

Combiner les données du recensement et de Manitoba Hydro pour comprendre la consommation d'électricité résidentielle

Christopher Duddek¹

Résumé

Depuis 2015, Statistique Canada reçoit des fichiers mensuels de Manitoba Hydro. Manitoba Hydro étant le fournisseur d'électricité de la plupart des Manitobains, il pourrait sembler facile d'obtenir la consommation totale d'électricité résidentielle annuelle. Cependant, la comparaison avec les estimations publiées relève des écarts. Pour mieux en comprendre les raisons, nous modifierons dans le présent article les variables géographiques du fichier de façon à comparer les données de Manitoba Hydro avec les comptes du Recensement de 2016. La comparaison nous permet de cerner des problèmes découlant du mode de création des fichiers de Manitoba Hydro. Nous étudierons une méthode d'évaluation de la couverture et de l'incidence des erreurs de couverture sur les estimations de la consommation obtenues.

Mots-clés : Validation des données; données administratives; recensement; couverture.

1. Pourquoi cette étude?

Le remplacement des données d'enquête par des données administratives fait partie de la vision stratégique de Statistique Canada (Cloutier, 2010). C'est ce qui explique que nous cherchions à obtenir des données sur la consommation électrique auprès de tous les principaux fournisseurs d'électricité du Canada. Il faut cependant s'interroger sur la qualité des données reçues.

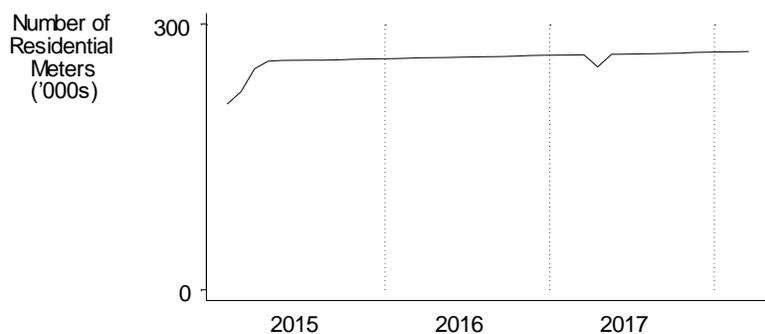
Pour commencer l'analyse des données de Manitoba Hydro, nous avons le luxe de pouvoir comparer nos estimations avec des données déjà publiées. L'un des résultats de l'Enquête sur l'approvisionnement et l'écoulement de l'électricité est la consommation résidentielle totale d'électricité (tableau 25-10-0029-01, anciennement CANSIM 128-0016). Si nous comparons ces résultats avec la sommation des kilowattheures (kWh) des résidences figurant dans le fichier d'Hydro, nous constatons que les estimations du fichier d'Hydro sont inférieures pour 2015 et 2016 (de 16 % et 8 % respectivement).

Nous savons qu'aucune des deux sources n'est parfaite. D'un côté, les estimations d'enquête s'accompagnent d'erreurs d'échantillonnage et d'erreurs non attribuables à l'échantillonnage, comme celles liées à la non-réponse. De l'autre, les données administratives peuvent comporter des erreurs de classification et de couverture. À titre d'exemple, la figure 1-1 montre le nombre mensuel de compteurs d'électricité à Winnipeg selon le fichier d'Hydro. Au début de la série et pour un mois de 2017, il y a moins de compteurs que ce qu'on pourrait attendre, si l'on suppose que le nombre de compteurs devrait être pratiquement constant et n'augmenter que légèrement au fil des ans.

Étant donné que le Manitoba compte des centaines de milliers de compteurs d'électricité, comment peut-on évaluer la couverture? Dans ce qui suit, nous tenterons de décomposer le problème en éléments plus petits et plus faciles à traiter.

Figure 1-1
Nombre mensuel de compteurs résidentiels à Winnipeg

¹Christopher Duddek, Statistique Canada, Immeuble R.-H.-Coats, 16^e étage, section B, Ottawa, Canada, K1A 0T6, Christopher.Duddek@canada.ca



2. Avantages des coordonnées GPS

Si nous examinons le nombre de ménages du Manitoba au moment du Recensement de 2016 et que nous le comparons avec le nombre de compteurs dans le fichier d'Hydro en mai 2016, nous observons 10 % de compteurs en moins par rapport aux ménages. Comment résoudre ces différences? Une partie de la réponse consiste à se concentrer sur des niveaux géographiques inférieurs.

En plus de fournir des renseignements sur les adresses, le fichier de Manitoba Hydro donne les coordonnées du système de positionnement mondial (GPS) de presque tous les compteurs. Cet extraordinaire degré de précision permet d'attribuer des classifications géographiques définies par le recensement à chaque compteur. Dans le cas étudié, nous avons cherché à nous placer à l'un des plus bas niveaux géographiques du Recensement de la population de 2016, à savoir l'aire de diffusion (AD). Le Manitoba comporte un peu plus de 2 000 AD. Nous avons pu, au moyen d'ArcGIS, attribuer la plupart des compteurs électriques à ce niveau. Pour les compteurs restants, nous avons utilisé différentes méthodes, comme l'utilisation du code postal connexe, pour faire correspondre le compteur à une AD.

Une fois que chaque compteur a été attribué à une région géographique du recensement, nous avons été en mesure de comparer le nombre de compteurs des fichiers d'Hydro avec le nombre de ménages dénombrés par le recensement. La tendance observée montre un compteur par logement classé comme logement individuel non attenant. Si un seul ménage habitait le logement, nous avons un rapport de 1:1 de compteurs par ménage.

Les observations étaient moins simples pour les appartements. Parfois, chaque unité avait son propre compteur. D'autres fois, il y avait seulement un ou deux compteurs pour desservir un immeuble à logements multiples. De plus, les compteurs desservant plusieurs unités étaient parfois classés comme compteurs commerciaux et non résidentiels. Cette constatation impliquait qu'il nous fallait trouver un moyen d'évaluer les compteurs par rapport aux logements à des niveaux géographiques bas.

3. Valider les aires de diffusion échantillonnées à Winnipeg

Afin d'évaluer la couverture du fichier d'Hydro, nous avons décidé de concentrer nos efforts sur la ville de Winnipeg. Notre choix s'explique par deux facteurs : d'une part, Winnipeg abrite plus de la moitié de la population du Manitoba et, d'autre part, il s'agit d'une région urbaine où les adresses et le type de bâtiment sont facilement vérifiables au moyen de Google Maps.

Figure 3-1
Échantillon sélectionné (grands centroïdes) des aires de diffusion de Winnipeg



Sur les 1115 AD composant Winnipeg, nous en avons sélectionné 27 au hasard (figure 3-1). Nous avons ensuite procédé à résolution de la répartition des types de logements, une AD après l'autre.

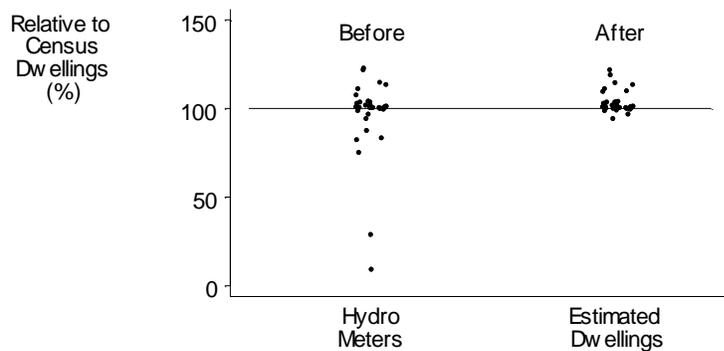
Pour chaque AD, nous avons obtenu le nombre de logements selon les données du recensement. Les principaux types de logements dans cette étude étaient (1) les logements individuels non attenants et (2) les appartements. Le Recensement de 2016 constituait notre source pour connaître le nombre de logements pour chaque type de logement. Si le nombre total de logements selon le recensement était proche du nombre de compteurs résidentiels d'électricité, nous supposions que la couverture était bonne. Quand ce n'était pas le cas, nous avons procédé à une vérification.

La première étape de la vérification a consisté à examiner les données des compteurs selon Hydro et à déterminer si certains des compteurs résidentiels étaient des compteurs d'entreprise au moyen du nom associé au compteur. Si le nom était celui d'une entreprise, nous avons changé la classification du compteur pour qu'il soit compté comme compteur commercial.

La deuxième étape a consisté à examiner les comptes du recensement par type de logement, puis à attribuer un type de logement à chaque compteur. Cette stratégie était un exercice manuel qui prenait beaucoup de temps, mais qui pouvait être réalisée au moyen de stratégies selon la répartition des logements dans une AD.

Si, par exemple, la majorité des logements étaient des logements individuels non attenants avec quelques appartements, nous cherchions les appartements et trouvions ensuite les compteurs qui correspondaient à leurs adresses. Nous n'avons pas essayé de couplage d'enregistrements direct entre les fichiers d'Hydro et le recensement. Nous avons plutôt examiné les compteurs à forte consommation et le nom qui leur était associé pour voir s'il s'agissait de toute évidence d'un appartement, ou nous observions si l'image de la rue sur Google Maps à cette adresse montrait un appartement.

Figure 3-2
Résolution du nombre de compteurs par rapport au nombre de logements estimés



Même si cela s’est révélé parfois difficile, nous avons constaté que nous pouvions atteindre nos objectifs en déployant suffisamment d’efforts. À l’occasion, nous trouvions qu’un compteur qui desservait un appartement était classé comme compteur commercial. Dans ces cas, nous modifions la classification à « résidentiel ». Nous avons fini par obtenir un tableau présentant les compteurs d’électricité répartis selon la même classification de types de logement du recensement.

La troisième étape a consisté à faire correspondre les compteurs au nombre estimé de logements. Nous avons supposé que les logements individuels non attenants auraient un compteur principal par logement. La correspondance était plus intéressante pour les appartements. Si un immeuble d’appartements contenait plusieurs compteurs, nous avons supposé que chaque appartement avait son propre compteur. Si, en revanche, l’immeuble d’appartements ne contenait qu’un compteur, nous utilisons à nouveau Google Maps pour estimer le nombre d’unités desservies. À la fin du processus, nous avons attribué un rapport de 1:1 ou de 1:N de compteurs par logement.

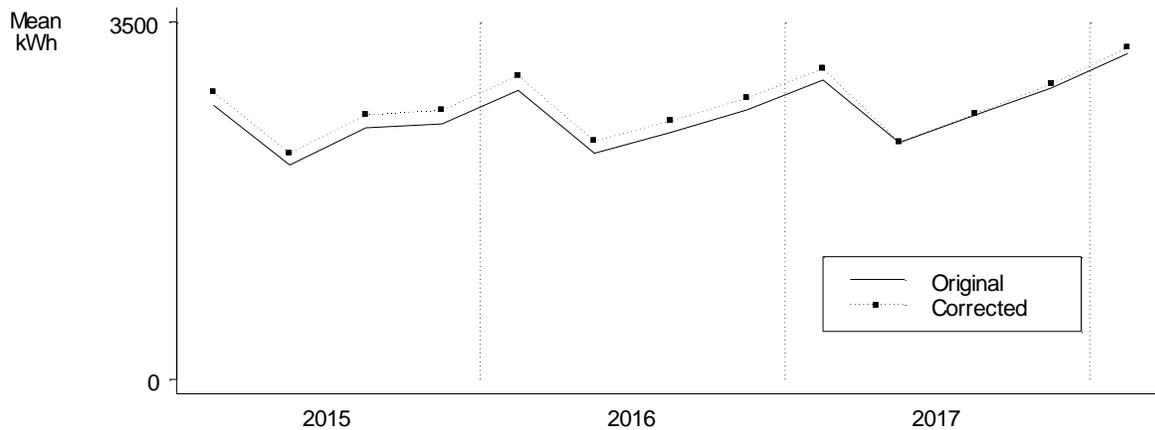
La dernière étape a consisté à calculer la somme des logements estimés desservis par les compteurs d’électricité et à la comparer avec les comptes du recensement. La figure 3-2 montre les répartitions « avant » et « après » des compteurs d’électricité et des logements estimés par rapport au nombre de logements selon le recensement pour les AD échantillonnées. Grâce à ce processus de résolution, nous sommes convaincus d’avoir une bonne couverture des logements dénombrés lors du Recensement de 2016. Si le processus n’avait pas été aussi laborieux, nous aurions appliqué cette même procédure de vérification aux autres AD de Winnipeg. Les résultats de l’échantillon que nous avons sélectionné étaient rassurants.

4. Estimation des totaux

Lavigne et Nadeau (2014) se servent d’un cadre pour déterminer l’adéquation des données administratives. L’une des dimensions de qualité qu’ils mentionnent est l’exactitude, car ils savent que les erreurs mènent à des erreurs de mesure. Après avoir prélevé un échantillon que nous avons ensuite évalué, nous pouvons procéder à l’évaluation de l’exactitude.

Nous savons que la classification erronée des compteurs influe sur notre estimation de la consommation d’électricité. D’une part, certains compteurs sont classés comme compteurs résidentiels, alors qu’ils desservent une entreprise. D’autre part, certains compteurs fournissent de l’électricité à des immeubles d’appartements, mais ils sont classés comme compteurs commerciaux. En quoi cette classification erronée pèse-t-elle sur les estimations de la consommation?

Figure 4-1
Incidence des erreurs de couverture sur les estimations trimestrielles de la consommation d’électricité



Les données d’Hydro sont surtout des données mensuelles dont les dates de début et de fin sont réparties uniformément au cours du mois. La calendarisation des données en intervalles mensuels signifie que nous devons attribuer une partie de la consommation mensuelle déclarée à différents mois, ce qui introduit une erreur de mesure (voir Fortier et coll., 2010). Pour atténuer cet effet, nous avons calendarisé les données d’Hydro sur des périodes trimestrielles.

Pour en revenir à notre échantillon, nous avons calculé la consommation selon la classification originale et corrigée. Nous avons ainsi obtenu une indication du biais attribuable à la classification erronée. La figure 4-1 montre que l’incidence est modeste. Cela s’explique en partie par le fait que les erreurs de classification s’annulent les unes les autres, dans une certaine mesure. D’après nos résultats, les différences trimestrielles varient de 1 % à 5 %, selon le trimestre.

5. Conclusion

Nous avons utilisé les coordonnées GPS des fichiers d’Hydro pour cartographier les compteurs en fonction des régions géographiques du recensement. Nous avons ainsi pu comparer les chiffres des logements selon le recensement avec le nombre de logements estimés tirés des fichiers d’Hydro. À partir de l’échantillon que nous avons sélectionné, nous avons été en mesure d’affirmer que la couverture des fichiers d’Hydro est bonne. Certes, certains compteurs résidentiels et commerciaux sont classés de manière erronée mais l’erreur étant présente dans les deux sens, elle a une incidence faible sur les estimations de la consommation.

Le travail présenté visait à comprendre la qualité de la couverture des fichiers d’Hydro. L’un des avantages procurés par ce travail est l’ajout d’une covariable à une petite partie du fichier qui peut nous servir à avoir une idée de la façon dont la consommation varie entre différents types de logements. Nous avons observé que la consommation d’électricité était nettement plus faible dans les appartements que dans les logements individuels non attenants. Cette observation sera utile dans l’évaluation des valeurs aberrantes d’une enquête comme l’Enquête sur les ménages et l’environnement.

Les travaux futurs devraient porter sur d’autres parties du Manitoba, situées à l’extérieur de Winnipeg, afin de s’assurer que la qualité du reste du fichier est aussi grande pour ce qui est de la couverture. Nous pourrions constater que l’erreur de classification est plus faible si la proportion d’appartements est moindre à l’extérieur de la ville de Winnipeg.

La notion selon laquelle les dossiers administratifs feraient partie intégrante des travaux des organismes statistiques nationaux a été cernée il y a trente ans au moins (Brackstone, 1987). Cependant, dans les faits, l’intégration d’autres sources de données a été plus lente que prévu en raison de problèmes d’accès, de protection des renseignements personnels et d’adéquation. Le climat actuel de réduction des taux de réponse et d’augmentation des coûts a ravivé

l'intérêt pour l'obtention de nouveaux flux de données préexistantes et exigera davantage que nous évaluions la qualité des données reçues. La validation des données au moyen d'une comparaison avec des sources de données comme le recensement aidera à atteindre cet objectif.

Bibliographie

Brackstone, G. J. (1987), « Issues in the Use of Administrative Records for Statistical Purposes », *Survey Methodology*, 13, p. 29-43.

Cloutier, M. (2010), « A Strategic vision for the use of administrative data at Statistics Canada », *Proceedings of Statistics Canada Symposium 2010*, p. 432-436.

Fortier, S., B. Quenneville, et F. Picard (2010), « Calendarization methods and applications », documentation préparée pour le Conseil consultatif canadien de la statistique, Ottawa, Canada: Statistique Canada.

Lavigne, M., et C. Nadeau (2014), « Un cadre d'évaluation des données administratives », *Recueil du Symposium de 2014 de Statistique Canada*.