

Compter les absents : estimer indirectement une diaspora insulaire

Éva Lelièvre

Ined, France

eva.lelievre@ined.fr

Thomas Merly-Alpa

Insee, France

thomas.merly-alpa@insee.fr

Célio Sierra-Paycha

Cridup, Université Panthéon Sorbonne, France

celio.sierra-paycha@univ-paris1.fr

Clément Digoin

ENSAI, France

clementdigoin@gmail.com

L'objectif de cette étude est d'évaluer le degré de fiabilité de l'estimation d'une diaspora insulaire à partir de l'information obtenue sur le lieu de départ où résident des apparentés (parents, frères et sœurs et enfants). Les diasporas insulaires, du fait de leur très petite taille, sont souvent inidentifiables dans les statistiques des pays d'arrivée car comptées dans des agrégats qui ne permettent pas de les identifier. Or, les mobilités économiques ou de plus en plus souvent liées aux aléas climatiques poussent une partie des insulaires à quitter de façon temporaire ou plus définitive leur région de naissance. Cette étude est conduite à partir des données de la première enquête Famille de la Polynésie française (*Feti'i e fenua*, ISPF-Ined, 2020). Nous ne pouvons pas estimer directement la diaspora à partir des données collectées, car un même individu peut être cité par plusieurs membres de sa famille interrogés dans l'enquête *Feti'i e fenua*. Nous utilisons et adaptons la Méthode Généralisée de Partage des Poids (MGPP) proposée par Deville et Lavallée (2006) pour produire une série d'estimations de la diaspora polynésienne dans l'Hexagone. Nous confrontons les estimations obtenues via la MGPP aux données du recensement 2019 de l'Hexagone qui permettent de cerner la population née en Polynésie vivant en France métropolitaine. Nous discutons alors de la nature des biais observés.

Estimation indirecte, méthode généralisée du partage des poids.

Indirect estimation, Generalized Weight Sharing Method.

Les petits pays insulaires sont caractérisés par deux spécificités : leur population est peu nombreuse et leur diaspora, au sens de natifs résidant en dehors de l'archipel ou de l'île d'origine, relativement importante (Poirine, 1994 ; Bertram, 1999 ; Rallu, 2010). En raison de l'ampleur de l'émigration, les politiques publiques et la recherche gagneraient à disposer de données quantitatives fiables sur les dimensions et la structure de ces diasporas. Toutefois, la petite taille des populations d'origine et *a fortiori* des populations migrantes ne garantit pas une bonne visibilité statistique dans les pays d'immigration (Sierra-Paycha, 2022). Si l'information sur les pays de naissance est bien collectée dans la plupart des recensements ou registres dans les pays d'accueil (Walmsley, Ahmed, Parsons, 2007 ; Özden, Parsons, Schiff, Walmsley, 2011), la façon dont est construite cette donnée peut conduire à une invisibilisation statistique des diasporas insulaires. En effet, certains petits États insulaires sont en situation d'autonomie politique relative vis-à-vis de plus grands États comme la Polynésie française avec la France, Palau avec les États-Unis ou les îles Cook avec la Nouvelle-Zélande. Cela a pour effet de fondre les effectifs de personnes originaires de ces pays insulaires dans des agrégats nationaux de rang supérieur : natifs de France, des États-Unis ou de Nouvelle-Zélande pour reprendre l'exemple précédent. Ainsi, ni la Polynésie française, ni les îles Cook ne figurent dans la base bilatérale des migrations rassemblant pour chaque pays d'origine la distribution des ressortissants dans tous les pays du monde à partir des statistiques dans les pays de résidence (Parsons et al., 2007). Quant à Palau, la base indique moins de 500 natifs résidant à l'extérieur pour l'année 2020, ce qui paraît bien peu, même pour une population nationale d'à peine quelques dizaines de milliers d'habitants. En raison de la taille très faible des effectifs des diasporas insulaires, les instituts statistiques des pays d'accueil de ces diasporas ont également tendance à agréger, dans un échelon national supérieur, les catégories liminales avec la modalité principale pour des raisons d'anonymisation ou de facilité de codification. Cette invisibilisation, qui s'explique par les faibles effectifs et par l'association politique à un État plus puissant, est d'autant plus problématique dans les contextes insulaires, car l'émigration y fait très souvent partie intégrante d'un modèle de développement. Cette place prépondérante de l'émigration a permis de qualifier ce modèle par l'acronyme MIRAB, qui signifie *Migration, Remittances, Aids and Bureaucracy* (Bertram, Watters, 1984, 1985, 1986 ; Bertram, 1986, 1993, 2004).

Cependant, au-delà même du repérage des diasporas entre petits et grands États associés, la migration reste le phénomène le plus difficile à cerner à partir des grandes sources de données. En effet, le rattachement des individus à un lieu unique de résidence habituelle invisibilise, d'une part, le fait que les individus circulent et que le « temporaire » est susceptible de s'allonger et, d'autre part, que les lieux d'attache, pas forcément fréquentés mais peuplés de liens et d'obligations, constituent également l'espace de vie des personnes (Courgeau, 1988 [2021] ; Robette, 2012 ; Lelièvre, Robette, 2010). De plus, derrière la question renvoyant à qui compte, où et comment sont comptés les individus en fonction de leur(s) résidence(s), de leur(s) nationalité(s) et de leurs lieux d'attachements, se jouent des rapports des pouvoirs sous-jacents, d'autant plus dans le cas des Outre-mer. L'exceptionnalité qui caractérise une partie de leur gestion étatique (Beauvallet et al., 2016), elle-même héritière d'un ancien rapport colonial entre la « Métropole » et son empire, se décline également sur le plan de la statistique censitaire. En Polynésie française, l'histoire récente des catégories révèle que l'intégration par le recensement des « indigènes » au champ statistique de la communauté nationale remonte à la collecte de 1881 (Rallu, 1998). Au vingtième siècle, les Outre-mer ont suivi des trajectoires statistiques différentes selon qu'ils ont connu ou non un processus de départementalisation. Ainsi, dans les Départements et régions d'Outre-Mer (DROM), le bulletin individuel du recensement est devenu quasi identique

à celui de la métropole. Dans les Collectivités d’Outre-Mer (COM), même si la structure des questionnaires est assez similaire, on constate toujours des différences, par exemple en ce qui concerne les catégories ethniques en Nouvelle-Calédonie, ou le confiage d’enfants en Polynésie française (Sierra-Paycha et al., 2018). De même, on constate des variations du traitement statistique des ressortissants des Outre-mer vivant en France hexagonale. Les personnes issues des DROM déclarant un numéro de département de naissance, leur présence sur le territoire « métropolitain » est assimilée à une migration interdépartementale. L’intégration des DROM dans le cadre d’une statistique « départementale » a ainsi permis de documenter les migrations de la période du Bureau des Migrations des Outre-Mer (BUMIDOM) à l’aide des sources statistiques censitaires (Condon, Ogden, 1991 ; Haddad 2018 ; Condon 2020). Quant aux ressortissants des COM, ils déclarent leur lieu de naissance comme le déclareraient des natifs de l’étranger, ce qui rapproche l’appréhension statistique de leur présence en France de la façon dont l’Institut National de la Statistique et des Études Économiques (INSEE) aborde les migrations internationales.

Cet article est une expérimentation visant à tester la possibilité d’étudier les diasporas, soit les absents, à partir de données d’enquête auprès des membres de la famille présents dans le pays insulaire d’origine. Faire un sondage indirect pour compter les absents permet en effet d’estimer l’émigration d’une région dont l’économie dépend des transferts et des possibles retours. Cette question importante de l’estimation de la diaspora demande néanmoins à être validée. L’article cherche à évaluer dans quelle mesure une enquête qui collecte la localisation des absents apparentés auprès des enquêtés permet une estimation de l’ensemble des absents et de sa marge d’erreur. Pour cette démonstration, nous étudions le cas des natifs de Polynésie française résidant dans l’Hexagone, dont on dispose par ailleurs de données exceptionnellement fiables grâce au recensement mené par l’INSEE, qui collecte des informations détaillées sur les lieux de naissance des populations ultramarines. En effet, dans les fichiers de micro-données de l’Insee, on trouve des variables spécifiques qui permettent d’identifier tous les ressortissants ultramarins de chacun des territoires, et ce quels qu’en soient les effectifs, alors que les nationalités étrangères sont agrégées lorsque les effectifs sont trop faibles.

La première étape de cette étude consiste à évaluer la taille de la population des natifs de Polynésie résidant dans l’Hexagone à partir de la première enquête Famille (*Feti’i e fenua*, ISPF-Ined, 2020) collectée sur ce territoire ultramarin, en utilisant la Méthode Généralisée de Partage des Poids (MGPP) proposée par Lavallée (1995). Puis, en procédant à la comparaison des estimations obtenues avec les données censitaires de France métropolitaine (RP-2019) dénombant les personnes nées en Polynésie, nous détaillons les biais et adaptons la MGPP aux spécificités des données de *Feti’i e fenua*. L’analyse des pyramides des âges issues de la MGPP nous montre que la structure des estimateurs obtenus diverge des valeurs du recensement. Nous discuterons des écarts observés et des explications pouvant les produire.

Présentation des données utilisées

L'enquête *Feti'i e fenua* (ISPF-Ined, 2020) en Polynésie française

L'enquête *Feti'i e fenua* (ISPF-Ined, 2020), littéralement « liens et territoire », a été menée en Polynésie française en 2019 et 2020 par l'Institut National d'Études Démographiques (Ined) et par l'Institut de la Statistique de la Polynésie française (ISPF). Cette enquête s'inspire de l'enquête *Famille et logements* (2011) de l'Insee en France métropolitaine. Elle a permis de recueillir différentes informations concernant l'organisation familiale des Polynésiens, les relations entre les membres d'une même famille et leur mobilité au sein du territoire et au-delà (Fardeau, Lelièvre, 2021 ; Fardeau et al., 2021).

Elle recueille la composition et la dispersion spatiale des familles à travers les lieux de résidence de leurs membres, y compris en dehors du territoire, ainsi que leurs occupations¹ et les liens qu'ils entretiennent avec les lieux d'origine (lieu de naissance, île ou région de socialisation avant 6 ans, lieux de scolarité, etc.). La collecte s'est déroulée auprès d'individus âgés de 40 à 59 ans tirés au hasard. Ils constituent les « adultes pivots » entre des parents vieillissants et des enfants en âge de décohabiter. Le périmètre de la famille qui fait l'objet du recueil porte donc sur trois générations.

Le plan de sondage procède en plusieurs étapes. La première est bâtie sur une classification des îles, basée sur le nombre de ménages éligibles résidant sur l'île (un ménage est éligible s'il compte au moins un individu qui a atteint un âge compris entre 40 et 59 ans à la date de l'enquête) et sur la présence d'un aéroport pour les plus petites îles. Chaque île de Polynésie française appartient ainsi à une strate (tableau 1)² : parmi les 74 îles habitées³, nous en avons ainsi échantillonné 31. Ces îles couvrent l'ensemble des cinq archipels.

La deuxième étape consiste à tirer au hasard, dans l'échantillon des ménages issus du recensement de 2017, des ménages parmi lesquels se trouve au moins un individu âgé de 40 à 59 ans. Sur 5 964 ménages échantillonnés, l'information sur les familles de 5 139 répondants a été collectée, ce qui correspond à un taux de réponse très élevé de 86 %.

Enfin, la dernière étape du plan de sondage est le tirage au hasard d'un des individus dans la tranche d'âge d'intérêt au sein du ménage.

1 L'enquête collecte non seulement les professions (emploi et statut) mais également l'ensemble des activités des personnes afin de cerner précisément le travail informel et d'autosubsistance.

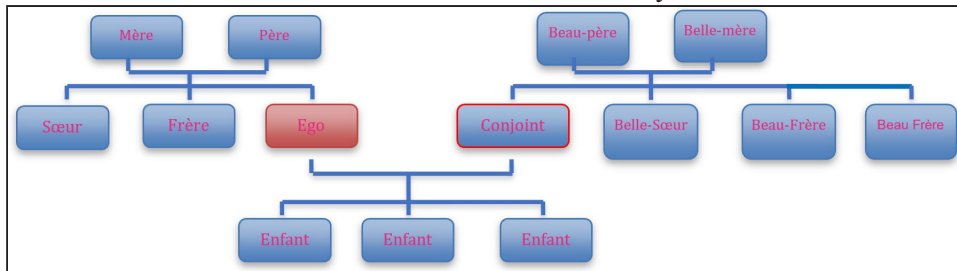
2 Pour les deux strates comptant les îles de moins de 200 ménages éligibles, nous avons opté pour un choix raisonné. Cependant, pour le calcul des poids de sondage, nous nous sommes replacés dans la situation d'un sondage aléatoire d'îles, en utilisant le nombre d'îles sélectionnées et le nombre d'îles dans la strate pour pondérer les îles sélectionnées.

3 Les îles comptant moins de 20 ménages éligibles ont été exclues de l'enquête.

Tableau 1. Strates du plan de sondage de l'enquête *Feti'i e fenua* (Ined-ISPF, 2000)

	Nombre d'îles	Nombre d'îles sélectionnées	Taux de sondage
Îles comptant plus de 500 ménages éligibles	7	7	100 %
Îles comptant entre 200 et 500 ménages éligibles	6	6	100 %
Îles comptant moins de 200 ménages éligibles et équipées d'un aéroport	16	7	43,75 %
Îles comptant moins de 200 ménages éligibles et non équipées d'un aéroport	44	11	25 %
Total	73	31	42,47 %

Le questionnaire est individuel, une seule personne (âgée de 40 à 59 ans) par ménage fournissant les informations concernant la situation de chacun sur trois générations : parents, frères et sœurs, enfants ainsi que, le cas échéant, famille du conjoint. On a ainsi le portrait spatialisé de ces familles qui résulte des migrations effectuées et les résume (univers familial présenté en figure 1).

 Figure 1. L'univers familial décrit par les répondants de l'enquête *Feti'i e fenua* (ISPF-Ined, 2020)


En raison du plan de sondage et de la non-réponse, il est nécessaire de calculer une pondération finale pour pouvoir extrapoler les résultats collectés à l'ensemble des individus de 40 à 59 ans en Polynésie Française. Pour cela, nous procédons en trois étapes. Tout d'abord, nous calculons les poids de sondage initiaux en tenant compte de la sélection des îles, des ménages puis des individus, comme exposé plus haut. À chaque étape, la probabilité d'être sélectionné est directement calculable, et le poids de sondage initial correspond à l'inverse du produit de ces trois probabilités. Nous envisageons ensuite le fait que certains individus n'ont pas répondu, ceci à travers une phase de correction de la non-réponse totale. Plusieurs méthodes sont possibles mais reposent sur le même principe : il s'agit de tenir compte du fait que le comportement de réponse peut varier selon différentes caractéristiques (âge, genre, type d'île) pour extrapoler les résultats. Enfin, la dernière étape est un calage sur marges (Deville, Sarndal, 1992) qui entraîne une légère altération des poids de sondages obtenus pour s'assurer qu'ils correspondent à des totaux connus (les « marges ») de la population polynésienne. À cette étape, on fait ainsi

correspondre des totaux de populations (comme le nombre d'hommes et de femmes de 40 à 59 ans) aux chiffres issus du recensement de la population polynésienne (RP, 2017).

Les données de recensement en France métropolitaine de 2019

On mobilise dans la suite de l'analyse, à des fins de comparaison, les données du recensement de la population en France métropolitaine de 2019 (Insee), qui résultent des enquêtes annuelles de recensement de 2017 à 2021. Nous utilisons ici le sous-échantillon des personnes nées en Polynésie française et résidant en France métropolitaine. On ne conserve donc que les individus qui vivent dans l'Hexagone et nés en Polynésie française. Par ce filtre, on obtient un échantillon de 5 530 individus vivant en France métropolitaine et qui ont déclaré être natifs de Polynésie française.

Ce chiffre ne correspond donc pas au nombre total de personnes vivant en France métropolitaine nées en Polynésie française. En effet, depuis 2004, la méthode de recensement en continu repose sur un échantillonnage rotatif de grande taille issu d'un plan de sondage complexe (Godinot, 2005). Ce n'est plus un recensement exhaustif comme cela était le cas jusqu'en 1999. Les estimations obtenues sont très précises, notamment en raison du fort taux de sondage (autour de 8 %), mais il est nécessaire d'utiliser une pondération afin de rendre l'échantillon représentatif de la population française. En effet, dans les plus grandes communes, la part de personnes recensées chaque année est inférieure à 8 % ; elle est supérieure dans les plus grands immeubles et les nouveaux logements. Quant aux petites communes, un cinquième d'entre elles sont recensées exhaustivement chaque année. Il est donc nécessaire de prendre en compte la façon dont les individus ont été sélectionnés à travers une pondération. Dans notre cas, la somme pondérée des 5 530 individus vivant en France métropolitaine et qui ont déclaré être natifs de Polynésie française conduit à une estimation de 17 278 individus nés en Polynésie française vivant en France métropolitaine. Cette estimation ainsi que sa répartition selon différentes caractéristiques (pyramides des âges, etc.) seront utilisées comme points de référence pour tester nos méthodes.

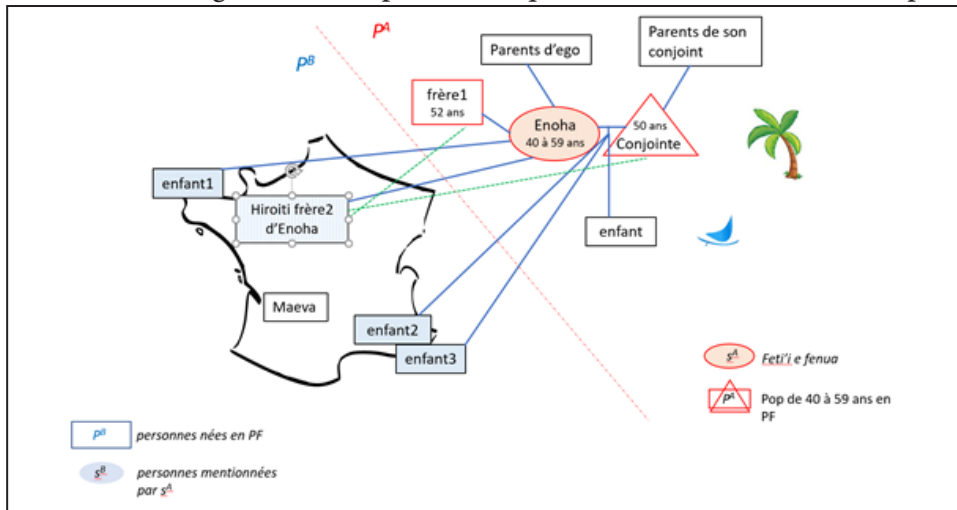
Le sondage indirect et la méthode généralisée de partage des poids

Pour évaluer la diaspora polynésienne, nous sommes dans le cadre d'un sondage dit « indirect ». En effet, on dispose d'un échantillon s^A de la population P^A des personnes âgées de 40 et 59 ans vivant en Polynésie française au moment de l'enquête, à partir duquel on cherche à estimer l'effectif de la population P^B , soit l'ensemble des individus ayant émigré de Polynésie française vers l'Hexagone.

Notons π_j^A la probabilité d'appartenir à l'échantillon s^A tiré dans la population P^A , appelée probabilité d'inclusion de l'individu j de P^A dans s^A . Les π_j^A sont connues, car nous avons conçu le plan de sondage de l'enquête *Feti'i e fenua* (voir Partie 1.1).

Les individus de la population P^B faisant partie de l'univers familial décrit par les individus de s^A constituent un échantillon noté s^B . Or, cette fois, les probabilités d'inclusion des individus i dans l'échantillon s^B de P^B sont généralement inconnues : il est difficile de calculer la probabilité qu'un individu ait été cité, d'autant plus quand les relations entre les individus de la population P^A et P^B ne sont pas bijectives (un individu de P^B n'est pas relié à un seul individu de P^A et inversement).

Figure 2. Le dispositif d'enquête et d'estimation de la diaspora



En effet, dans les données de *Feti'i e fenua*, un individu ayant entre 40 et 59 ans (population $s^A \subset P^A$) peut avoir plusieurs membres de sa famille qui ont émigré (population $s^B \subset P^B$). Inversement, un individu qui a émigré (population P^B) peut être décrit par plusieurs personnes de la population P^A pas nécessairement enquêtées (Figure 2). Ainsi Enoha répondant à l'enquête *Feti'i e fenua* (\subset à l'échantillon s^A) peut citer plusieurs membres de sa famille dans l'Hexagone (son frère en région parisienne, ses enfants à Brest, Lyon et Toulon appartenant à l'échantillon s^B). Notons également, nous y reviendrons, que Maeva est susceptible de n'être citée par personne et que le frère d'Enoha aurait potentiellement également été cité par deux personnes mentionnées par Enoha (son autre frère et sa conjointe), toutes deux ayant entre 40 et 59 ans.

Méthode naïve

La méthode naïve consiste à donner à chaque individu de s^B le poids de l'individu de s^A qui l'a cité. En d'autres termes, le lien qui lie les individus de P^A et de s^B est le suivant : un individu de s^B est lié à un individu de P^A si et seulement si l'individu i de s^B est cité par l'individu j de P^A dans *Feti'i e fenua*. La valeur de ce lien est de 1 si cela est vrai, 0 sinon. Le nombre total de liens L_i , pour un individu cité, est égal au nombre d'individus de s^A l'ayant cité comme faisant partie de son univers familial.

Dans le cas particulier où l'individu i n'est cité que par un seul individu j , son poids peut s'écrire comme l'inverse de la probabilité d'inclusion de l'individu j de P^A dans s^A , soit : $w_i = 1 / \pi_j^A$

Dans la famille d'Enoha (Figure 3), seul Enoha a été interrogé et a cité Hiroiti, donc Hiroiti a un seul lien au total ($L_i = 1$). Si Hiroiti n'est mentionné par aucun autre individu de s^A , son poids vaut donc $w = 1 / (\pi)$ avec π la probabilité d'inclusion d'Enoha dans l'enquête *Feti'i e fenua*.

Si cette première méthode naïve est plutôt simple, elle ne donne pas de résultats satisfaisants.

En effet, on obtient après pondération 27 729 individus nés en Polynésie et vivant en France métropolitaine, soit un écart de 9 951 avec les effectifs du recensement (17 278 personnes).

Cette méthode conduit à surestimer grandement le nombre d'individus vivant en France métropolitaine. Cette surestimation s'explique par la définition des liens que nous avons donnée : ne prendre en compte que les individus de s^A pour compter le nombre de liens conduit à augmenter artificiellement le poids de chaque individu de s^B .

Une autre façon de le voir est de dire que cette hypothèse conduit à une forte « non-réponse de liens » (Xu, Lavallée, 2009), ce qui implique un biais de surestimation important. Nous pouvons également le voir comme une mauvaise standardisation de la matrice de liens liée à la méconnaissance du nombre total de liens (Medous et al., 2023). Pour améliorer nos résultats, il faut raffiner la prise en compte des liens familiaux par la méthode généralisée du partage des poids.

Méthode généralisée du partage des poids

La présence de ces liens non bijectifs nous conduit à utiliser la Méthode Généralisée de Partage des Poids (MGPP) présentée par J.C. Deville et P. Lavallée en 2006 (Deville, Lavallée, 2006). Comme son nom l'indique, cette méthode repose sur l'idée que nous allons répartir le poids des individus de s^A parmi les individus de s^B de telle sorte que nous puissions obtenir une estimation sans biais. Un cas d'application simple est celui des relations « parents-enfants », pour lesquelles chaque enfant possède au plus deux parents et donc deux liens. Les enfants n'ayant qu'un seul parent dans la population initiale ont deux fois moins de chance d'être sélectionnés : il est nécessaire de prendre cela en compte, ce qui se fait en divisant par deux les poids des parents des enfants ayant leurs deux parents dans la population initiale.

De façon mathématique, les poids w_i obtenus par la MGPP sont calculés ainsi :

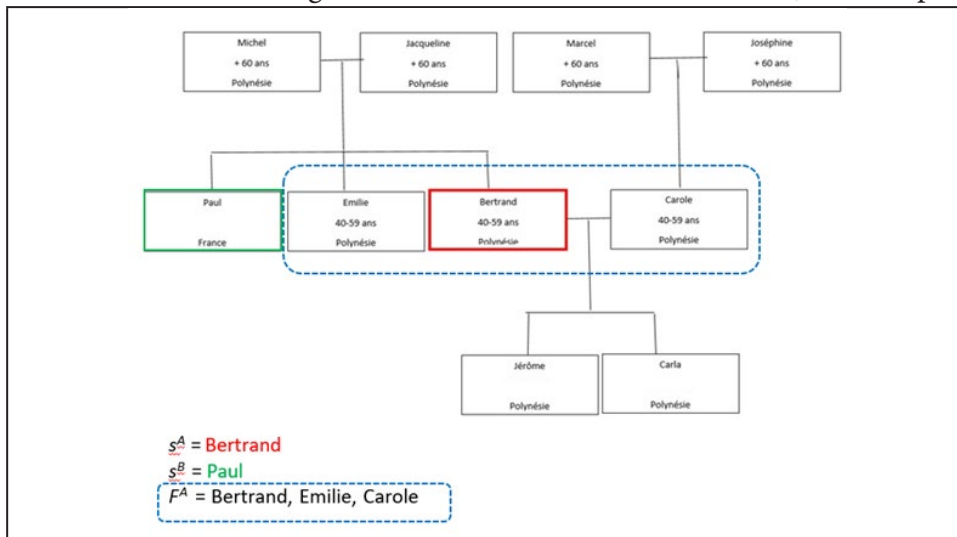
$$w_i = \sum_{j=1}^{n^A} \frac{L_{ij}}{\pi_j^A \times L_i}$$

où L_{ij} est l'indicatrice de la présence d'un lien entre l'individu i de P^A et j de P^B , qui vaut 1 si l'individu j de s^A est lié à l'individu i de s^B , 0 sinon, et L_i est le nombre total de liens pour un individu i de s^B avec des individus de P^A .

Méthode inclusive tenant compte de l'univers familial

Pour implémenter la MGPP dans notre cas, il faut définir ce que nous entendons par « lien » entre les individus de P^A et s^B . En effet, ce choix conditionne le mode de calcul des indicatrices L_{ij} et du nombre total de liens L_i . Pour illustrer cela, prenons un exemple de famille pour un individu enquêté nommé Bertrand. Supposons que tous les membres de la famille de Bertrand sont nés en Polynésie française. En suivant les notations que nous avons définies, nous obtenons la Figure 3.

Figure 3. L'univers familial de Bertrand (cas-exemple)



Compte tenu de la façon dont l'univers familial est collecté dans l'enquête (Figure 1), nous ne connaissons pas l'ensemble des liens reliant les individus de s^B aux individus de P^A , mais il est possible de les déduire. En effet, il est possible qu'un individu vivant dans l'Hexagone présent dans s^B soit relié à d'autres individus de P^A qui appartiennent à l'univers familial des répondants de s^A . Par exemple, dans l'exemple détaillé en Figure 2, Paul pourrait être cité également par Emilie (sa sœur) et par Carole (la femme de son frère) vivant en Polynésie Française et dans la bonne tranche d'âge, qui auraient pu être enquêtées et auraient pu le citer. Les informations dont nous disposons sont donc le lien familial entre les individus de l'ensemble des univers familiaux des répondants de s^A , et de s^B ; nous noterons cet ensemble F^A . Dans l'exemple présenté en Figure 2, $F^A = \text{Enoha, sa conjointe et son premier frère}$, et on a bien la relation :

$$s^A \subset F^A \subset P^A$$

La MGPP requiert de définir les « liens » entre les individus de P^A et de s^B . Pour les approximer, on compte le nombre de liens pour chaque individu de s^B en utilisant les liens que l'on connaît entre les personnes constituant F^A et les personnes décrites de s^B soit l'ensemble des liens familiaux entre les personnes de 40 à 59 ans vivant en Polynésie française et ayant été enquêtées ou mentionnées dans l'enquête *Feti'i e fenua*.

En résumé, nous considérons les ensembles suivants :

P^A : ensemble des individus ayant entre 40 et 59 ans et vivant en Polynésie française ;

P^B : ensemble des individus nés en Polynésie française et vivant en France métropolitaine ;

s^A : individus enquêtés de *Feti'i e fenua* ;

s^B : individus mentionnés par des individus de s^A nés en Polynésie française et vivant en France métropolitaine ;

F^A : ensemble des individus ayant entre 40 et 59 ans, vivant en Polynésie française et ayant été enquêtés ou mentionnés dans l'enquête *Feti'i e fenua*. Ils vérifient $s^A \subset F^A \subset P^A$.

Puisque la première méthode conduisait à une surestimation importante, il convient donc d'affiner l'identification des liens entre les apparentés.

La définition du lien entre les individus ne change pas ; mais nous allons considérer maintenant l'ensemble des individus de F^A pour calculer le nombre total de liens L_i . Pour cela, nous évaluons pour chaque individu de F^A et de s^B la présence d'un lien de la façon suivante : un individu de s^B est lié à un individu de F^A si et seulement si :

(1) l'individu j de F^A qui a participé à l'enquête *Feti'i e fenua* a mentionné l'individu i de s^B . La valeur de ce lien sera de 1 si cela est vrai, 0 sinon.

En se reportant à l'univers d'Enoha (Figure 2), trois personnes de 40 à 59 ans sont liées à Hiroiti, une enquêtée et deux éligibles à l'enquête. Dans la famille de Bertrand (Figure 3), trois personnes ont entre 40 et 59 ans (Bertrand, Émilie et Carole) et vivent en Polynésie française : elles sont donc éligibles.

Si Émilie était interrogée, elle citerait Paul car c'est son frère. Si Carole était interrogée, elle citerait Paul car c'est son beau-frère, donc Paul a trois liens au total ($L_i = 3$). Si Paul n'est mentionné par aucun autre individu de s^A , d'après la formule (1) son poids vaut $w = 1 / (3\pi)$ avec π la probabilité d'inclusion de Bertrand dans l'enquête *Feti'i e fenua*.

Prise en compte des fratries recomposées

La prise en compte de ces liens permet de réduire la non-réponse de liens. Il reste cependant à gérer quelques cas spécifiques. Par exemple, dans les cas où figurent aussi dans l'univers familial de la personne interrogée (respectivement de son conjoint) des demi-frères et des demi-sœurs, il est difficile de savoir avec assurance si ces demi-frères et demi-sœurs de la personne interrogée pourraient citer la personne appartenant à s^B . Une possibilité serait de les exclure, mais, dans ce cas, nous réintroduirions du biais de surestimation. Il est donc décidé d'allouer un lien de force plus faible à ces individus, conformément à la proposition de J.C. Deville et P. Lavallée (2006) : entre $\frac{1}{2}$ (pour les cas où l'individu dans s^B est un parent de la personne interrogée) et $\frac{2}{3}$ (pour les cas où l'individu dans s^B fait partie de la fratrie de la personne interrogée). Le choix de ces coefficients repose sur l'énumération des cas possibles. Supposons un enquêté ayant un demi-frère faisant partie de F^A . Pour la mère de l'individu enquêté, deux situations sont possibles : elle est soit également la mère du demi-frère, soit elle ne l'est pas. La demi-sœur de l'individu enquêté, quant à elle, peut partager les deux mêmes parents avec le demi-frère, un seul des deux ou aucun, d'où un choix de pondération du lien plus fort. Ces pondérations sont prises en compte dans le calcul du lien L_{ij} et du nombre total de liens L_i .

Dans la famille de Bertrand (Figure 3), supposons maintenant qu'Émilie soit la demi-sœur de Bertrand. Dans ce cas, Paul a trois liens au total, mais l'un d'eux ne compte que pour $\frac{2}{3}$ ($L_i = 2 + \frac{2}{3}$). Si Paul n'est mentionné par aucun autre individu de s^A , son poids vaut donc $w = 1 / ((2 + \frac{2}{3}) \pi)$ avec π la probabilité d'inclusion de Bertrand dans l'enquête *Feti'i e fenua*.

On utilise cette méthode dans la suite de cet article.

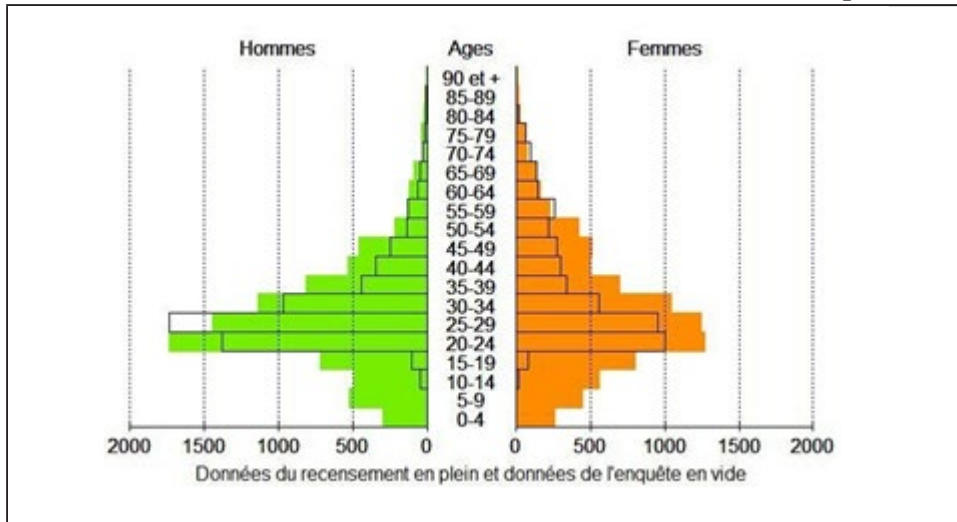
Comparer les pyramides des âges (Estimation / RP) et identifier les biais

En appliquant la MGPP, on obtient après pondération un effectif de 10 240 Polynésiens vivant en France métropolitaine. Il manque donc 7 038 individus par rapport au recensement⁴. En tenant compte du sexe et de l'âge de chaque individu, on peut construire les pyramides des âges des personnes nées en Polynésie française et vivant en France à partir, d'une part, des absents pondérés en s'appuyant sur les données de *Feti'i e fenua* et, d'autre part, du

4 Rappelons que l'effectif après pondération est de 17 278 personnes.

recensement de la France métropolitaine. La Figure 4 montre la superposition de la pyramide des âges issue du recensement (en couleur) et de la diaspora estimée par la MGPP (encadrée) en France métropolitaine. Des tranches d'âge de cinq ans sont utilisées pour pouvoir analyser en particulier la population jeune. Pour les âges les plus élevés, l'incertitude des estimations ne permet pas une analyse très fine.

Figure 4. Pyramide des âges des personnes nées en Polynésie française et vivant en France (diaspora estimée comparée au recensement en France métropolitaine)



Cette population est globalement jeune et se concentre autour des 20-35 ans, puis décroît aux âges plus élevés. La prépondérance de cette tranche d'âge n'est pas étonnante, car un certain nombre de jeunes Polynésiens quittent la Polynésie et rejoignent l'Hexagone dans le cadre de l'armée ou pour y poursuivre leurs études (Fardeau et al., 2021). La MGPP permet d'obtenir une structure proche de celle du recensement au vu de la figure 4, mais de nombreux écarts existent.

La différence principale est l'absence des moins de 20 ans, qui ne sont quasiment pas cités dans l'enquête. Cela résulte du sondage particulier de *Feti'i e fenua*, constitué d'un échantillon représentatif des 40-59 ans : d'une part, les jeunes enfants des enquêtés vivent encore avec eux et, d'autre part, les jeunes enfants nés en Polynésie et présents dans l'Hexagone sont souvent partis avec leurs parents (qui ne peuvent donc pas avoir été enquêtés en Polynésie). Il convient donc de supprimer ces tranches d'âges car, du fait du plan de sondage de *Feti'i e fenua*, l'estimation par proxy est largement déficiente les concernant.

Cette suppression des moins de 18 ans de l'étude conduit à l'estimation par la méthode MGPP d'une diaspora de 10 151 individus (correspondant aux 13 958 personnes de plus de 18 ans recensées dans l'Hexagone). En ne prenant plus en compte les mineurs, on réduit donc l'écart entre la diaspora estimée et les données du recensement de 7 000 à 3 700, une estimation améliorée.

Les biais subsistants sont les suivants. Tout d'abord, indifféremment pour les deux sexes, les personnes n'ayant pas de membres de leur famille éligible à l'enquête *Feti'i e fenua* sont absentes de l'estimation de la diaspora. Ensuite, la population masculine des 25-29 ans est surestimée (+286) par l'estimation indirecte de la diaspora : cela peut provenir d'une sous-estimation des militaires par le recensement dans l'Hexagone, ceux-ci résidant en caserne ou étant partis en mission à l'étranger. Enfin, les femmes sont sous-estimées par la MGPP, notamment les 25-40 ans, bien que l'on n'observe aucune sur-citation

des jeunes hommes par les enquêté-es de *Feti'i e fenua*. On peut seulement supposer que ce biais résulte d'une moindre affiliation de ces jeunes femmes à des personnes éligibles de l'enquête.

L'écart entre les populations étant une sous-évaluation du nombre de personnes par la MGPP, cela nous rassure sur les risques liés à la non-réponse de liens, qui conduirait systématiquement à une surestimation. Ces écarts sont donc potentiellement liés, d'une part, à une incertitude probabiliste, qui pourrait être estimée à partir du plan de sondage de *Feti'i e fenua* et, d'autre part, à la présence d'individus dans l'Hexagone n'ayant plus de liens avec la Polynésie.

Conclusion

En conclusion, la méthode d'estimation que nous avons mise en place pour évaluer la diaspora, à partir de la MGPP adaptée à notre situation, est plutôt efficace sur l'exemple de la Polynésie française. En effet, la pondération par la MGPP des individus cités dans l'enquête permet d'approximer de façon satisfaisante le nombre de Polynésiens vivant en France métropolitaine.

L'écart constaté avec le recensement, de 7 000 individus en moins pour les absents estimés par les données de l'enquête, est certes relativement important (pour une population totale respective d'environ 17 200) mais imputable en grande partie à la construction de l'enquête. Si le fait de n'avoir interrogé que les 40-59 ans, ces adultes « pivots », permet de capter une grande partie de la population émigrante, les absents des classes d'âges extrêmes sont difficilement atteignables, notamment les plus jeunes. En ne prenant plus en compte les mineurs, notre estimation est grandement améliorée (on passe d'un écart de 7 000 à 3 700 par rapport aux données du recensement).

Le reste de l'écart est explicable en très grande partie par le phénomène que l'on étudie : la migration. En effet, la MGPP ne permet que de retrouver les personnes ayant encore des parents proches (et dans la tranche d'âge interrogée) sur le territoire d'origine. Si certains émigrants n'ont plus aucun parent proche (du moins dans la tranche d'âge interrogée) dans le lieu de départ, la MGPP ne permet pas de les prendre en compte. Ce phénomène explique donc en partie l'écart avec le recensement.

Le caractère encore expérimental de cette méthode d'estimation est une invitation à poursuivre d'autres recherches méthodologiques visant à affiner la connaissance statistique des diasporas insulaires à partir de données indirectes. Pour mieux connaître leur diaspora, d'autres instituts de statistiques nationaux de petits pays insulaires pourraient reproduire cette méthode d'estimation indirecte des caractéristiques de l'émigration. Nous avons utilisé ici une enquête dont le plan de sondage particulier impacte grandement l'estimation, mais si, par exemple, les questions concernant la localisation des membres de la famille étaient incluses dans les recensements des petits États insulaires, seuls les natifs n'ayant plus de famille résidente dans leur région de naissance seraient invisibilisés.

Une autre piste de travail serait d'utiliser les résultats sur la diaspora pour améliorer la précision des données collectées sur le territoire d'origine. En suivant l'approche de Deville (1998) et de Caron et Sautory (2004), nous pourrions envisager de réaliser un calage simultané sur les poids des répondants de *Feti'i e fenua* en prenant des marges sur les habitants de Polynésie française et sur les Polynésiens vivant en France métropolitaine, via les recensements respectifs de ces deux populations (Digoin et al., 2023). Pour reprendre les concepts théoriques, les unités primaires seraient les répondants de *Feti'i e fenua* et les unités secondaires les personnes que ces répondants citent comme vivant en France métropolitaine. Ce calage simultané permettrait

aux répondants de *Feti'i e fenua* de représenter correctement la structure de la population en Polynésie française, mais aussi celle des Polynésiens vivant en France métropolitaine.

Bibliographie

- BEAUVALLLET Willy, CÉLESTINE Audrey, ROGER Aurélie, 2016, « L'État outre-mer », *Politix*, vol. 116, n° 4, p. 139-161.
- BERTRAM Geoff, 1986, « Sustainable development in Pacific micro-economies », *World Development*, vol. 14, n°7, p. 809-822.
- BERTRAM Geoff, 1993, « Sustainability, aid, and material welfare in small South Pacific Island economies, 1900–1990 », *World Development*, vol. 21, n° 2, p. 247-258.
- BERTRAM Geoff, 1999, « The MIRAB model twelve years on », *The Contemporary Pacific*, vol. 1, n° 1, p. 105-138.
- BERTRAM Geoff, 2004, « On the convergence of small island economies with their metropolitan patrons », *World development*, vol. 32, n° 2, p. 343-364.
- BERTRAM Geoff, WATTERS Ray, 1984, « New Zealand and its small island neighbours: A review of New Zealand policy toward the Cook Islands, Niue, Tokelau, Kiribati and Tuvalu », *Institute of public policies working paper*, n° 1.
- BERTRAM Geoff, WATTERS Ray, 1985, « The MIRAB economy in South Pacific microstates », *Pacific Viewpoint*, vol. 26, n° 3, p. 497-519.
- BERTRAM Geoff, WATTERS Ray, 1986, « The MIRAB process: Earlier analyses in context », *Pacific Viewpoint*, vol. 27, n° 1, p. 47-59.
- CARON Nathalie, SAUTORY Olivier, 2004, « Calages simultanés pour différentes unités d'une même enquête », *Document de travail Méthodologie statistique INSEE*, n°0403.
- CASSIN Lesly, MELINDI-GHIDI Paolo, PRIEUR Fabien, 2022, « Confronting climate change: Adaptation vs. migration in Small Island Developing States », *Resource and Energy Economics*, vol. 69, article n° 101301.
- CONDON Stéphanie A., 2020, « Continuité coloniale et gestion démographique des Antilles françaises, 1950-1980 », *Migrations Société*, vol. 182, n° 4, p. 43-57, <https://doi.org/10.3917/migra.182.0043>.
- CONDON Stéphanie A., OGDEN Philip E., 1991, « Emigration from the French Caribbean : the Origins of an Organized Migration », *International Journal of Urban and Regional Research*, vol. 15, n° 4, p. 505-523, <https://doi.org/10.1111/j.1468-2427.1991.tb00655.x>.
- COURGEAU Daniel, 1988 (réed. 2021), *Méthodes de mesure de la mobilité spatiale*, Paris, Éditions de l'Ined, coll. « Méthodes et Savoirs », n° 11, 368 p.
- DEVILLE Jean-Claude, 1998, « Les enquêtes par panel : en quoi diffèrent-elles des autres enquêtes ? », suivi de « Comment attraper une population en se servant d'une autre », *Actes des Journées de Méthodologie Statistique*, Paris, Insee, coll. « Insee Méthodes », p. 84-86, 63-82.
- DEVILLE Jean-Claude, SARNDAL Carl-Erik, 1992, « Calibration Estimators in Survey Sampling », *Journal of the American Statistical Association*, n° 87, p. 376-382.
- DEVILLE Jean-Claude, LAVALLÉE Pierre, 2006, « Sondage indirect : les fondements de la méthode généralisée du partage des poids », *Techniques d'enquête*, vol. 32, n° 2, p. 185-196.
- DIGOIN Clément, LELIÈVRE Éva, MERLY-ALPA Thomas, SIERRA-PAYCHA Celio, 2023, « Compter les absents : estimation de la fiabilité d'une observation indirecte », Communication au Colloque *Sondages*, Société française de statistique, Aubervilliers, 24 mars 2023.

- FARDEAU Leïla, LELIÈVRE Éva, ÉQUIPE ATOLLS, 2021, « L'enquête Feti'i e fenua (Enquête Famille, territoire et relations intergénérationnelles en Polynésie française) : Apurement et imputation des données », Éditions de l'Ined, coll. « Documents de travail », n° 262.
- FARDEAU Leïla, LELIÈVRE Éva, SIERRA-PAYCHA Celio, 2021, « La première enquête Famille en Polynésie française : Feti'i e fenua », *Points études et bilans de la Polynésie française*, n° 1276, p. 1-4.
- GODINOT Alain, 2005, « Pour comprendre le recensement de la population », *Insee Méthodes*, <https://www.insee.fr/fr/information/2579979>.
- HADDAD Marine, 2018, « L'effet d'une politique d'État sur les migrations DOM-métropole. Les enseignements des recensements de 1962 à 1999 », *Population*, vol. 73, n° 2, p. 191-224.
- LAVALLÉE Pierre, 1995, « Pondération transversale des enquêtes longitudinales menées auprès des individus et des ménages à l'aide de la méthode du partage des poids », *Techniques d'enquête*, vol. 21, n° 1, p. 27-35.
- LELIÈVRE Éva, ROBETTE Nicolas, 2010, « A Life Space Perspective to Approach Individual Demographic Processes », *Canadian Studies in Population*, vol. 37, n° 1-2, p. 207-244.
- MEDOUS Estelle, GOGA Camelia, RUIZ-GAZEN Anne, BEAUMONT Jean-François, DESSERTAINE Alain, PUECH Pauline, 2023, « Many-to-one indirect sampling with application to the french postal traffic estimation », *The Annals of Applied Statistics*, vol. 17, no 1, p. 838-859.
- ÖZDEN Çağlar, PARSONS Christopher R., Schiff Maurice, WALMSLEY Terrie L., 2011, « Where on earth is everybody? The evolution of global bilateral migration 1960-2000 », *The World Bank Economic Review*, vol. 25, n° 1, p. 12-56.
- PARSONS Christopher R., SKELDON Ronald, WALMSLEY Terrie L., WINTERS L. Alan, 2007, « Quantifying international migration: A database of bilateral migrant stocks », *World Bank Policy Research Working Paper*, n° 4165.
- POIRINE Bernard, 1994, « Rent, emigration and unemployment in small islands: The MIRAB model and the French overseas departments and territories », *World development*, vol. 22, n° 12, p. 1997-2009.
- ROBETTE Nicolas, 2012, « Les espaces de vie individuels : de la géographie à une application empirique en démographie », *Cybergeo : revue européenne de géographie*, n° 605, <https://doi.org/10.4000/cybergeo.25332>.
- RALLU Jean-Louis, 2010, « The Demography of Oceania from the 1950s to the 2000s », *Population*, vol. 65, n° 1, p. 9-115.
- RALLU Jean-Louis, 1998, « Les catégories statistiques utilisées dans les DOM-TOM depuis le début de la présence française », *Population*, vol. 53, n° 3, p. 589-608.
- SIERRA-PAYCHA Celio, 2022, « L'expansion du champ migratoire polynésien au XXIe siècle : le fait d'une jeunesse qualifiée », *L'Espace géographique*, vol. 51, n° 1, p. 74-94.
- SIERRA-PAYCHA Celio, LELIÈVRE Éva, TRABUT Loïc, 2018, « Le fa'a'amura'a : Confier et recevoir un enfant en Polynésie française », *Points forts de la Polynésie française*, n° 4, p. 1-12.
- WALMSLEY Terrie L., AHMED S. Amer, PARSONS Christopher R., 2007, « A global bilateral migration data base : skilled labor, wages and remittances », *Global Trade Analysis Project*, n° 1880, p. 1-32.
- XU Xiaojian, LAVALLÉE Pierre, 2009, « Traitements de la non-réponse de lien dans l'échantillonnage indirect », *Techniques d'Enquêtes*, vol. 35, n° 2, p. 165-178.