



Note d'orientation technique

Date : Août 2020 (version n° 2 mise à jour)

Mots clés : Recensement de la population et de l'habitat, estimations modélisées de la population, couverture du recensement

Valeur des estimations modélisées de la population pour la planification et la préparation des recensements

Le recensement de la population et de l'habitat figure parmi les exercices les plus complexes et les plus vastes entrepris par un pays en temps de paix. Dans la plupart des pays à faibles et moyens revenus, le recensement de la population et de l'habitat constitue la principale source de données sur le nombre d'habitants, leur répartition spatiale, leur répartition par âge et par sexe, leurs conditions de vie et d'autres caractéristiques socio-économiques essentielles. Il fournit en outre un aperçu exhaustif de l'habitat et de ses conditions dans le pays. Les données du recensement sont essentielles pour assurer une bonne gouvernance, le développement du pays, la réduction des risques et la réponse aux crises, les programmes de protection sociale et les études du marché des entreprises. Toutefois, un recensement reste un exercice coûteux et difficile sur le plan logistique et est idéalement mené une fois tous les dix ans.

Les Instituts Nationaux de Statistique (INS) sont souvent confrontés à des obstacles qui peuvent avoir une incidence sur la mise en œuvre, la qualité des données et la réussite globale d'un recensement national de la population. Des approches statistiques innovantes sont développées et évaluées pour favoriser, et éventuellement améliorer, la planification et la mise en œuvre des recensements. Les estimations modélisées de la population, avec leurs estimations de la densité de population sur de petites zones, peuvent participer à la planification du recensement et en améliorer la qualité.

Des approches de modélisation statistique pour l'estimation des populations sont développées sur la base de la corrélation entre la densité de population et les couches de données géospatiales. Ces corrélations peuvent être exploitées dans un modèle statistique pour prédire la densité de population, avec un certain degré d'incertitude, en fonction des données d'entrée initiales. Ces modèles présentent l'avantage d'être relativement peu coûteux et peuvent fournir d'autres moyens de déduire des estimations démographiques plus récentes afin de soutenir la planification et la mise en œuvre du recensement. Cependant, en raison de leur caractère probabiliste, les estimations basées sur des modèles ne doivent pas être traitées de la même manière que les recensements de la population conventionnels, mais plutôt comme une opportunité d'affiner la planification lorsque le dernier recensement est largement dépassé. Concrètement, les estimations reposent sur des enquêtes par sondage auprès des ménages existantes ou spécifiquement mises en œuvre, telles que celles produites lors des dénombrements des ménages servant aux enquêtes. Ces échantillons sont ensuite combinés avec des co-variables géospatiales pour former la base du modèle statistique. Une fois le modèle produit, d'autres sources de données indépendantes, telles que les



Note d'orientation technique

enquêtes auprès des ménages, peuvent être utilisées pour évaluer les estimations de population nouvellement modélisées.

La population est modélisée en fonction de co-variables basées sur l'imagerie satellite ou dérivées de données géospatiales, telles que la distance par rapport aux routes, l'intensité de l'éclairage nocturne ou la classification de la couverture terrestre. Souvent, les meilleurs indicateurs de la densité de population sont la présence, la densité et la configuration des infrastructures. En supposant que les populations sont principalement situées dans des zones habitées, nous pouvons estimer la population d'un pays en combinant les densités de population avec les cartes des zones habitées. Pour ce faire, il est nécessaire de cartographier les zones de peuplement, le plus souvent à l'aide d'images satellites récentes à très haute résolution. Le modèle peut alors prévoir les densités de population ou le nombre de personnes par unité de surface pour toutes les unités géographiques, telles que les unités administratives ou les zones de dénombrement, ou même des quadrillages de 100 mètres sur 100 mètres.

En matière de fiabilité, les estimations modélisées contiennent un degré d'incertitude qui peut se traduire par des intervalles de confiance de différents ordres de grandeur autour de la densité de population prévue. À ce jour, aucun mécanisme ne permet de déterminer le degré d'incertitude acceptable, ni même s'il doit être inclus dans la prise de décision. Pour les indicateurs généraux basés sur des données d'enquête, il est recommandé de considérer qu'un indicateur n'est pas fiable lorsque l'écart type relatif (écart type par rapport à la moyenne) est inférieur à 33 %. Néanmoins, même si les estimations de la population sont sujettes à des imprécisions, elles peuvent :

- contribuer à la planification et à la cartographie du recensement
- se substituer au recensement dans les zones inaccessibles
- ensembles de données opérationnelles communes sur les statistiques de la population
- servir de données pour mettre à jour la base de sondage de l'échantillon maître
- permettre d'évaluer la couverture du recensement
- permettre d'anonymiser les résultats du recensement
- permettre de combiner les recensements de la population et de l'habitat avec les recensements agricoles

Contribution à la planification et à la cartographie du recensement

Le processus de planification du recensement et du travail cartographique sur le terrain repose fondamentalement sur les projections (ou prévisions) de population par région, municipalité, village et zone de dénombrement (ZD). Ces prévisions de population donnent des informations sur l'allocation des ressources relative à la mise en œuvre du recensement, y compris les fonds, le temps alloué pour les différentes étapes du recensement, le personnel et les opérations nécessaires. Cependant, les projections démographiques officielles produites juste après le



Note d'orientation technique

recensement peuvent devenir de plus en plus dépassées au fil du temps si elles ne sont pas continuellement mises à jour avec les données des enquêtes plus récentes. En outre, si les projections démographiques peuvent être précises aux niveaux national et régional, elles peuvent être inexactes à des niveaux inférieurs, notamment au niveau de la ZD. Comme les estimations démographiques modélisées sont basées sur les co-variables géospatiales les plus récentes disponibles, elles fournissent souvent une estimation plus précise au niveau du district, du village ou de la ZD.

Les estimations modélisées de la population, combinées aux données géospatiales des réseaux routiers et fluviaux, peuvent être utilisées pour planifier les itinéraires les plus efficaces des cartographes. Un travail de terrain plus efficace est susceptible de réduire les coûts globaux et le temps nécessaire pour réaliser la cartographie. En raison de l'important coût financier que représente un recensement, il est essentiel de prendre en compte les possibilités d'économies potentielles. Ces économies peuvent être réalisées grâce au temps nécessaire au cartographe ou à l'agent recenseur pour effectuer son travail sur le terrain, grâce à un acheminement plus efficace, à une planification plus précise du nombre de cartographes nécessaires pour chaque ZD, etc. En outre, la cartographie des lieux d'habitation basée sur l'imagerie satellite améliore la planification et la mise à jour de la cartographie car elle permet d'identifier de petites zones de peuplement qui n'apparaissent pas auparavant sur les cartes. Enfin, les estimations modélisées de la population ainsi que les données sur les réseaux fluviaux et routiers peuvent être utilisées pour déduire de nouvelles zones de dénombrement.

Alternative au recensement dans les zones inaccessibles

Lors d'un recensement, certains pays sont confrontés à des problèmes de sécurité ou d'accès dans certaines parties de leur territoire, où il est impossible d'envoyer des agents recenseurs. Dans ce cas, il est possible de modéliser les chiffres de population et les densités des zones inaccessibles, en utilisant soit les données de recensement qui ont été collectées, soit des données d'enquête spécialement conçues (Schéma n° 1). L'exactitude des estimations dépendra de la qualité du dénombrement dans les zones accessibles, des performances intrinsèques du modèle et de la vraisemblance des hypothèses de modélisation. Le modèle de population développé est toujours adapté à la situation en question (c'est-à-dire à la disponibilité des données et au contexte de la population) dans le but de produire les meilleures estimations de population possibles compte tenu des données en entrée et du temps disponibles.

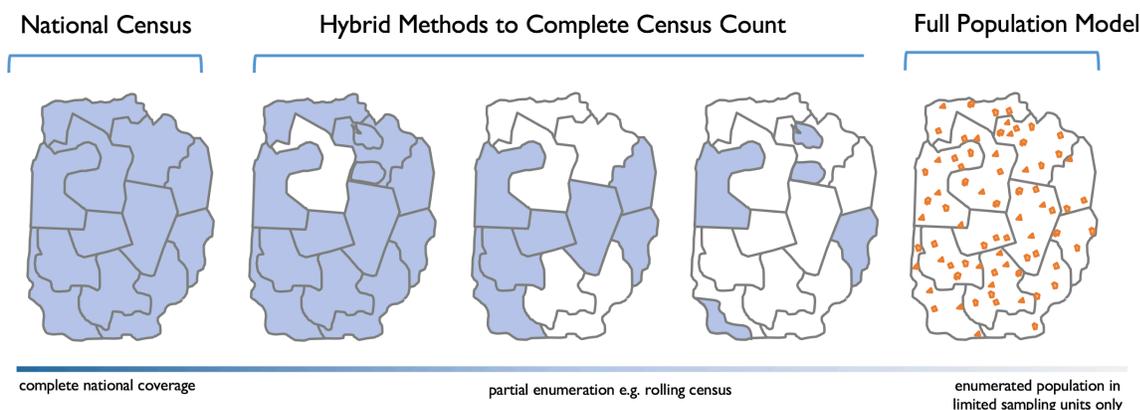


Schéma n° 1

Ensembles de données opérationnelles communes sur les statistiques de la population

Les ensembles de données opérationnelles communes (COD) sont des ensembles de données de référence faisant autorité, nécessaires pour soutenir les opérations et la prise de décision de tous les intervenants humanitaires. Les COD-PS (ensembles de données opérationnelles communes sur les statistiques de populations) sont des ensembles de données démographiques actualisées, ventilées par sexe et par âge, au niveau géographique le plus bas disponible. Elles doivent être utilisées dans les contextes humanitaires uniquement pour répondre aux besoins de données afin de réagir et d'intervenir en cas de crise. En tant que telles, elles préconisent de choisir le critère plus souple et pragmatique de « meilleures données disponibles » plutôt que les normes statistiques officielles. Dans ce cas, il est possible d'estimer les COD-PS et les densités des zones inaccessibles en utilisant les données du recensement qui ont été partiellement collectées pour établir un modèle. Ainsi, lorsque les données démographiques modélisées constituent la meilleure source disponible pour des données actualisées sur le sexe, l'âge et la répartition géographique, elles peuvent être utilisées comme COD-PS pour soutenir les opérations et la prise de décision dans les contextes humanitaires.

Base de sondage de l'échantillon maître et estimations de la population

Une fois les données du recensement validées, un échantillon maître constitué des ZD est généralement tiré de la base de données du recensement. Cet échantillon est utilisé pendant la période intercensitaire pour mener les principales enquêtes socio-économiques et démographiques du pays. Les paramètres de cette base de sondage de référence ont tendance à s'écarter de ceux de la population au fil du temps, en raison de la croissance démographique et des changements dans la distribution spatiale. Selon le type de cartographie de recensement à partir duquel l'échantillon de référence initial a été tiré, les estimations de population carroyées peuvent être



Note d'orientation technique

utilisées pour mettre à jour la base de sondage de l'échantillon maître de deux façons. Si la cartographie du recensement était numérique, il serait possible de mettre à jour la base de sondage de l'échantillon maître en actualisant les populations des ZD à travers les estimations carroyées. Si la cartographie n'était pas numérique, la mise à jour des populations des ZD à travers les estimations carroyées ne pourrait pas être appliquée car les limites des ZD ne sont pas numériques et il n'est donc pas possible de superposer les couches des estimations de populations carroyées aux couches des ZD. Dans cette situation, il est conseillé de prélever un nouvel échantillon de référence dans la population carroyée, en considérant chaque grille ou groupe de grilles contiguës comme des unités d'échantillonnage primaires. Cette deuxième stratégie est une amélioration du fameux échantillonnage aréolaire, puisque dans ce cas, nous avons une estimation plus précise de la population des aires (qui dans ce cas sont des grilles ou des groupes de grilles contiguës). Comme pour les autres utilisations, cette approche n'est possible que si les estimations modélisées sont suffisamment fiables pour refléter les changements des paramètres de la population.

Évaluation de la couverture du recensement

En général, après la phase de dénombrement, la couverture du recensement est évaluée à l'aide de différentes méthodes et sources de données (projections démographiques, données de registres, données sur la santé, données sur l'éducation, etc.), qui pourraient inclure la mise en œuvre d'une enquête post-dénombrement.

Grâce aux estimations modélisées de la population, nous disposons d'une source de comparaison supplémentaire. Pour que les estimations modélisées de la population puissent servir de source d'évaluation de la couverture du recensement, l'estimation doit être réalisée en utilisant des données provenant de zones de dénombrement dont la couverture et la qualité des données sont de bonne qualité. Une fois de plus, les estimations modélisées ne doivent être prises en compte que si la performance du modèle est élevée et les hypothèses sous-jacentes réalistes.

Anonymisation des résultats du recensement

Les méthodes statistiques peuvent être utilisées pour anonymiser l'ensemble des données du recensement (Schéma n° 1, dessin de gauche). Les observations individuelles peuvent être agrégées en totaux de population carroyées à haute résolution à l'aide de logiciels SIG et de méthodes statistiques. Grâce à ces méthodes, les observations (par exemple les cumuls administratifs) peuvent être désagrégées à une résolution de 100 x 100 m et ainsi créer une couche de données anonymes utilisable par le gouvernement pour la prise de décision et la planification d'enquêtes et par le grand public, si elles sont rendues publiques.

Combinaison avec les recensements agricoles

La combinaison du recensement de la population et de l'habitat avec le recensement agricole au sein du système statistique national est conforme à la stratégie globale pour l'amélioration des



Note d'orientation technique

statistiques agricoles et rurales¹. Au niveau opérationnel, cette combinaison peut être réalisée 1) en établissant un cadre de recherche intégré, 2) en développant une base de sondage de l'échantillon maître pour l'agriculture, et 3) en intégrant les systèmes de gestion des données. L'expérience montre que la collecte de données agricoles limitées et bien définies lors du recensement de la population et de l'habitat peut contribuer de manière substantielle à la mise en place d'un cadre général efficace pour les recensements et les sondages agricoles dans de nombreux pays en développement. En outre, les estimations modélisées de la population permettent d'affiner le cadre d'échantillonnage pour les recensements agricoles, car les densités de population reflètent l'occupation du sol, ce qui fournit des informations sur les zones agricoles potentielles à faible densité de population et les zones non agricoles potentielles à forte densité de population. Ces informations pourraient ensuite être utilisées pour affiner la base de sondage agricole :

- en combinant des estimations modélisées avec des informations géographiques provenant du recensement agricole, et
- en classant les zones urbaines ou peuplées qui doivent être exclues de l'établissement de la base de sondage pour les enquêtes agricoles, car cela permettrait d'éviter l'utilisation de ressources pour des opérations sur le terrain dans des zones non agricoles.

Conclusion

Les estimations modélisées de la population sont potentiellement utiles à différentes étapes d'un recensement, de la planification à la mise en œuvre, de l'évaluation de la qualité à l'utilisation des données. Toutefois, il est essentiel de noter que ces estimations doivent être utilisées avec prudence : différents modèles produiront différentes estimations de la population. Heureusement, il existe des mécanismes permettant de « choisir » objectivement le modèle le plus précis sur le plan statistique et de le valider. Même s'il est possible de choisir le meilleur modèle, il est impératif de rappeler qu'un modèle géospatial de population ne peut pas refléter correctement le monde réel. Comme l'a dit George Box, « N'oubliez pas que tous les modèles sont faux par nature ; en pratique, il convient de se demander jusqu'à quel point ils doivent être erronés pour être inutiles ».

Ce dossier technique (Version n° 2) a été préparé par Sabrina Juran, Mathias Kuepie, Maureen Jones, Catalina Valle, Kathrin Weny, Mathias Nathan du Fonds des Nations Unies pour la population en collaboration avec Andy Tatem, Attila Lazar, Douglas Leasure et Heather Chamberlain de WorldPop à l'Université de Southampton.

¹ Organisation des Nations unies pour l'alimentation et l'agriculture. Stratégie globale pour l'amélioration des statistiques agricoles et rurales. 2010
http://www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/ICAS5/Ag_Statistics_Strategy_Final.pdf