Literature Review: Population Estimation in Northern Canada

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).

This literature review will explore existing work related to census-independent population estimation and how it could be extended to northern Canada. First, we will review common tools for census-independent population estimation, focusing on remote sensing and Bayesian hierarchical modelling. Since population estimation research has not been conducted in

Canada, we will then review how similar methods have been applied in other fields in the Canadian context. Finally, we will discuss the integration of Indigenous Knowledge and Indigenous Data Sovereignty principles with geospatial research, which is essential for culturally aligned research in Northern Canada.

Population Estimation using Satellite Imaging and Geospatial Data

Remote sensing (satellite imaging) and geospatial data are essential tools for population estimation in regions where traditional census is outdated or missing. Most often, this work is conducted in nations or regions where the local government does not have the resources to conduct a full census, such as Addis Ababa, Ethiopia (Dittakan et al., 2013) or Bo, Sierra Leone (Hillson et al., 2019). In regions where population growth is high and census data is old or altogether unavailable, reasonable population estimates are crucial for health planning, refugee support, and appropriate allocation of resources and services. Usually some combination of remote sensing (satellite) imagery and survey data are used to produce these estimates.

Combining satellite data with additional data sources that provide population information at a sub-regional scale allows for the production of finer estimates. For example, Leasure et al. (2019) combined Nigerian microcensus data outlining population by settlement type with satellite images of various Nigerian states. In another study mapping refugee populations, regularly collected refugee registration data was used in tandem with satellite derived settlement maps (Darin et al., 2024) to train the model. Depending on the predictive goal, the microcensus may also capture the number of people per household or structure (Dittikan et al., 2013; Weber et al., 2018; Boo et al., 2022). While incorporating an additional data source may provide more accurate estimates, this approach can be time- and resource-intensive as it not only requires images to be manually segmented or labelled, but also depends on the availability of microcensus data.

Alternatively, some researchers trade off accuracy and precision for efficiency by using automated segmentation methods or relying solely on the images for data. Dittakan et al. (2013) employed an automated quad-tree processing technique to segment images into individual households, which when combined with a household size microcensus provided reasonable accuracy measures. Hillson et al. (2019) investigated how accurate a model can be without the use of an additional data layer, only utilizing the satellite images and their inherent properties. Using a set of preestablished geometric covariates for population estimation, the researchers found the method to be less computationally intensive, though less accurate. In instances where microcensus data is not available or manual image segmentation is not viable, these alternative approaches should be considered.

Irrespective of including additional data layers, quality satellite data itself is not widely nor easily accessible. While image data can be obtained from sources like Google Earth or Landsat (Hillson et al., 2019; Dittikan et al., 2013), free satellite data tends to be at a coarser resolution (30-50m), resulting in less accurate estimates. High resolution data is available through private platforms such as Maxar Technologies (Boo et al., 2022) or Ecopia AI (Darin et al., 2024), though costly.

While producing reasonable spatial estimates of long-term, stable populations is a challenging task, providing estimates for transient populations, such as refugees, is even more difficult. Darin et al. (2024) used humanitarian administrative data, high-resolution settlement maps derived from satellite imagery, and state-provided gridded population estimates to produce grid-cell estimates of refugee counts in Cameroon. This research provides a framework for connecting manually-inputted spatial textual information to geocoded information, and how routinely collected operational data can be repurposed for research. This could be a useful resource for future work on population modelling of transient or migratory populations in other contexts -- for example, internal climate-forced displacements of Indigenous peoples in Canada.

Population Estimation using Bayesian Hierarchical Modelling Frameworks

Another popular technique for population estimation is some variation of Bayesian modelling, a flexible modelling framework wherein some prior information (eg. census or survey data) can be incorporated into the model. Bayesian modelling techniques for population estimation are often referred to as “bottom-up” techniques as they rely on small area enumerations or microcensuses to predict populations in unsampled areas. This approach differs from “top-down” methods that disaggregate census data into smaller gridded areas. Bottom-up approaches are useful in areas where censuses are incomplete or enumeration is not possible (Nilsen et al., 2021).

One particularly useful Bayesian modelling technique for population estimation is hierarchical modelling. Originally used in ecology, these nested models are suitable for mapping populations in data-sparse environments and producing uncertainty measures (Leasure et al., 2020). The hierarchical structure allows the model to “borrow” estimation strength from national or regional data to improve prediction in areas where data is more limited (Leasure et al., 2020; Alexander & Alkema, 2022). Further, the hierarchical framework allows population density to vary across space and socioeconomic context, which can be captured and explained by incorporating informative covariates in the model (eg., settlement type or region).

As stated, the flexibility of Bayesian hierarchical modelling makes them a useful tool in population estimation, as they can be integrated with a variety of other data sources. Boo et al. (2022) integrated a hierarchical model with high-resolution building footprint data and household surveys to produce a very strong model (r-squared = 0.79), indicating a high level of explanatory power for predicting populations gridded at 100m in the Democratic Republic of the Congo. Other work integrates Bayesian modelling with settlement maps and microcensus data (Weber et al., 2018; Leasure et al., 2020).

While Bayesian techniques are often coupled with remote sensing data, images are not required to produce subnational population estimates. Alexander and Alkema (2022) produced estimates of women of reproductive age in Kenya using a Bayesian cohort component projection model, incorporating both data from the census, household surveys, and administrative records, as well as components of population change. Mortality schedules, fertility, and migration are included in the modelling framework to reflect natural population changes and processes that affect women of reproductive age over time. The Bayesian framework allows for the production of population estimates with associated uncertainty metrics, where uncertainty may come from sampling or nonsampling error.

A notable advantage of Bayesian modelling is this ability to produce quantitative measures of uncertainty in the form of prediction intervals and posterior probabilities (Leasure et al., 2020). Model uncertainty comes from limitations in its ability to explain features of the data, perhaps due to observational error or small sample sizes (Boo et al., 2022). Weber et al. (2018)

developed a novel technique for providing population estimates as prediction intervals specific to the size and composition of the region of interest.

Spatial and Temporal Analyses in Canada: Environment and Epidemiology

Because most population estimation research is done in regions that do not have the resources to conduct regular censuses, there does not yet exist examples of this research applied to the Canadian context. However, given the unique challenges imposed by the Canadian landscape, it may be worthwhile to consider research in other fields that use techniques like remote sensing and Bayesian modelling.

Satellite imaging has been leveraged in ecology research to track and monitor environmental changes in remote Canadian regions. Fraser et al. (2011) used Landsat images to detect changes in vegetation in four northern national parks over a 25 year period. Similar to previous population research, the inputs to the model were the geometric and spectral qualities of the images, and the researchers found that the 30m Landsat resolution allowed for easier regional-level change detection compared to finer resolution imagery. Landsat data was also used by Travels-Smith et al. (2024) to model vertical and horizontal vegetation structure in northern forests. In addition to Landsat surface reflectance data, the researchers integrated spatial data concerning vegetation heights from ICESat-2. The researchers note a difficulty in using satellite imagery for northern climates is greater reflectivity of snow covered surfaces, impacting consistency of canopy height estimates. Bartsch, Strozzi, and Nitze (2023) encountered similar issues in permafrost modelling, noting that satellite imaging research in circumpolar regions is lacking due to the need of higher spatial resolution data in heterogeneous tundra environments.

While both groups of researchers encountered issues related to estimation of environmental factors, spectral difficulties as a result of snow cover may also prove to be a challenge in population estimation, perhaps complicating structure identification or settlement classification.

Olthof & Fraser (2024) also worked with a time-series mapping problem when mapping small-scale water features of the Hudson Bay Lowlands. The researchers used 30m resolution Landsat data and a sub-pixel mapping approach to capture finer-grain water dynamics. The researchers compared data-driven machine learning methods to physical linear unmixing models, and found the linear models to outperform the machine learning methods. Given that most open-access satellite data is at a similar resolution, similar sub-pixel mining approaches may be applicable to map settlement patterns in sparsely-populated regions.

Spatial data has also been used in epidemiological research concerning northern regions. Pardhan-Ali et al. (2012) employed spatio-temporal cluster analysis to study notifiable gastrointestinal illness in the Northwest Territories from 1991-2008. The researchers note that NWT provides a unique study area as most communities have fewer than 1,000 residents, and are only accessible by air or ice roads. Additionally, cluster analysis is made more challenging due to the large geographic area and sparse population distribution typical of northern communities, and results may be influenced by underreporting or underdiagnosis in the north.

Incorporating Indigenous Knowledge and Data Sovereignty in Spatial Analysis

It must be acknowledged that the Canadian census is an inherently colonial practice. Given that Indigenous peoples make up the largest shares of the population in northern Canada (Statistics Canada, 2021) and the federal government's history of using data to perpetuate anti-Indigenous stereotypes and justify colonial policies (Laboucan, 2024), it is imperative that work conducted in northern regions concerning Indigenous peoples and their data considers Indigenous Data Sovereignty and Indigenous Knowledge. Indigenous Data Sovereignty (IDS) refers to the ability for First Nations to have control over how data is collected about them, and that they control how this information is used or distributed (First Nations Information Governance Centre). Indigenous Knowledge (IK) refers to the complex knowledge systems based on the worldviews of Indigenous Peoples.

There are specific challenges encountered when integrating Indigenous Knowledge with spatial data. Primarily, Indigenous knowledge of “place” is not well-suited to geospatial technologies, like satellite imaging (Briggs et al., 2020). Most geospatial tools are better suited to data about space, rather than developing an understanding of place. Indigenous people’s knowledge of place often reflects the cultural and environmental context of Indigenous communities. In spite of the place-space gap, Briggs et al. (2020) identify areas in which Indigenous Knowledge can potentially be integrated into geographic information science. One method is through “volunteered geographic information”, wherein private citizens create and make publicly available geographic information, exemplified on platforms like OpenStreetMap, which is prevalent in population estimation techniques that integrate building footprints. These data sources have the potential to empower marginalized citizens and capture aspects of knowledge of “place”.

There is a growing body of work exploring how IK can be integrated with GIS tools and technologies. Mackenzie et al. (2017) explores these efforts, such as maps that accurately depict Indigenous conceptions of landforms. The authors discuss various challenges encountered in this integration, such as reconciling the GIS tendency for fixed, precise boundaries with more relative, fluid boundaries in some Indigenous areas. A more seamless integration may involve mapping migratory populations, such as herds of animals. Gagnon et al. (2020) integrates data about caribou migration from an Indigenous community-based monitoring program with satellite-based climate data to understand how environmental factors affect caribou health and growth. The authors found that incorporating Indigenous knowledge remarkably improved ecological modelling.

A common error in efforts to integrate IK with western geospatial science is the oversimplification of similarities, overemphasis of differences, and entrenchment of false dichotomies (Briggs et al., 2020). Specifically, given that IK is often encoded and practiced in Indigenous languages but research projects are often conducted in English, IK becomes oversimplified (Cannon et al., 2024), undermining the acquisition of accurate Indigenous data. Further, Indigenous data analyzed solely with western methodologies may produce deficit framings (Cannon et al., 2024). By ensuring Indigenous researchers are valuable collaborators in projects concerning Indigenous data, we may reduce bias in research design -- though non-Indigenous partners should be wary of burdens on knowledge-holders. Non-Indigenous researchers who anticipate working with Indigenous data or collaborating with Indigenous partners are advised to take the time to read about IDS and familiarize themselves with tools for supporting IDS in Canada, like OCAP (FNIGC).

References

Ahn, D., Yang, J., Cha, M., Yang, H., Kim, J., Park, S., Han, S., Lee, E., Lee, S., & Park, S. (2023). A human-machine collaborative approach measures economic development using satellite imagery. *Nature Communications*, *14*(1), 6811. https://doi.org/10.1038/s41467-023-42122-8

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

An update on the socio-economic gaps between Indigenous Peoples and the non-Indigenous population in Canada: Highlights from the 2021 Census. (2021). Statistics Canada. https://www.sac-isc.gc.ca/eng/1690909773300/1690909797208

Bartsch, A., Strozzi, T., & Nitze, I. (2023). Permafrost Monitoring from Space. *Surveys in Geophysics*, *44*(5), 1579–1613. https://doi.org/10.1007/s10712-023-09770-3

Bondarenko, M., Kerr, D., Sorichetta, A., & Tatem, A. J. (2020). *Census/projection-disaggregated gridded population datasets for 189 countries in 2020 using Built-Settlement Growth Model (BSGM) outputs.* [Dataset]. https://doi.org/10.5258/SOTON/WP00684

Boo, G., Darin, E., Leasure, D. R., Dooley, C. A., Chamberlain, H. R., Lázár, 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. *Nature Communications*, *13*(1), 1330. https://doi.org/10.1038/s41467-022-29094-x

Briggs, C., Burfurd, I., Duckham, M., Guntarik, O., Kerr, D., McMillan, M., & San Martin Saldias, D. (2020). Bridging the geospatial gap: Data about space and indigenous knowledge of place. *Geography Compass*, *14*(11), e12542. https://doi.org/10.1111/gec3.12542

Canada's Indigenous peoples increasingly at risk of disaster displacement. (n.d.). IDMC - Internal Displacement Monitoring Centre. Retrieved November 7, 2024, from https://www.internal-displacement.org/expert-analysis/canadas-indigenous-peoples-increasingly-at-risk-of-disaster-displacement

Chambers, K., Corbett, J., Keller, C., & Wood, C. (2004). Indigenous Knowledge, Mapping, and GIS: A Diffusion of Innovation Perspective. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *39*(3), 19–31. https://doi.org/10.3138/N752-N693-180T-N843

Conservation and Society. (n.d.). Retrieved November 6, 2024, from https://journals.lww.com/coas/fulltext/2017/15010/spatio_temporal_visualisation_and_data_exploration.5.aspx

Darin, E., Dicko, A. H., Galal, H., Jimenez, R. M., Park, H., Tatem, A. J., & Qader, S. (2024). Mapping refugee populations at high resolution by unlocking humanitarian administrative data. *Journal of International Humanitarian Action*, *9*(1). https://ora.ox.ac.uk/objects/uuid:9044b83d-b914-452e-83ea-58808afd0382

Detecting long-term changes to vegetation in—ProQuest. (n.d.). Retrieved November 6, 2024, from https://www.proquest.com/docview/2551247814?sourcetype=Scholarly%20Journals

Dittakan, K., Coenen, F., Christley, R., & Wardeh, M. (2013). Population Estimation Mining Using Satellite Imagery. In L. Bellatreche & M. K. Mohania (Eds.), *Data Warehousing and Knowledge Discovery* (Vol. 8057, pp. 285–296). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40131-2_25

Gagnon, C. A., Hamel, S., Russell, D. E., Powell, T., Andre, J., Svoboda, M. Y., & Berteaux, D. (2020). Merging indigenous and scientific knowledge links climate with the growth of a large migratory caribou population. *Journal of Applied Ecology*, *57*(9), 1644–1655. https://doi.org/10.1111/1365-2664.13558

Google Colab. (n.d.). Retrieved November 7, 2024, from https://colab.research.google.com/drive/11Zc3ejIprdWFhcp6Ix8QiY0yxLrkOlha#scrollTo=yay4BS-j-Kzg

Hillson, R., Coates, A., Alejandre, J. D., Jacobsen, K. H., Ansumana, R., Bockarie, A. S., Bangura, U., Lamin, J. M., & Stenger, D. A. (2019). Estimating the size of urban populations using Landsat images: A case study of Bo, Sierra Leone, West Africa. *International Journal of Health Geographics*, *18*(1), 16. https://doi.org/10.1186/s12942-019-0180-1

Indigenous Knowledge, Mapping, and GIS: A Diffusion of Innovation Perspective | Cartographica. (n.d.). Retrieved November 6, 2024, from https://utppublishing.com/doi/10.3138/N752-N693-180T-N843

Jochem, W. C., Bird, T. J., & Tatem, A. J. (2018). Identifying residential neighbourhood types from settlement points in a machine learning approach. *Computers, Environment and Urban Systems*, *69*, 104–113. https://doi.org/10.1016/j.compenvurbsys.2018.01.004

Laboucan, A. (2024). Https://beyond.ubc.ca/what-is-indigenous-data-sovereignty-and-why-does-it-matter/. *The University of British Columbia*.

Leasure, D., Jochem, W., Weber, E., Seaman, V., & Tatem, A. (n.d.). National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty. *PNAS*. https://doi.org/10.1073/pnas.1913050117

Mackenzie, K., Siabato, W., Reitsma, F., & Claramunt, C. (2017). Spatio-temporal Visualisation and Data Exploration of Traditional Ecological Knowledge/Indigenous Knowledge. *Conservation and Society*, *15*(1), 41. https://doi.org/10.4103/0972-4923.201391

Mapping surface water dynamics (1985–2021) in the Hudson Bay Lowlands, Canada using sub-pixel Landsat analysis—ScienceDirect. (n.d.). Retrieved November 6, 2024, from https://www.sciencedirect.com/science/article/pii/S0034425723004467

Mapping vegetation height and identifying the northern forest limit across Canada using ICESat-2, Landsat time series and topographic data—ScienceDirect. (n.d.). Retrieved November 6, 2024, from https://www.sciencedirect.com/science/article/pii/S0034425724001081

Neal, I., Seth, S., Watmough, G., & Diallo, M. S. (2022). Census-independent population estimation using representation learning. *Scientific Reports*, *12*(1), 5185. https://doi.org/10.1038/s41598-022-08935-1

Nilsen, K., Tejedor-Garavito, N., Leasure, D. R., Utazi, C. E., Ruktanonchai, C. W., Wigley, A. S., Dooley, C. A., Matthews, Z., & Tatem, A. J. (2021). A review of geospatial methods for population estimation and their use in constructing reproductive, maternal, newborn, child and adolescent health service indicators. *BMC Health Services Research*, *21*(1), 370. https://doi.org/10.1186/s12913-021-06370-y

Pardhan-Ali, A., Berke, O., Wilson, J., Edge, V. L., Furgal, C., Reid-Smith, R., Santos, M., & McEwen, S. A. (2012). A spatial and temporal analysis of notifiable gastrointestinal illness in the Northwest Territories, Canada, 1991-2008. *International Journal of Health Geographics*, *11*(1), 17. https://doi.org/10.1186/1476-072X-11-17

(PDF) Detecting long-term changes to vegetation in northern Canada using the Landsat satellite image archive. (n.d.). Retrieved November 6, 2024, from https://www.researchgate.net/publication/231097802_Detecting_long-term_changes_to_vegetation_in_northern_Canada_using_the_Landsat_satellite_image_archive

Robinson, C., Hohman, F., & Dilkina, B. (2017). A Deep Learning Approach for Population Estimation from Satellite Imagery. *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*, 47–54. https://doi.org/10.1145/3149858.3149863

Saku, J. (1999). ABORIGINAL CENSUS DATA IN CANADA: A RESEARCH NOTE. *The Canadian Journal of Native Studies*, *XIX*(2(1999)), 365–379.

Spatial and temporal persistence of nearshore kelp beds on the west coast of British Columbia, Canada using satellite remote sensing—Schroeder—2020—Remote Sensing in Ecology and Conservation—Wiley Online Library. (n.d.). Retrieved November 6, 2024, from https://zslpublications.onlinelibrary.wiley.com/doi/full/10.1002/rse2.142

Taking care of knowledge, taking care of salmon: Towards Indigenous data sovereignty in an era of climate change and cumulative effects. (n.d.). Retrieved November 6, 2024, from https://www.facetsjournal.com/doi/10.1139/facets-2023-0135

Travers-Smith, H., Coops, N. C., Mulverhill, C., Wulder, M. A., Ignace, D., & Lantz, T. C. (2024). Mapping vegetation height and identifying the northern forest limit across Canada using ICESat-2, Landsat time series and topographic data. *Remote Sensing of Environment*, *305*, 114097. https://doi.org/10.1016/j.rse.2024.114097

Understanding and Strengthening Indigenous Communities using GIS. (n.d.). Retrieved November 6, 2024, from https://resources.esri.ca/education-and-research/understanding-and-strengthening-indigenous-communities-using-gis

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

Yagoub, M. M., Tesfaldet, Y. T., AlSumaiti, T., Al Hosani, N., & Elmubarak, M. G. (2024). Estimating population density using open-access satellite images and geographic information system: Case of Al Ain city, UAE. *Remote Sensing Applications: Society and Environment*, *33*, 101122. https://doi.org/10.1016/j.rsase.2023.101122

Young, T. K., & Bjerregaard, P. (n.d.). Towards estimating the indigenous population in circumpolar regions. *International Journal of Circumpolar Health*, *78*(1), 1653749. https://doi.org/10.1080/22423982.2019.1653749