1,7 %	-1,0 %	2,0 %	-5,2 %	1,2 %	-5,1 %	1,9 %
Koungou	-6,2 %	1,4 %	-2,2 %	1,1 %	-5,9 %	0,9 %	-5,1 %	1,4 %
Mamoudzou	-2,0 %	0,9 %	-0,6 %	0,8 %	-2,1 %	0,8 %	-1,7 %	1,0 %
Ouangani	-0,5 %	2,1 %	-0,3 %	2,2 %	-0,7 %	1,5 %	-1,0 %	2,4 %
Pamandzi	-7,0 %	2,2 %	-5,6 %	2,9 %	-5,6 %	1,7 %	-10,5 %	2,9 %
Sada	-2,1 %	4,0 %	-1,3 %	2,0 %	-1,7 %	1,6 %	-7,2 %	4,1 %
Tsingoni	-1,3 %	2,0 %	-0,9 %	2,0 %	-1,2 %	1,5 %	-1,4 %	2,1 %
Ensemble des GC	-3,8 %	0,5 %	-1,5 %	0,5 %	-3,6 %	0,4 %	-3,9 %	0,6 %

Lecture : Avec le plan de sondage de la méthode DOM classique sous hypothèse optimiste effectuée sur les « faux entrants », on s'attend à un biais en population de -2,0 % du défaut de couverture pour Mamoudzou.

TABLE 16 – Impact sur les estimateurs d'un plan de sondage classique dans les DOM (scénario 1) pour toutes les grandes communes de Mayotte.

<i>Grandes communes de Mayotte Scénario 2 Tirage d'îlots</i>	Population		Pourcentage d'étrangers		Nombre de logements		Pourcentage des habitations de fortune	
	Biais	CV	Biais	CV	Biais	CV	Biais	CV
Bandraboua	-0,5 %	2,7 %	-2,0 %	7,4 %	0,0 %	1,8 %	-3,8 %	12,2 %
Bandrele	0,1 %	2,7 %	-0,3 %	6,4 %	0,1 %	1,9 %	-0,3 %	10,3 %
Dembeni	-0,2 %	4,1 %	-0,5 %	4,4 %	0,0 %	2,1 %	-0,8 %	7,0 %
Dzaoudzi	-0,3 %	4,0 %	-0,1 %	3,5 %	-0,2 %	3,1 %	-0,7 %	3,9 %
Koungou	0,3 %	3,5 %	-0,1 %	3,3 %	0,1 %	1,9 %	0,2 %	7,5 %
Mamoudzou	0,0 %	2,2 %	-0,1 %	2,4 %	0,0 %	1,2 %	-0,2 %	5,3 %
Ouangani	-0,2 %	3,7 %	-0,4 %	5,5 %	0,0 %	2,4 %	-0,5 %	11,0 %
Pamandzi	-0,1 %	4,9 %	-0,4 %	6,4 %	-0,1 %	2,9 %	-0,2 %	9,9 %
Sada	0,1 %	4,3 %	-0,1 %	6,5 %	0,0 %	2,5 %	-0,2 %	10,5 %
Tsingoni	0,1 %	3,7 %	-0,7 %	4,9 %	0,3 %	2,8 %	-1,7 %	10,7 %
Ensemble des GC	0,0 %	1,2 %	-0,2 %	1,4 %	0,0 %	0,7 %	-0,5 %	2,8 %

Lecture : Avec le plan de sondage des îlots à recenser exhaustivement, on s'attend à un biais en population de 0,0 % du défaut de couverture pour Mamoudzou.

TABLE 17 – Impact sur les estimateurs du plan de sondage avec des îlots à recenser exhaustivement (scénario 2) pour toutes les grandes communes de Mayotte.

<i>Grandes communes de Mayotte Scénario 3 Recensement exhaustif des habitations de fortune</i>	Population		Pourcentage d'étrangers		Nombre de logements		Pourcentage des habitations de fortune	
	Biais	CV	Biais	CV	Biais	CV	Biais	CV
Bandraboua	-1,0 %	1,4 %	0,4 %	1,6 %	-1,3 %	0,9 %	1,3 %	1,0 %
Bandrele	-1,9 %	1,6 %	0,5 %	2,1 %	-2,0 %	1,0 %	1,9 %	1,1 %
Dembeni	-1,8 %	1,3 %	0,3 %	1,1 %	-3,1 %	1,1 %	2,8 %	1,1 %
Dzaoudzi	-3,6 %	1,3 %	1,4 %	1,4 %	-3,9 %	1,1 %	3,3 %	1,2 %
Koungou	-2,2 %	1,1 %	0,6 %	0,8 %	-2,9 %	0,8 %	2,8 %	0,9 %
Mamoudzou	-1,7 %	0,8 %	0,5 %	0,6 %	-2,2 %	0,7 %	2,1 %	0,7 %
Ouangani	-0,4 %	1,7 %	0,1 %	1,8 %	-0,4 %	1,4 %	0,4 %	1,3 %
Pamandzi	-5,3 %	1,8 %	0,8 %	2,1 %	-5,6 %	1,6 %	4,5 %	1,7 %
Sada	-1,3 %	2,4 %	0,1 %	3,5 %	-1,5 %	1,5 %	1,4 %	1,8 %
Tsingoni	-1,5 %	1,5 %	0,1 %	1,5 %	-2,2 %	1,3 %	2,0 %	1,3 %
Ensemble des GC	-2,0 %	0,4 %	0,6 %	0,4 %	-2,5 %	0,4 %	2,3 %	0,4 %

Lecture : Avec le plan de sondage d'un recensement exhaustif des habitations de fortune, on s'attend à un biais en population de -1,7 % du défaut de couverture pour Mamoudzou.

<i>Grandes communes de Mayotte Scénario 3 « optimiste » Recensement exhaustif des habitations de fortune</i>	Population		Pourcentage d'étrangers		Nombre de logements		Pourcentage des logements de fortune	
	Biais	CV	Biais	CV	Biais	CV	Biais	CV
Bandraboua	-0,1 %	1,4 %	0,0 %	1,6 %	-0,1 %	0,9 %	0,1 %	0,9 %
Bandrele	-0,3 %	1,6 %	-0,1 %	2,2 %	-0,2 %	1,0 %	0,2 %	1,1 %
Dembeni	-0,5 %	1,3 %	0,0 %	1,1 %	-1,2 %	1,1 %	1,1 %	1,1 %
Dzaoudzi	-0,9 %	1,4 %	0,6 %	1,4 %	-1,0 %	1,1 %	0,9 %	1,1 %
Koungou	-0,1 %	1,1 %	0,3 %	0,8 %	-0,7 %	0,8 %	0,7 %	0,9 %
Mamoudzou	0,2 %	0,7 %	0,1 %	0,6 %	-0,4 %	0,7 %	0,4 %	0,7 %
Ouangani	0,0 %	1,7 %	0,0 %	1,8 %	0,0 %	1,4 %	0,0 %	1,4 %
Pamandzi	0,0 %	1,8 %	0,0 %	2,0 %	0,0 %	1,6 %	0,0 %	1,6 %
Sada	-0,2 %	2,2 %	0,1 %	3,6 %	-0,3 %	1,5 %	0,3 %	1,8 %
Tsingoni	0,0 %	1,5 %	0,1 %	1,5 %	0,0 %	1,2 %	0,0 %	1,3 %
Ensemble des GC	-0,3 %	0,4 %	0,1 %	0,4 %	-0,5 %	0,4 %	0,4 %	0,4 %

Lecture : Avec le plan de sondage d'un recensement exhaustif des habitations de fortune sous hypothèse optimiste effectuée sur les « faux entrants », on s'attend à un biais en population de 0,2 % du défaut de couverture pour Mamoudzou.

TABLE 18 – Impact sur les estimateurs du plan de sondage avec un recensement exhaustif des habitations de fortune (scénario 3) pour toutes les grandes communes de Mayotte.

<i>Grandes communes de Mayotte Scénario 4 Tirage mixte</i>	Population		Pourcentage d'étrangers		Nombre de logements		Pourcentage des logements de fortune	
	Biais	CV	Biais	CV	Biais	CV	Biais	CV
Bandraboua	-1,7 %	3,4 %	-0,6 %	3,8 %	-1,6 %	2,7 %	0,8 %	3,5 %
Bandrele	-2,0 %	2,6 %	0,2 %	3,7 %	-1,9 %	1,7 %	1,9 %	2,7 %
Dembeni	-2,1 %	3,3 %	0,0 %	3,0 %	-3,3 %	2,3 %	2,7 %	2,2 %
Dzaoudzi	-3,7 %	3,4 %	1,4 %	2,5 %	-4,0 %	2,6 %	3,1 %	3,6 %
Koungou	-2,3 %	2,4 %	0,4 %	1,7 %	-3,0 %	1,7 %	2,8 %	1,8 %
Mamoudzou	-1,6 %	1,4 %	0,5 %	1,3 %	-2,2 %	1,0 %	2,1 %	1,1 %
Ouangani	-0,9 %	2,8 %	-0,4 %	2,9 %	-0,5 %	1,9 %	0,1 %	2,4 %
Pamandzi	-5,6 %	3,9 %	0,5 %	4,0 %	-5,7 %	2,9 %	4,3 %	4,8 %
Sada	-1,3 %	2,7 %	0,1 %	4,1 %	-1,5 %	1,8 %	1,5 %	5,1 %
Tsingoni	-2,0 %	2,9 %	-0,1 %	2,9 %	-2,4 %	2,2 %	1,6 %	2,9 %
Ensemble des GC	-2,2 %	0,9 %	0,4 %	0,8 %	-2,6 %	0,6 %	2,2 %	0,8 %

Lecture : Avec le plan de sondage d'un tirage mixte, on s'attend à un biais en population de -1,6 % du défaut de couverture pour Mamoudzou.

<i>Grandes communes de Mayotte Scénario 4 « optimiste » Tirage mixte</i>	Population		Pourcentage d'étrangers		Nombre de logements		Pourcentage des habitations de fortune	
	Biais	CV	Biais	CV	Biais	CV	Biais	CV
Bandraboua	-0,7 %	2,9 %	-1,0 %	3,4 %	-0,4 %	2,0 %	-0,4 %	2,7 %
Bandrele	-0,5 %	2,4 %	-0,8 %	3,6 %	-0,2 %	1,7 %	0,2 %	2,5 %
Dembeni	-0,8 %	3,0 %	-0,3 %	3,0 %	-1,4 %	2,2 %	0,9 %	2,1 %
Dzaoudzi	-1,1 %	3,4 %	0,5 %	2,5 %	-1,1 %	2,6 %	-0,7 %	3,5 %
Koungou	-0,4 %	2,4 %	0,1 %	1,7 %	-0,7 %	1,6 %	0,7 %	1,7 %
Mamoudzou	-0,2 %	1,4 %	0,0 %	1,3 %	-0,4 %	1,0 %	0,4 %	1,1 %
Ouangani	-0,4 %	2,7 %	-0,7 %	2,9 %	0,0 %	1,8 %	-0,1 %	2,4 %
Pamandzi	-0,2 %	3,7 %	-0,1 %	3,8 %	-0,1 %	2,7 %	0,0 %	4,1 %
Sada	-0,3 %	2,6 %	0,1 %	4,0 %	-0,3 %	1,7 %	0,5 %	5,0 %
Tsingoni	-0,1 %	2,7 %	-0,1 %	2,9 %	0,1 %	2,1 %	0,0 %	2,6 %
Ensemble des GC	-0,4 %	0,8 %	-0,1 %	0,8 %	-0,5 %	0,6 %	0,4 %	0,7 %

Lecture : Avec le plan de sondage d'un tirage mixte sous hypothèse optimiste effectuée sur les « faux entrants », on s'attend à un biais en population de -0,2 % du défaut de couverture pour Mamoudzou.

TABLE 19 – Impact sur les estimateurs du plan de sondage avec un tirage mixte (scénario 4) pour toutes les grandes communes de Mayotte.

