Do Remote Sensing Population Datasets Provide a Realistic Picture of the Spatial Distribution of Populations in the Arctic? Empirical Comparison with Data from the Swedish Population Register.

Author: Thibaud Bureau du Colombier [1, 2] ✉

[1] Université de Strasbourg, France

[2] Ecole des Hautes Etudes en Démographie (EHED), France

✉ t.bureau@unistra.fr

Abstract

Populations living in the Arctic are subject to increasing risks to their lifestyles and health due to climate change and economic development. However, from the definition of the region itself to the renewed risk of not being able to access quality demographic data, it is difficult to study their responses to current challenges. Thus, it is necessary to find alternatives.
Global population datasets built using remote sensing covariates such as land use, night lighting or water bodies, aim to provide a realistic spatial distribution of population and population change over time globally. These promising datasets could enable the development of spatial population distribution perspective models, improving our understanding of past and future demographic trends and enabling the development of public policies, particularly in contexts where we lack data.
Using Sweden's highly reliable register-based grid data as a reference of the spatial distribution of its population, we aim to identify how close three of the main population gridded datasets are from the reference and quantify this distance.
Our results highlight the improvements made in these grids in the last few years. We detail their performances in identifying population distribution for the years 2015 and 2020, and for population change in 2015-2020 period. Finally, we discuss their fitness for use in the Arctic context.

Keywords: Arctic, Gridded population datasets, Population distribution, Population estimates, Sweden

Introduction

The Arctic has long been neglected by demographers and population experts. In the collective imagination, it remains for many a cold, icy and therefore inhospitable place for human life, where only a few individuals survive. It is difficult to quantify exactly the population living in the Arctic, whether indigenous or non-indigenous, but one thing is certain: it is populated by several million individuals, about whom we still know very little.

One of the best examples of our lack of knowledge, which is at the same time a source of explanations for the difficulty of estimating the population of the Arctic, lies in the definition of this territory. Even today, the Arctic is very loosely defined as the region around the North Pole, in and around the Arctic Circle. It is a land and sea area that includes the territories of eight countries that are members of the Arctic Council, founded in 1996 by the Ottawa Declaration: Canada, Denmark, Finland, Iceland, Norway, Russia, Sweden and the United States. However, the boundary remains unclear as to which region(s) of their territories should be considered part of this area.

For this reason, numerous delimitations have been created and are used today to define the Arctic [Vaguet, 2021]. These include geophysical boundaries such as the 10°C isotherm in July, the southern limit of permafrost and the limit of tree flora. The main problem with these geophysical delimitations lies in their variability over time. The same applies to Louis-Edmond Hamelin's Nordicity index [Hamelin, 1968], which considers environmental constraints and adds indicators of geographical position (latitude) and human desert.

What is indirectly assumed behind these limits is geographical disparity but stability of conditions over time, whether in terms of environmental constraints or human settlement. This is tantamount to considering the populations of the Arctic as stable and/or immobile, thereby reducing the interest in studying them in population studies.

To remedy this problem, various Arctic Council working groups have defined their own boundaries for what should be considered the Arctic. Examples include the Conservation of Arctic Flora and Fauna, the Arctic Monitoring and Assessment Programme, the Emergency Prevention, Preparedness and Response, and the Arctic Human Development Report (AHDR) [Einarsson et al. 2004].

In the past, for political reasons or because of the difficulty of collecting data, the Arctic territories were less quickly included in censuses and registers than the more southerly provinces, both in North America and Eurasia [Dolson, 2010]. This differentiation is no longer relevant today, except for certain surveys such as the Canadian Community Health Survey.

The challenge now lies in harmonising data at supranational level, to treat the Arctic as a region, as called for by the Arctic Council, to identify the specific characteristics of these populations.

However, to my knowledge, no comprehensive surveys were ever conducted on Arctic peoples cross-border. Studies working on the evaluation of Arctic population counts often use censuses and register data. Their results are consistent with one another and vary mainly from the use of various delimitation of what should be considered the Arctic [Heleniak, 2021; Smirnov, 2020]. In 2015, Larsen and Fondhal updated the AHDR and estimated the population of the Arctic region according to this definition at over 4 million people in 2013, a far cry from the few hundred or thousand people cut off from the world in the collective imagination. For their part, Ramage and colleagues (2021) estimated the population living in the Arctic Circumpolar Permafrost Region (ACPR) at nearly 5 million individuals, including 1 million living on the coast, based on data from national statistical offices in 2016-2017. Using broader definitions and based on the administrative boundaries of individual countries, the Arctic population can also be estimated at 8.9 million inhabitants in 2001, rising to 9.0 million twenty years later [Bureau du Colombier, 2022]. This apparent demographic stability masks major geographical disparities in the evolution of its

population and their characteristics. In particular, there are still today considerable gaps in knowledge in assessing counts of the variety of indigenous people in the Arctic [Young and Bjerregaard, 2019; Burtseva et al., 2019], and thus their diversity "in population size and composition, growth rates, settlement patterns, and economic structures" [Emelyanova, 2022].
The relative lack of knowledge we have on the Arctic and its peoples, compared with the other territories of the Global North, should be reduced in the light of the region's current challenges. Indeed, the climate of the Arctic is undergoing a change that is unprecedented compared to the rest of the world, leading to a change in the potential for settlement in the entire area. In particular, the increase in surface temperature is expected to be almost three times higher than that of the rest of the globe between 2000 and 2100 [IPCC, 2022]: this is the phenomenon of Arctic amplification [Serreze and Barry, 2006].
This particularity provides a new prism through which to focus on Arctic populations, whether indigenous or not, who will be subject to increasing risks to their lifestyles and health (displacement of wildlife towards the north, risks of floods and droughts, soil destabilization), or, more broadly, on changes in the settlement of the region. An increasing number of observers are also looking to the Arctic for the economic opportunities it promises.
Some studies have developed in the last decade to study the response of Arctic populations to the climatic, economic and political challenges they face on a local scale [Chi et al., 2024], but we lack large-scale studies (and data) capable of informing us about the region as a whole

In Northern Europe, the Nordic Council aims to encourage scientific collaboration between its member states. However, this impetus is hampered, in particular but not exclusively by the difficulty of accessing Russian data. This has been particularly true since the beginning of the invasion of Ukraine by the Russian authorities, which highlights the dependence of academic research on international agreements, particularly in certain regions of the world, or in certain political contexts that are not conducive to academic research on climate change topics.
Faced with the need to study individuals who have been left out of demographic and population studies for too long, especially in view of the environmental changes occurring on their territories and faced with growing uncertainty over access to certain data, it is necessary to find alternatives for studying these populations globally.

Population datasets built from remote sensing data could provide a solution. They have been developing significantly over the past few decades [Archila Bustos et al., 2020]. They are a great tool to examine population distribution based on several covariates such as land cover, built structures, night-time lights or water bodies for example. As of June 2025, we listed eleven major open access global gridded population datasets: HYDE (history database of the global environment) [Klein Goldewijk et al., 2010], GDGPS (global dataset of gridded population and GDP scenarios) [Murakami and Yamagata, 2016], GHS-POP (global human settlement population) [JRC, 2024], GlobPOP (global gridded population dataset) [Liu et al., 2024], GPWv4 (gridded population of the world, version 4) [Doxsey-Whitfield et al., 2015], GRUMP (global rural-urban mapping project) [Balk, 2009], HRSL (high resolution settlement layer) [Tiecke et al., 2017], Kontur [Kontur, 2020], LandScan Global [Dobson et al., 2000], WorldPop [Tatem et al., 2017] and WPE (world population estimate) [Frye et al., 2018]. These datasets are largely differing in the complexity of their methodology, but they all have in common to create products distributing population counts or density on grid cells across the globe.
These promising datasets are enabling researchers to better understand the evolution of population distribution in the past and its mechanisms in different contexts, especially remote communities for which we lack population data, in the Arctic but also elsewhere in the world. This knowledge would enable the development of spatial population distribution perspective models, improving our understanding of past and future demographic trends and enabling the

development of public policies. Some studies are beginning to use population grids in the Arctic, notably in Sweden, where Karagiorgos and colleagues (2024) sought to specifically assess the reliability of population grids regarding flood exposure settlements. Thus, understanding how well remote sensing-based population grids represent the spatial distribution and the evolution of population over an entire country with reliable data, such as Sweden, is key for their development in population studies. This work is included in a global research trend of assessing and comparing remote sensing-based population grids in various contexts, whether in rural settings [Láng-Ritter et al., 2025], low- and middle-income countries [Thomson et al., 2022], fast growing population regions [Yin et al., 2021] or high-income countries [Archila Bustos et al., 2020].

Data and Methods

The aim of this study is to assess the quality of open access population grids using remote sensing data, compared with reference data from a reliable source, the Swedish population register. This work involves both qualifying the proximity between the population numbers for the years 2015 and 2020 from our different sources, and assessing the change in population between these two dates. The latter objective is in line with studying populations and their movements over time.
We selected three of the main population grid open access datasets, which are the Global Human Settlement Population Grid (GHS-POP), from JRC and CIESIN, LandScan Global from ORNL, and WorldPop gridded population counts. These grids, which are among the most highly modelled, are the most widely used in academic research [Láng-Ritter et al., 2025]. They have been developed with a view to being used as tools for public policy [ORNL, n.d.; WorldPop, n.d.; Schiavina et al., 2023].

The geospatial datasets are included in a 'top-down' modelling framework, as opposed to 'bottom-up'. Bottom-up methods use estimated total number of people per grid-cell to get count of people for sample locations. These methods are mainly used for countries where there is a lack of reliable and/or updated data [WorldPop, n.d]. On the other hand, top-down methods are used to study global coverage, over an entire country or region such as the Arctic, and changing population over time. From count of people per administrative unit, top-down methods estimate total number of people per grid-cell, with total population matching official counts.
This data can be accessed in the form of spatial raster datasets that depict the distribution of population, expressed as the number of people per cell. To obtain it, global population datasets providers use total population counts from varying sources. Then, they build a model using a remote sensing-based global data modelling and mapping approach to identify where people are more likely to reside (or live), and which results in creating a 'weighted' population density grid. Finally, this grid is used to project total population counts on the cells.

The GHS-POP, LandScan and WorldPop grids we used in this study are downloadable freely for academic purposes either in 3 arc-seconds or 30 arc-seconds in WGS84 coordinate system. It corresponds to approximately 100m and 1km respectively at the equator in Mollweide coordinates. This work is a continuation of that carried out by Archila Bustos et al. (2020) in Sweden and aims to update their results and to identify, if necessary, the potential causes of the differences obtained between our results. Indeed, updates in candidate grid cells may not lead to consistent results over time and past studies comparing datasets may need to be updated regularly to maintain their relevance for current and future use of population grids. For this reason, we will use names and methods like those used in their work in order to facilitate comparison between our results. Thus, the three gridded population datasets will hereafter be called 'candidate population grids', while

the reference will be called the 'known population grid'. A final aim is to identify a best fitting resolution, if any, that is reducing approximations for each data source.
We will now briefly present the different candidate and known population grids, and then detail the methods used to compare them. For a more complete description of existing candidate population grids, see Leyk et al. (2019), Archila Bustos et al. (2020), Láng-Ritter et al. (2025), and POPGRID Data Collaborative (2023).

Statistics Sweden register-based population grid

Sweden has a wide range of high-quality demographic data. One of these is the Swedish Total Population Register, a yearly updated database containing a wealth of socio-demographic information on Sweden's inhabitants, whether Swedish or non-Swedish citizens with a residence permit of at least one year. In addition to usual socio-demographic information, the register includes the place of residence of each individual, in the form of geographical coordinates. Having the location of its population makes possible for Statistics Sweden (SCB), the national population statistics office of Sweden, to produce register-based population grids, on a yearly basis. To generate the grids, each individual in the population register is assigned to its place of residence. Then, the persons in a 1 x 1km cell are summed up to obtain the number of inhabitants per cell. Before making this data available, Statistics Sweden anonymises it so that individuals cannot be identified. This consists of assigning a value of 0 or 3 for all cells with a population equal to 1 or 2. In this way, the final downloadable product contains 1 x 1km cells for all populated areas of the country. A value of zero in an existing cell therefore indicates a cell originally populated by one or two individuals. We identify the bias as minor since most people in Sweden, even in the Arctic, are concentrated in urban areas, which reduces the number of cells involved. On the other hand, the grid does not cover the whole of Sweden, since only cells where an individual's place of residence has been identified are considered to be populated and are thus created [SCB, n.d.]. Archila Bustos and colleagues (2020) compared official population counts from Statistics Sweden to the register-based grids in 1990, 1995, 2000, 2005, 2010 and 2015, for a resolution of 100 x 100m, and concluded the grid data were highly accurate, with less than 0.1% of the population missing each year on average.
Since the beginning of 2025, the Swedish register-based population grids have been available free of charge for each of the years from 2015 to 2024. They are produced yearly with a resolution of 1 x 1km, in EPSG:3006 – SWEREF99 coordinate system. They consist of three different layers: one is the total population in each grid cell, then the population by five-years age range in each grid cell, and finally the population by sex in each grid cell. Academics and students may obtain the layers for previous years, up until 1980, or for a 100 x 100m resolution, in return for a commission fee asked by Statistics Sweden.

For this study, in order to compare the register-based grids with remote sensing datasets, we are using Sweden's total population layers for the years 2015 and 2020, in the 1 x 1km resolution. We downloaded the vector files on 16/04/2025 and assigned the known population in each grid cell to its centroid for later treatments.

GHS-POP

The Global Human Settlement Population Grid (GHS-POP), from the European Commission's Joint Research Centre (JRC) is a global multi-temporal population grid. It is one of the many data produced by the Global Human Settlement Layer (GHSL) [Schiavina et al., 2023]. It combines population estimates from finest administrative units, derived from the Gridded Population of the

World (GPW) project, and spatial extents of human settlements to produce spatial raster datasets available between 1975 and 2030 in 5-year steps [Doxsey-Whitfield et al., 2015]. GHS-POP products are using a so-called ‘constrained’ method to spread the population. Unlike the ‘unconstrained’ method, where the assumption is made that no settlement dataset is accurate enough to identify all residential buildings globally and therefore be used as a mask to map uninhabited areas, the ‘constrained’ method projects population only in cells where residential buildings have been observed from remote sensing imagery. Because their original constrained model was solely based on the presence of built structures [Freire et al., 2016], using a dasymetric mapping approach, GHS-POP was considered a ‘lightly modeled’ population grid [POPGRID, 2023]. But the model has been complexified and improved in several ways in the last couple of years, including the use of built-up volume maps, systematic improvement of census coastlines, or integration of non-residential built-up volume information [JRC, 2024]. Therefore, the latest release (R2023A) introduced residential built-up volume as a predictor for population downscaling, thereby depicting population place of residence [Pesaresi et al, 2024]. These improvements should rise the percentage of cells correctly identified as populated or unpopulated. GHS-POP products are freely accessible for spatial resolutions of 3 arc-seconds and 30 arc-seconds in the WGS84 projection system, as well as 100m and 1km equivalent in Mollweide system.

We conducted our work using the latest version of GHS-POP (GHS-POP R2023A). We downloaded the raster files on 16/04/2025 for the tiles corresponding to Sweden’s extent (R3_C20, R3_C21 and R4_C20) in 30 arc-seconds (WGS84), for 2015 and 2020. For each year, we merged the three rasters into one global. We then clipped the global raster with the administrative boundaries of Sweden, obtained from the GADM Database [GADM, 2015], but we obtained a slight underestimation (1-2 %) of the overall Swedish population. We chose to use the unclipped global raster (1-2 % overestimation) since it was reducing the error percentage in our comparison statistics.

LandScan

The LandScan Global population distribution is part of the LandScan Program, initiated at Oak Ridge National Laboratory (ORNL) and developed on behalf of the U.S. federal government. It uses spatial data, high-resolution imagery exploitation, and a multi-variable dasymetric modelling approach to disaggregate census counts normalised to the CIA World Factbook within an administrative boundary [ORNL, n.d.; POPGRID, 2023]. Variables input included in the model are land cover, roads, built structures, a dichotomy variable to distinguish urban areas, infrastructure, environmental data, protected areas and water bodies [Mesev, 2003]. The dataset is developed annually and made publicly available through ORNL’s LandScan Portal, in 30 arc-seconds spatial resolution (WGS84), from 2000 to 2023. LandScan Global methodology is improved every year by using the highest quality input data available, giving better estimates every year [Lebakula et al., 2025]. However, previous products are not updated, which makes cell comparisons over time less – if at all – relevant. Unlike many population grids which estimate the resident population, LandScan Global models the ambient population, i.e. the 24-hour average [Dobson et al., 2000], with a view to capturing “the full potential activity space of people throughout the course of the day and night” [ORNL, n.d.]. This choice has been made to better address where people are when a natural or manmade disaster occurs. But it also leads to greater uncertainty when it comes to assessing the intrinsic quality of the data in relation to a reference, usually ‘resident’, population. For example, with this difference in definition alone, the ambient population in city centres is expected to be higher than the residential population, due to employment dynamics [Dobson et al., 2000].

We downloaded the 2015 and 2020 LandScan Global ambient population raster grids on 16/04/2025. The global grids were then clipped to the administrative boundaries of Sweden.

WorldPop

The WorldPop research programme, initiated in 2013 from the University of Southampton, produces data on population distributions and other demographic characteristics at high spatial resolution [WorldPop, n.d.]. It results from the combination of AfriPop, AsiaPop and AmeriPop projects and is now producing, among other things, population estimates with age and sex breakdowns for each 100 x 100m grid square globally [Lloyd et al., 2017]. WorldPop products are considered being one – if the most – highly modelled population grid [POPGRID, 2023]. They are using country-official population estimates as well as UNPD estimates to produce two grid types. Their model includes a bunch of geospatial covariates, from sources like, in the example of Sweden, ESA Climate Change Initiative (urban areas, woody vegetation, etc.), OpenStreetMap (major roads and waterways, road intersections), WorldClim (average time invariant temperature and precipitation), Viewfinder Panoramas (slope and elevation), Socio Economic Data and Applications Center, from CIESIN (distance to coastline), Visible Infrared Imaging Radiometer Suite (night lights), and the World Database on Protected Areas (distance to category 1 protected areas) [WorldPop, n.d., Lloyd et al., 2019]. However, inputs variables are not annual, which certainly provides an evolving distribution of the population over time, but which does not correspond exactly to any particular year. For example, the Swedish reference population was selected from SCB for the year 2010, and the exponential growth rate calculated for the period 2005-2010 to then be simply applied for the entire 2000-2020 period [Lloyd et al., 2019]. While it is common practice in gridded population datasets to have differences in total population at local or sub-national level, this also creates a difference at national level. In 2015, the difference between the reference population provided by SCB and that estimated by WorldPop was 0.4%, rising to 1.3% in 2020. The population density is then obtained from a flexible model algorithm using the Random Forest-based dasymetric approach developed by Stevens et al. (2015). Finally, the total population counts are projected onto the weighted grids in constrained and unconstrained datasets. But constrained datasets are only available in 2020, as opposed to unconstrained data which are annual time series between 2000 and 2020 available in 3 and 30 arc-seconds. Unconstrained datasets are therefore the ones we used, even if this method produces a non-zero allocation of population to all land grid cells, resulting in misallocation of population to uninhabited areas[1], and under-estimates of urban populations in some areas.

We retrieved the already clipped 1 x 1km unconstrained aggregated UN-adjusted population data for Sweden on 16/04/2025, for 2015 and 2020. We chose to work with the UN-adjusted grid because the final aim of this study is to possibly use these candidate data for all Arctic countries and territories, with the most homogenous data possible, the UN estimates fulfilling this role better.

Methodology

We created a virgin rectangular grid of 1km-by-1km cells, aligned with the known population grid, in EPSG:3006 – SWEREF99 TM (to create the grid with meter units, and not degree), using a geographic information system (QGIS). The cells which overlap (at least partially) with the

[1] Except water, missing land cover or areas outside census unit boundaries where the population density is defined as zero.

administrative boundaries of Sweden were selected by localisation, exported and reprojected to WGS84.
Each of the following steps have been done identically for 2015 and 2020. We calculated zonal statistics in each cell and for each candidate population raster grid and attributed their population sums to the virgin grid. We reprojected the layer in SWEREF99 to join by nearest localisation the attributes of the known population vector.
Our final 1 x 1km grid contains 460,068 cells and a total known population of 9,833,597 individuals in 2015 and 10,363,164 in 2020. Compared to SCB total population counts, we lost less than 0.2% of population in both year through anonymisation.
We then created virgin rectangular grids for smaller resolutions (2.5 x 2.5km, 5 x 5km, 10 x 10km and 25 x 25km cells, aligned with the known 1 x 1km population grid), and reproduced the firsts previous steps. To sum the values of 1 x 1km vector layer known grid into 5 x 5km, 10 x 10km and 25 x 25km cells, we 'joined the attributes by localisation (summary)'.
Finally, we exported the attribute tables to obtain, for each spatial resolution, a .csv file operable in R and containing for each cell the known and candidate population estimates in 2015 and 2020. The final smaller resolution grids contain respectively 75,035 cells (2.5 x 2.5km), 19,178 cells (5 x 5km), 4,976 cells (10 x 10km) and 864 cells (25 x 25km).

Comparison statistics

To measure the differences between the known and candidate population grids for each year we selected indicators commonly used in academic research and in particular by Archila Bustos et al. (2020), in order to better compare our results with theirs. The first statistics used are the percent mean absolute error (%MAE) and the percent root mean square error (%RMSE), which penalises extreme values more than %MAE. They are calculated in percent to allow for cross-dataset comparisons. Another comparison statistic is a measure of the linear association between the known and candidate grids, namely Pearson's R, which ranges between -1 (strong negative relation) and 1 (strong positive relation). To assess the gaps between our population grids differently, we used the percentage of cells correctly identified as populated and unpopulated. Candidate grid cells with population greater than zero were identified as populated, while cells with population equal to zero were considered unpopulated. For the reference population, each existing cell in the known population grid, whether it has a value of zero or more, was identified as populated. Cells that did not exist in the known population grid are the ones considered unpopulated. Then, as in Archila Bustos et al. (2020), "the percent correctly populated is the number of cells that were identified as populated in both the candidate and known datasets, divided by the number of known populated cells" (respectively for the unpopulated cells). Lastly, we calculated the relative difference between the known and candidate grids in each cell, which ranges between -1 (unpopulated in candidate but populated) and 1 (populated in candidate but unpopulated), and produced six associated maps of Sweden.

We then calculated the known population changes between 2015 and 2020. We identified three types of cells. *Growth* cells were selected when the population was higher in 2020 than in 2015, or when the cell was populated in 2020 and was not in 2015. *Stable* cells were selected when the population was the same in 2015 and 2020, or when the cell was unpopulated for those two years. *Decline* cells were selected when the population was lower in 2020 than in 2015, or when the cell was populated in 2015 and was not in 2020.
Then, for each cell type and each candidate population, we calculated their population change. We measured the linear association between the known population change and the candidate populate change, as well as the percentage of correctly growing, stable, and declining cells.

Lastly, we conducted two final analyses by calculating the same comparison statistics as described above for round values from the candidate population grid, and by selecting the closest rounded value in each cell. Rounding the candidate population estimates is a way to reduce the side effect of models using a non-zero allocation in rural areas, where cells having less than 0.5 inhabitants would count as zero. Finally, by selecting the nearest rounded value from the candidate grids, we establish a measure of how close (or far) a candidate grid combining the best of each dataset is to reality.

Hypotheses

Hypothesis 1:
LandScan is expected to perform less well in 2015 and 2020 in the 1 x 1km grid than other candidate population grids because it focuses on the "ambient population" and is compared with a reference "resident population".

Hypothesis 2:
GHS-POP should achieve significantly better results in our study than in that of Archila Bustos et al. (2020), following the improvements implemented since 2020 by the JRC.

Hypothesis 3:
Increasing cell size will proportionally increase comparative indicator scores for all candidate grids.

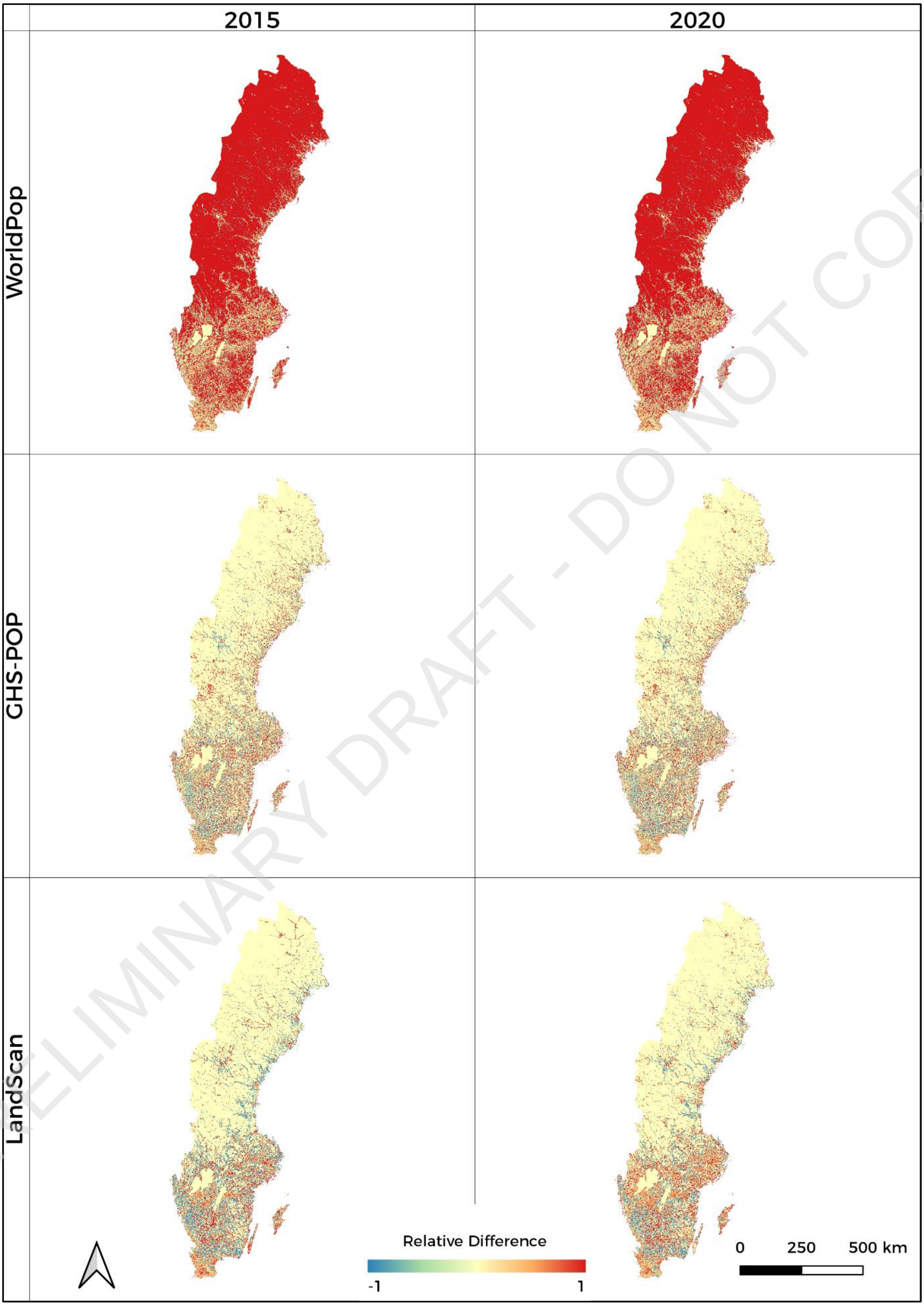
Hypothesis 4:
The comparison of population change between candidate grids and the known grid should be less accurate for LandScan (because the methodology changes each year) and WorldPop (because 2005-2010 population growth is projected from 2010), than for GHS-POP.

Hypothesis 5:
Assigning rounded values and in particular zero to almost-zero value cells should greatly improve WorldPop's performance in cell-by-cell comparisons.

Results

As we described previously, highly modelled candidate grids have major differences in their methodologies. One of them is to 'constrain' or not the distribution of the population only in urban extents and can be seen in Figure 1.

Figure 1. Relative difference between known and candidate population grids by year



Generally, all comparison statistics for the 1km² grid cells are improving for all datasets between 2015 and 2020, with the exception of the percentage of well-attributed unpopulated cells in LandScan (Table 1).

Table 1. Comparison statistics for 1km^2 cells by dataset and year

Dataset	Year	%MAE	%RMSE	Pearson's R	Percent correct		
					Populated	Unpopulated	Total
GHS-POP	2015	0.601	5.71	0.870	83.9	89.2	87.9
	2020	0.597	5.69	0.873	84.0	89.5	88.1
LandScan	2015	0.742	7.52	0.754	68.3	91.4	85.7
	2020	0.718	7.45	0.756	73.8	90.7	86.5
WorldPop	2015	0.699	6.18	0.837	99.9*	2.14	26.2
	2020	0.679	6.06	0.843	99.9*	2.15	26.5

** Truncated values. Reading note: 83.9% of 2015 populated cells were correctly identified by GHS-POP.*

In both years, LandScan's 1km² cell by cell comparison statistics have a lower performance than GHS-POP and WorldPop grids. Its %MAE and %RMSE are higher, meaning their absolute fit with the known grid is not as good. This is consistent with a lower linear association in LandScan. For each of these three indicators, and for both years of the study, GHS-POP obtained the best results, followed by WorldPop and then LandScan.

When looking at the cells correctly identified as populated, LandScan is also less performing, whereas GHS-POP obtain good results and WorldPop obtain almost perfect results with only 10 cells out of more than 110 000 which are not populated when they should have been. However, WorldPop has almost no unpopulated cells, leading to a very low total percent of well-attributed cells. LandScan has a better fit for unpopulated cells than any other grids, with more than 90%, just above GHS-POP 89%. In total, the latter is performing the best for a 1km² grid cells with a total percent of correct cells at 88%, but LandScan is very close. More than the attribution of populated or unpopulated cells (even if the former is lower than that of its competitors), it is the values inside the cells, especially the extreme values, that lower LandScan's precision.

The annual improvements of LandScan methodology may explain the increase in the percent correct populated cells (+5.5 percentage points) as well as the decline for unpopulated ones (-0.7 p.p.), leading to an overall improvement between 2015 and 2020. Focusing on ambient population rather than resident population is certainly the cause for having lower populated cells estimates than in other constrained grid (i.e. GHS-POP). We thus validate our first hypothesis. In GHS-POP and WorldPop, since no methodological improvements have been made between the 2015 and 2020 grids, improved statistics may be due either to a change in the distribution of the population across Sweden between 2015 and 2020, which fortuitously leads to an improvement in the fit between estimated and known data, or to an improvement in the prediction of 2020 input data compared with 2015 ones. With the information available to us, we cannot draw any conclusions on this point.

Our second hypothesis aims to test the gap between the results of currently available GHS-POP 2023 version and the 2019 one. Major improvements have been made between GHS-POP R2019A and R2023A products. The only indicator that is performing less well in the latter is the percentage of correctly identified unpopulated cells (-7.2 p.p., Table 2). This performance gap is largely offset by the enhancements of populated cells (+55.0 p.p.). All the other indicators point to greater proximity to the known population grid. Therefore, our second hypothesis is validated, and results obtained by Archila Bustos and colleagues (2020) may be updated for GHS-POP Global dataset.

Table 2. Comparison statistics for 1km² cells in GHS-POP 2015 by version

Version	%MAE	%RMSE	Pearson's R	Percent correct		
				Populated	Unpopulated	Total
GHS-POP R2019A	0.72	5.84	0.85	28.9	96.4	81.3
GHS-POP R2023A	0.60	5.71	0.87	83.9	89.2	87.9

GHS-POP is performing best for 1km² cells (see Table 1) after receiving improvements in its methodology, which has repercussions on the quality of its most recent products, for population grids of current (as would be the case for LandScan) and past years.
However, these improvements in the GHS-POP products due to the improvement in its methodology must be qualified in relation to the improvement achieved through the use of better input data. This feature is common to the WorldPop, LandScan and GHS-POP products, and therefore may lead to an improvement for each of these files.
Between our results and the ones calculated by Archila Bustos and colleagues (2020), the %MAE and the %RMSE are both decreasing (i.e. improving) for all 2015 candidate grids. The improvement of the former is three to four times better in GHS-POP than in LandScan and WorldPop grids. Pearson's R is also increasing more in GHS-POP (+0.02) than in WorldPop (+0.01) and LandScan (stable). However, the %RMSE is decreasing eight to nine times less in GHS-POP than in the other candidate grids but still remains the lowest overall. Finally, where GHS-POP massively improved its percent of correct populated cells, LandScan products see a diminution in this category (-1.4 p.p.) as well as in the total percent of correctly identified cells (-2.8 p.p.). In the case of WorldPop, while the percentage of populated cells correctly identified is similar in our work to previous one, and the percentage of non-populated cells decreases slightly (-0.8 p.p.), the percentage of cells correctly identified by WorldPop found by Archila Bustos and colleagues (2020) was 16%, whereas it reaches 26.5% in our case (+10.2 p.p.). After multiple checks, we were unable to explain this difference, our total percent being consistent with all our other produced results.

We noted earlier (see hypothesis 2) that the values obtained by our various indicators were better in 2020 than in 2015, especially for GHS-POP, but also for LandScan and WorldPop. This is still true only for 1x1km grids: for other resolutions, %MAE and %RMSE reflect a deterioration in the fit with the known population in 2020 compared with 2015 (Table 3, 3B and 3C).

Table 3A. Comparison statistics of GHS-POP by year and spatial resolution

Year	Resolution	%MAE	%RMSE	Pearson's R	Percent correct		
					Populated	Unpopulated	Total
	1 x 1km	0.601	5.71	0.870	83.9	89.2	87.9
	2.5 x 2.5km	0.466	3.83	0.913	90.5	80.5	84.9
2015	5 x 5km	0.204	1.08	0.990	93.7	69.0	84.3
	10 x 10km	0.109	0.484	0.997	96.3	54.3	87.1
	25 x 25km	0.124	1.59	0.944	98.7	20.8	91.8
	1 x 1km	0.597	5.69	0.873	84.0	89.5	88.1
	2.5 x 2.5km	0.471	4.02	0.907	91.0	80.6	85.2
2020	5 x 5km	0.215	1.25	0.988	94.3	68.7	84.3
	10 x 10km	0.124	0.592	0.997	96.4	54.2	87.0
	25 x 25km	0.137	1.60	0.947	98.8	20.7	91.4

Table 3B. Comparison statistics of LandScan by year and spatial resolution

Year	Resolution	%MAE	%RMSE	Pearson's R	Percent correct		
					Populated	Unpopulated	Total
2015	1 x 1km	0.742	7.52	0.754	68.3	91.4	85.7
	2.5 x 2.5km	0.562	4.24	0.886	75.8	90.7	84.1
	5 x 5km	0.341	2.08	0.962	79.5	88.9	83.1
	10 x 10km	0.222	1.04	0.986	86.0	83.6	85.5
	25 x 25km	0.105	0.308	0.999	95.4	76.6	93.8
2020	1 x 1km	0.718	7.45	0.756	73.8	90.7	86.5
	2.5 x 2.5km	0.552	4.26	0.886	80.7	87.9	84.7
	5 x 5km	0.345	2.15	0.961	84.4	81.8	83.4
	10 x 10km	0.230	1.06	0.986	90.3	67.8	85.3
	25 x 25km	0.124	0.355	0.999	98.6	35.4	92.6

Table 3C. Comparison statistics of WorldPop by year and spatial resolution

Year	Resolution	%MAE	%RMSE	Pearson's R	Percent correct		
					Populated	Unpopulated	Total
2015	1 x 1km	0.699	6.18	0.837	99.9*	2.14	26.2
	2.5 x 2.5km	0.500	3.82	0.909	99.9*	2.18	45.4
	5 x 5km	0.237	1.21	0.988	99.9*	2.33	62.6
	10 x 10km	0.121	0.450	0.997	100	2.38	78.6
	25 x 25km	0.049	0.119	0.999*	100	0	91.1
2020	1 x 1km	0.679	6.06	0.843	99.9*	2.15	26.5
	2.5 x 2.5km	0.492	3.94	0.905	99.9*	2.18	45.2
	5 x 5km	0.240	1.29	0.986	100	2.30	61.8
	10 x 10km	0.127	0.488	0.997	100	2.36	78.4
	25 x 25km	0.058	0.145	0.999*	100	0	90.5

Improvement in comparison statistics scores by resolution was tested by dividing the gain in each indicator by the ratio between two close resolutions. Constant improvement values would have led to conclude increase in cell size leads to a proportional improvement in comparison statistics scores. But this hypothesis is not conclusive for LandScan, WorldPop, and even less so for GHS-POP.

For GHS-POP, Table 3A shows an improvement in the first three indicators (%MAE, %RMSE and Pearson's R) only up to the 10x10km grid. The 25x25km grid obtained poorer results than the higher resolution grid, and poorer %RMSE and Pearson's R than the 5x5km grid. On the other hand, the percentage of correctly identified cells increased for the two least precise resolutions (as did LandScan, see Table 3B), indicating an improvement in the predictive power of these grids. However, this result is masked by the fact that almost all the cells become populated at low resolutions. Thus, the simple fact of declaring all 25x25km cells as populated (and therefore no non-populated cells) results in a total predictive percentage of more than 90% (see Table 3C). There is also a massive loss of accuracy in the predictability of GHS-POP non-populated areas between the 1x1km grid and the 25x25km grid, both in 2015 and 2020.

Table 3B provides extensive information on LandScan's performance. The changes in methodology between 2015 and 2020 have led, as we have seen in hypothesis 1, to a clear improvement in the recognition of populated cells for 1x1km resolution at the expense of a slight loss in unpopulated cells. However, for less precise resolutions, notably 25x25km in 2020, the identification of non-populated areas becomes poor (only 35.4%), which was not the case in 2015. Finally, it should be noted that none of the candidate datasets manages to correctly identify all the cells, even for a low resolution such as 25x25km.

Then, to test our fourth hypothesis, we compared the change of population in the cells of each candidate grids to the one in the cell of the known grid. We found in Table 4A that Pearson's R values are positive for all candidate grids but LandScan's one shows no correlation (<0.10), whereas GHS-POP and WorldPop values are weak (0.20-0.39). Also, when looking at the fit between the direction of population change (either growth, stable or decline), constrained datasets are performing way better than the unconstrained one.

Table 4A. Population change comparison statistics in 1x1km grid by dataset, between 2015 and 2020

Dataset	Pearson's R	Total percent correct
GHS-POP	0.387	77.3
LandScan	0.067	78.2
WorldPop	0.318	11.8

Moreover, in all three candidate datasets, we see a weak to very-weak but positive correlation between growing cells in candidate grids compared to the known grid (see Table 4B). However, all candidate datasets have also very-weak to weak negative correlation levels in declining cells. The difference in performance between GHS-POP and other candidate grids may be explained by the light overestimation of the total population we obtained by not clipping the raster since this result is consistent over resolution. Results in population change cells prediction confirm this finding with low to very low correct percentage in growing and declining candidate grid cells. Still, we find very good to excellent results for stable cells. This is mainly due to the large proportion of unpopulated cells in both years (>75%), and, for WorldPop, to a very low number of identified unpopulated cells. Yet, the few cells that WorldPop identifies as stable, because they are not populated, give excellent results.

Table 4B. Comparison statistics in 1x1km grid by dataset and population change, between 2015 and 2020

Dataset	Population change	Cell counts	Pearson's R	Total percent correct
	Growth	70296	0.399	35.2
GHS-POP	Stable	327212	NA	96.1
	Decline	62560	-0.225	24.5
	Growth	48950	0.133	33.2
LandScan	Stable	363855	NA	90.7
	Decline	47263	-0.159	28.2
	Growth	212601	0.364	11.8
WorldPop	Stable	7424	NA	99.8
	Decline	240043	-0.166	9.00

R.n. 35.2% of the growing cells of GHS-POP are indeed growing in the known grid between 2015 and 2020.

This hypothesis 4 is not validated. Indeed, even if it is difficult to separate the impact of the change in LandScan's methodology from the difference in definition (ambient versus resident), we do not identify a clear disadvantage in LandScan's performances compared with the other candidate grids. In fact, LandScan obtains the best (or less bad) results for declining cells, whereas differences between GHS-POP and WorldPop are mainly explained by their constrained/unconstrained intrinsec methodology gaps.

Our last hypothesis considered that assigning rounded values and in particular zero to almost-zero value cells should greatly improve WorldPop's performance in cell-by-cell comparisons. Indeed, this is vastly improving the prediction capacity of WorldPop products in 2015 and 2020 (Table 5). Although this reduces the performance of populated cells at the same time, the improvement in overall predictability is striking. This result is all the more interesting in that it coexists with a maintenance (or even a very slight improvement) in the other performance indicators. These figures could also be slightly improved by redistributing in populated cells the 22,986 and 25,005 individuals (for 2015 and 2020 respectively) that this manipulation has removed. They represent 0.23% and 0.25% of the population for these years.

Table 5. Difference between the performance of indicators in traditional population grids and those with rounded values. Comparison by year and dataset.

Year	Dataset	Difference in %MAE (p.p.)	Difference in %RMSE (p.p.)	Difference in Pearson's R values	Difference in correct percentage (p.p.)		
					Populated	Unpopulated	Total
2015	GHS-POP	-0.000	+0.000	+0.000	-7.21	+3.92	+1.18
	LandScan	-0.000	-0.000	+0.000	-0.049	+0.025	+0.007
	WorldPop	-0.002	-0.000	+0.000	-3.35	+53.4	+39.4
2020	GHS-POP	-0.000	+0.000	+0.000	-7.24	+3.92	+1.15
	LandScan	-0.000	+0.000	-0.000	-0.036	+0.022	+0.007
	WorldPop	-0.002	-0.001	+0.000	-3.20	+55.2	+40.7

R.n. Assigning rounded values increased the correct percentage of unpopulated cells in WorldPop by 53.4 percentage points in 2015. Shaded cells indicate an improvement in the indicator thanks to the rounded values.

For its part, LandScan obtains almost unchanged results, since the majority of the values in its cells are already integer values. Thus, in 2015, the equivalent of just 22 individuals (and 19 in 2020) were lost through this manipulation.

In contrast to WorldPop, the results of assigning rounded values for GHS-POP are mixed, with a slight improvement in total predictability and for non-populated cells, but a drop in predictability for populated cells. The loss of individuals was only 0.02% for both years.

If we look at the change in population between 2015 and 2020, we see almost no loss of information for Pearson's R by rounding the values and a very sharp improvement in the total percentage of correctly identified cells, not only for WorldPop but also for the other candidate grids. However, as we detailed in hypothesis 4, this good performance masks the difficulties in identifying decreasing cells. And, even if the results for these cells with rounded values are better than the original results, their correlation remains negative, whatever the dataset.

Finally, for each 1km² cell, we chose the candidate grid value closest to that of the known population. This allows us to assess how close this simulated 'best population grid' is to reality. As this grid was created using a known reference population, it is not intended to be replicated in other contexts. Table 6, 6B and 6C present our results.

Table 6A. Comparison statistics for nearest values in 1km² cells by year

Year	%MAE	%RMSE	Pearson's R	Percent correct		
				Populated	Unpopulated	Total
2015	0.333	3.74	0.944	91.8	96.4	95.3
2020	0.329	3.71	0.944	93.0	95.0	94.5

The %MAE, %RMSE and Pearson's R all show an excellent correlation between this grid and the reference, both in 2015 and 2020. The percentages of correctly identified cells, whether populated or not, are also very high. However, there is a slight drop in the predictability of our grid in 2020, in line with the trends observed at LandScan.

Table 6B. Comparison statistics for nearest values in 1km² cells between 2015 and 2020

Period	Pearson's R	Total percent correct
2015-2020	0.423	84.1

For the 2015-2020 study period, Pearson's R has a moderate value due to an average correlation (0.503) for cells with a population identified as increasing and an absence of correlation (0.029) for cells with a population identified as decreasing. However, the percentage of cells correctly identified as increasing and decreasing is almost identical (44-46%), which suggests that as well as not accurately identifying the cells undergoing a population change, it is the very value of these changes that is not well captured. Thus, the cells identified as decreasing describe a stable population on average.

Table 6C. Comparison statistics for nearest values 1km² cells by population change

Population change	Cell counts	Pearson's R	Total percent correct
Growth	67321	0.503	45.9
Stable	336105	NA	98.4
Decline	56642	0.029	44.6

Discussion

The great development of global gridded population datasets in the last decades have permit to, for example, estimate population at risk of coastal hazard and sea level rise, and mapping sub-national heterogeneities in health, wealth, and resource access [MacManus et al., 2021; Lloyd et al., 2019; Mondal and Tatem, 2012].

In order to continue to develop these products and ensure their widespread use in all settings and regions of the world, academic researchers have been looking at their performance. It is difficult to determine that one grid is 'the best' since none of them performs better than the others in all the tests we have carried out. We therefore reiterate the precautions formulated by Leyk and colleagues (2019) who, after a detailed examination of the methodology used for each grid, concluded that the choice of data grid by users would depend on their objective(s).

LandScan Global performed less well than GHS-POP and WorldPop overall, which is consistent with the results of Archila Bustos and colleagues (2020) who identified LandScan as the worst performing of the highly modelled grids in the comparative statistics they used for the annual analyses. But these discrepancies are due in particular to the type of population projected (i.e. ambient rather than resident). This point also constitutes the strength and uniqueness of LandScan, which may be a preferred option when one wants to be able to quantify where individuals are during daytime, for emergency response purposes for example.

For a study looking at urban/rural population, it is better to use a constrained grid where urban extent has been used in the modelling, such as GHS-POP. The constrained modelling may however overconcentrate population in urban extent, whereas unconstrained might underestimate it, e.g. WorldPop [Leyk et al., 2019]. Also, an endogeneity problem may arise when the candidate grid includes in its population projection model one or more variables which constitute the subject of study. This occurs for the study of urban extent with a constrained grid but can also occur for the study of the links between population and climate change with respect to climate input data.

Lately, population grids have gained significant improvements in their precision, being often available at 1x1km and even 100x100m cell size. In some of their applications, climate scenario

modelling for example, such precision is not necessarily needed, and lower resolution of general spatial population distribution may be preferable, especially if it provides a better assessment of the actual population size. Our results show that reconstructing these grids for smaller resolutions has a beneficial effect on the metrics we used. But future users should remember that this is not a systematic effect. When LandScan and GHS-POP statistics were the best for 25x25km grids, WorldPop was performing best at 10x10km. In addition, our results show that improvements in matching between two annual grids for a given resolution cannot be inferred to other resolutions.
Our study is one of the first to evaluate the fit between different grids in relation to a reference population, and then to compare these results with a similar study carried out with an earlier version of one of the grids (i.e. GHS-POP R2019A). Indeed, our results update those of Archila Bustos and colleagues (2020) using the most recent versions of the grids, both in terms of the data inputs used and the changes in methodology. These results show that studies comparing global gridded datasets are very important but have a very short period of validity. Another example is the work from Kuffer and colleagues (2022) that evaluated causes of uncertainty in GHS-POP products prior to the 2023 version. Thus, when releasing a new product version, updated with new methodologies, it would be precious to generalise its comparison not only with previous product(s) which these methodologies improve but also with other comparable population grids.
Our analyses concluded that progress have been made in recent years, allowing grids to get closer to a realistic population distribution for specific years (here 2015 and 2020). However, the results are less enthusiastic when we look at changes in population distribution over time. The grids we have studied do a poor job of identifying cells whose population is decreasing. These data are therefore not yet able to replace more traditional demographic data for longitudinal studies, particularly in specific contexts such as territories experiencing known demographic decline (e.g., Russian Arctic territories), or to specifically study population decline. However, it should be noted that these population grids offer an additional possibility when few data are available. In this case, for long-term population change, it is preferable to use GHS-POP, which obtains the best results using a non-extrapolated 'resident' population every five years.
Other structural problems with gridded population datasets have been identified in the literature. A lack of performance has been shown in low population density settings (i.e. <50 people/km²) [Archila Bustos et al., 2020]. Láng-Ritter and colleagues (2025) found massive and systematic underestimation of the rural population in gridded datasets for all tested countries. However, in Sweden and Finland, GHS-POP products were only tested for years before 1990, and we cannot say whether they used the 2023 version or an earlier version. Conversely, GHS-POP has been shown to outperform WorldPop but still overestimates flood exposure in Sweden [Karagiorgos et al., 2024].

Finally, Liu and colleagues (2024) have created a composite 31-year population grid to maximise the performance and precision of the three highly modelled population grids we analysed, as well as GPWv4 and GRUMP. They found that their dataset, GlobPOP, performs consistently better than the products they combined, in developed countries (Japan, Germany, USA and Portugal) and developing ones (China, Liberia, Guyana and Lebanon). There is a need to continue testing this dataset in order to develop its potential future use, particularly, but not only, because it combines datasets using different definitions and which originally have specific applications.
In addition to continuing to develop these comparative analyses in various contexts (urban, rural, LMICs, HICs, etc.) to improve the performance of existing grids across the globe and develop their applications for public policies, it is important to propose ways to expand their scope. Thus, it is necessary to extend the population distribution of the entire population to distribution in gender and age structures. Data on gender and age distribution from the Swedish register could also be used to test these products in future research.

Conclusion

With a view to finding alternatives to traditional demographic data for studying Arctic populations in a context of climate change and threats to academic research, we sought to examine the proximity of three of the main open-access population grids (namely GHS-POP, LandScan Global and WorldPop) to the Swedish total population register. The development of these global gridded datasets is a superb opportunity to study populations that are not easily accessible.

We draw readers' attention to the fact that the results obtained in this and in other studies are comparable only when the same versions of the products are used: their performance can vary greatly depending on their updates, and so can the results obtained by testing them. We remind that the choice of grid must be made according to one's object and field of study, since no grid performs better than the others for all our indicators; plus, their methodologies differ and therefore do not allow the same treatments. For use in the Arctic, we recommend using GHS-POP because it detects residential populated and non-populated areas best and obtains better results over the two years tested (2015 and 2020). However, these data would need to be improved for longitudinal use, in the Arctic as elsewhere, as their results remain mediocre, particularly in rural areas.

References

Archila Bustos, Maria Francisca, Ola Hall, Thomas Niedomysl, et Ulf Ernstson. « A pixel level evaluation of five multitemporal global gridded population datasets: a case study in Sweden, 1990–2015 ». Population and Environment 42, nº 2 (décembre 2020): 255-77. https://doi.org/10.1007/s11111-020-00360-8.

Balk, D. More than a name - why is global urban population mapping a GRUMPy proposition? In Global Mapping of Human Settlement (CRC Press, 2009).

Bureau du Colombier, Thibaud. « Etudier la vulnérabilité de la population arctique aux changements climatiques : une mesure impossible ? », 2022.

Bright, E., Rose, A., & Urban, M. (2016). *LandScan Global 2015* [Data set]. Oak Ridge National Laboratory. https://doi.org/10.48690/1524210

Burtseva, Evdokia I., M. K. Ammosov North-Eastern Federal University, Sleptsov, Anatoly N., M. K. Ammosov North-Eastern Federal University, Velichenko, Valery V., M. K. Ammosov North-Eastern Federal University, Potravny, Ivan M., Plekhanov Russian University of Economics, Gassiy, Voletta V., et Kuban State University. « Issues of estimating and compensating for losses to indigenous peoples in the conditions of industrial development of the Arctic ». Arctic: Ecology and Economy, nº 1(33) (5 mars 2019): 34-49. https://doi.org/10.25283/2223-4594-2019-1-34-49.

Chi, Guangqing, Shuai Zhou, Megan Mucioki, Jessica Miller, Ekrem Korkut, Lance Howe, Junjun Yin, et al. « Climate Impacts on Migration in the Arctic North America: Existing Evidence and Research Recommendations ». Regional Environmental Change 24, nº 2 (16 mars 2024): 47. https://doi.org/10.1007/s10113-024-02212-9.

Dobson, J., Bright, E., Coleman, P., Durfee, R. & Worley, B. LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Remote Sens. 66, 849–857 (2000).

Dolson, David. « Census Coverage Studies in Canada: A History with Emphasis on the 2011 Census », 2010.

Doxsey-Whitfield, E. et al. Taking advantage of the improved availability of census data: a first look at the gridded population of the World, Version 4. Pap. Appl. Geogr. 1, 226–234 (2015).

Einarsson, Níels, Joan Nymand Larsen, Annika Nilsson, et Oran R. Young. *Arctic Human Development Report*. Stefansson Arctic Institute, 2004. https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-32208.

Emelyanova, Anastasia. « The Arctic Region and Its Inhabitants ». In Arctic One Health: Challenges for Northern Animals and People, edited by Morten Tryland, 3-20. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-030-87853-5_1.

Freire, Sergio, Erin Doxsey-Whitfield, Kytt MacManus, Jane Mills, et Martino Pesaresi. « Development of New Open and Free Multi-Temporal Global Population Grids at 250 m Resolution », 2016.

Frye, C., Wright, D. J., Nordstrand, E., Terborgh, C., & Foust, J. (2018). Using classified and unclassified land cover data to estimate the footprint of human settlement. Data Science Journal, 17. https://doi-org.scd-rproxy.u-strasbg.fr/10.5334/dsj-2018-020.

GADM, version 2.5, July 2015. Consulté le 10 avril 2025. https://gadm.org/maps/SWE.html.

Hamelin, Louis-Edmond. « Un indice circumpolaire ». *Annales de Géographie* 77, n° 422 (1968): 414-30.

Heleniak, Timothy. « The future of the Arctic populations ». Polar Geography 44, n° 2 (3 avril 2021): 136-52. https://doi.org/10.1080/1088937X.2019.1707316.

Intergovernmental Panel On Climate Change (Ipcc). *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 1re éd. Cambridge University Press, 2023. https://doi.org/10.1017/9781009325844.

Joint Research Centre (JRC), European Commission, GHSL Data Package 2023, Publications Office of the European Union, Luxembourg, 2024, doi:10.2760/098587, JRC133256

Karagiorgos, K., Georganos, S., Fuchs, S. et al. Global population datasets overestimate flood exposure in Sweden. Sci Rep 14, 20410 (2024). https://doi.org/10.1038/s41598-024-71330-5

Klein Goldewijk, K., Beusen, A. & Janssen, P. Long-term dynamic modeling of global population and built-up area in a spatially explicit way: HYDE 3.1. Holocene 20,565–573 (2010).

Kontur Inc. Kontur Population dataset. Kontur Inc. https://www. kontur.io/portfolio/population-dataset/ (2020).

Kuffer, Monika, Maxwell Owusu, Lorraine Oliveira, Richard Sliuzas, and Frank van Rijn. 2022. "The Missing Millions in Maps: Exploring Causes of Uncertainties in Global Gridded Population Datasets" *ISPRS International Journal of Geo-Information* 11, no. 7: 403. https://doi.org/10.3390/ijgi11070403

Láng-Ritter, Josias, Marko Keskinen, et Henrikki Tenkanen. « Global Gridded Population Datasets Systematically Underrepresent Rural Population ». Nature Communications 16, n° 1 (18 mars 2025): 2170. https://doi.org/10.1038/s41467-025-56906-7.

Larsen, Joan Nymand, et Gail Fondahl. *Arctic Human Development Report: Regional Processes and Global Linkages*. Nordic Council of Ministers, 2015.

Lebakula, Viswadeep, Kelly Sims, Andrew Reith, Amy Rose, Jake McKee, Phil Coleman, Jason Kaufman, et al. « LandScan Global 30 Arcsecond Annual Global Gridded Population Datasets from 2000 to 2022 ». Scientific Data 12, nº 1 (24 mars 2025): 495. https://doi.org/10.1038/s41597-025-04817-z.

Leyk, Stefan, Andrea E. Gaughan, Susana B. Adamo, Alex de Sherbinin, Deborah Balk, Sergio Freire, Amy Rose, et al. « The Spatial Allocation of Population: A Review of Large-Scale Gridded Population Data Products and Their Fitness for Use ». *Earth System Science Data* 11, nº 3 (11 septembre 2019): 1385-1409. https://doi.org/10.5194/essd-11-1385-2019.

Liu, Luling, Xin Cao, Shijie Li, et Na Jie. « A 31-Year (1990–2020) Global Gridded Population Dataset Generated by Cluster Analysis and Statistical Learning ». Scientific Data 11, nº 1 (24 janvier 2024): 124. https://doi.org/10.1038/s41597-024-02913-0.

Lloyd, Christopher T., Alessandro Sorichetta, et Andrew J. Tatem. « High Resolution Global Gridded Data for Use in Population Studies ». Scientific Data 4, nº 1 (31 janvier 2017): 170001. https://doi.org/10.1038/sdata.2017.1.

Lloyd, Christopher T., Chamberlain ,Heather, Kerr ,David, Yetman ,Greg, Pistolesi ,Linda, Stevens ,Forrest R., Gaughan ,Andrea E., et al. « Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets ». Big Earth Data 3, nº 2 (3 avril 2019): 108-39. https://doi.org/10.1080/20964471.2019.1625151.

MacManus, K., Balk, D., Engin, H., McGranahan, G., and Inman, R.: Estimating population and urban areas at risk of coastal hazards, 1990–2015: how data choices matter, Earth Syst. Sci. Data, 13, 5747–5801, https://doi.org/10.5194/essd-13-5747-2021, 2021.

Mesev, Victor, éd. « LandScan: A Global Population Database for Estimating Populations at Risk ». In Remotely-Sensed Cities, 0 éd., 301-14. CRC Press, 2003. https://doi.org/10.1201/9781482264678-24.

Mondal P, Tatem AJ (2012) Uncertainties in Measuring Populations Potentially Impacted by Sea Level Rise and Coastal Flooding. PLoS ONE 7(10): e48191. https://doi.org/10.1371/journal.pone.0048191

Murakami, D. and Yamagata, Y. (2016) Estimation of gridded population and GDP scenarios with spatially explicit statistical downscaling, *ArXiv*, 1610.09041, URL: https://arxiv.org/abs/1610.09041.

Oak Ridge National Laboratory (ORNL). Documentation. *About LandScan*. https://landscan.ornl.gov/about

Pesaresi, Martino, Marcello Schiavina, Panagiotis Politis, Sergio Freire, Katarzyna Krasnodębska, Johannes H. Uhl, Alessandra Carioli, et al. (2024). Advances on the Global Human Settlement Layer by Joint Assessment of Earth Observation and Population Survey Data. International Journal of Digital Earth 17 (1). doi:10.1080/17538947.2024.2390454

POPGRID Data Collaborative. 2023. Consulté le 21/06/2025. Global Population Grids: Summary Characteristics. https://popgrid.org/data-docs-table1

Rose, A., McKee, J., Sims, K., Bright, E., Reith, A., & Urban, M. (2021). *LandScan Global 2020* [Data set]. Oak Ridge National Laboratory. https://doi.org/10.48690/1523378

Schiavina, Marcello; Freire, Sergio; Alessandra Carioli; MacManus, Kytt (2023): GHS-POP R2023A - GHS population grid multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset] doi: 10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE PID: http://data.europa.eu/89h/2ff68a52-5b5b-4a22- 8f40-c41da8332cfe

Serreze, Mark, et Jennifer Francis. « The Arctic Amplification Debate ». *Climatic Change* 76 (1 juin 2006): 241-64. https://doi.org/10.1007/s10584-005-9017-y.

Smirnov, Andrey V., and Institute of Social, Economic and Energy Problems of the North, Komi Science Centre, Ural Branch of the Russian Academy of Sciences. « The Arctic Population: Dynamics and Centers of the Settlement System ». Arctic and North, nº 40 (23 septembre 2020): 270-90. https://doi.org/10.37482/issn2221-2698.2020.40.270.

Statistikmyndigheten SCB. « Open Data for Grid Statistics ». Consulté le 20 juin 2025. https://www.scb.se/en/services/open-data-api/open-geodata/grid-statistics/.

Stevens, F. R., Gaughan, A. E., Linard, C. & Tatem, A. J. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLoS ONE 10, e0107042 (2015). https://esa.un.org/unpd/wpp/Download/Standard/Population/

Tatem, A. J. WorldPop, open data for spatial demography. Sci. Data 4, 170004 (2017).

Tiecke, T. G. et al. Mapping the world population one building at a time. Preprint at https://doi.org/10.48550/arXiv.1712.05839 (2017).

Thomson, Dana R., Douglas R. Leasure, Tomas Bird, Nikos Tzavidis, et Andrew J. Tatem. « How Accurate Are WorldPop-Global-Unconstrained Gridded Population Data at the Cell-Level?: A Simulation Analysis in Urban Namibia ». *PLOS ONE* 17, nº 7 (21 juillet 2022): e0271504. https://doi.org/10.1371/journal.pone.0271504.

Vaguet, Yvette. « Fronts et frontières en Arctique, quelle singularité ? » Document. Géoconfluences. École normale supérieure de Lyon, décembre 2021. ISSN : 2492-7775. https://geoconfluences.ens-lyon.fr/informations-scientifiques/dossiers-regionaux/arctique/articles-scientifiques/fronts-et-frontieres-en-arctique.

WorldPop. Documentation. « About us ». https://www.worldpop.org/about/

WorldPop. Documentation. "Sweden population map metadata report" https://data.worldpop.org/GIS/Population/Global_2000_2020_Constrained/2019/SWE/swe_ppp_2019_metadata.html

Yin, X.; Li, P.; Feng, Z.; Yang, Y.; You, Z.; Xiao, C. Which Gridded Population Data Product Is Better? Evidences from Mainland Southeast Asia (MSEA). ISPRS Int. J. Geo-Inf. 2021, 10, 681. https:// doi.org/10.3390/ijgi10100681

Young, T. K., & Bjerregaard, P. (2019). Towards estimating the indigenous population in circumpolar regions. International Journal of Circumpolar Health, 78(1). https://doi.org/10.1080/22423982.2019.1653749