Adele Lauzon CGS-M Research Proposal

Census-Independent Population Estimation in Northern Canada

Background: Census data that is accurate at the regional level is essential for appropriate policy formulation and resource allocation (Alexander & Alkema, 2022). In remote, rural regions of Northern Canada where the internet is unreliable, census takers must travel door-to-door to produce regional population estimates. In 2016, Statistics Canada employees visited 30,000 homes in northern regions spanning 40% of Canada's land mass (DeCoste, 2016). Gathering information in the north is exorbitantly expensive compared to the rest of the country (DeCoste, 2016). Moreover, census takers have been impacted by increasingly unpredictable weather conditions, which make traditional methods of data collection less reliable (DeCoste, 2016; McBride, 2018).
Despite these efforts, nonresponse rates in the North are higher than the rest of the country, particularly in the territories (Statistics Canada, 2021). Additionally, though statistical methods like imputation are used to calculate population estimates, there are no measures of uncertainty published in census reports. If population is estimated in a Bayesian framework, uncertainty in data and spatial processes can be taken into account (Alexander & Alkema, 2022).

Objective: In this study, we propose using Bayesian hierarchical modelling in tandem with satellite data to estimate regional populations in the Canadian territories. Census-independent population estimation can mitigate the costs of door-to-door canvassing and provide helpful uncertainty measurements.

Theoretical Approach: The Bayesian hierarchical modelling framework will adapt previous research focused on low-income countries to a Canadian context (Boo et al., 2022; Leasure et al., 2020; Weber et al., 2018). Population densities will be estimated using data from WorldPop Global and geospatial covariates and validated on community estimates from the 2021 Canadian Census (Leasure et al., 2020). Geospatial covariates will be derived from the Settlement Mapper Tool (Cheriyadat, 2007), which produces a settlement map using a support vector machine classifier trained on high-resolution satellite imagery (Weber et al., 2018). Geospatial covariates will include WorldPop estimates, school densities, and household sizes (Leasure et al., 2020).
At a high level, the hierarchical modelling framework will be as follows: the number of people in each community will be estimated with a Poisson model of density and area, and density will be modelled using the Log-Normal of the log-scale density estimate based on geospatial covariates. Log-scale density will be modelled hierarchically by settlement type, community, region, and territory, and parameters will be estimated using the Markov chain Monte Carlo method, validated with 10-fold cross-validation against 2021 Census data.

Implications: Population estimates complete with error measurements can provide a more complete picture of Northern Canada, especially when used in conjunction with traditional census data. Lawmakers can be more certain that appropriate resources are directed towards geographically vulnerable communities.

Bibliography:

Alexander, M., Alkema, L. (2022). A Bayesian Cohort Component Projection Model to Estimate Women of Reproductive Age at the Subnational Level in Data-Sparse Settings. *Demography, 59(5)* 1713-1737. https://doi.org/10.1215/00703370-10216406

Boo, G., Darin, E., Douglas, R. L., Dooley, C. A., Chamberlain, H. R., Lazar, A. N., Tschirhart, K., Sinai, C., Hoff, N. A., Fuller, T., Musene, K., Batumbo, A., Rimoin, A. W., Tatem, A. J. (2022). High-resolution population estimation using household survey data and building footprints. *Nat Commun 13*, 1330. https://doi.org/10.1038/s41467-022-29094-x

Cheriyadat, A., Bright, E., Potere, D., Bhaduri, L. (2007). Mapping of settlements in high-resolution satellite imagery using high performance computing. *GeoJournal (69)*, 119-129. 10.1007/s10708-007-9101-0.

DeCoste, L. (2016). In Northern Canada, the census can be a matter of life and death. *The Globe and Mail*.

Leasure, D. R., Jochem, W. C., Weber, E. M., Tatem, A. J. (2020). National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty. *PNAS, 117 (39)* 24173-24179. https://doi.org/10.1073/pnas.1913050117

McBride, K. (2018). Data Resources and Challenges for First Nations Communities. *Alberta First Nations Information Governance Centre.*

Statistics Canada (2020). The road to the 2021 Census. *The Government of Canada.*

Statistics Canada (2023). Census of Population. *The Government of Canada.*

Weber, E., Seaman, V. Y., Stewart, R. R., Bird, T. J., Tatem, A. J., McKee, J. J., Bhaduri, B., Moehl, J. J., Reith, A. E. (2018). Census-independent population mapping in northern Nigeria. *Remote Sensing of Environment, 204*: 786-798. https://doi.org/10.1016/j.rse.2017.09.024