A Geosimulation for the Future Spatial Distribution of the Global Population

Carsten Keßler
Department of Planning
Aalborg University Copenhagen
Copenhagen, Denmark
kessler@plan.aau.dk

Peter J. Marcotullio
Department of Geography
Hunter College
City University of New York
New York City, NY, USA
pmarcotu@hunter.cuny.edu

Abstract

The United Nations provide detailed projections for the size of the urban and rural populations for every country in the world up to the end of the century. This paper presents results from the development of a geosimulation that predicts the future geographic distribution of the global population on a grid at 1km resolution. Starting from the status quo, the simulations project the development of population numbers for every grid cell, as well as its future land cover class between the classes urban, suburban, and rural. Our initial results indicate that we will see massive population growth in areas close to the equator, which will cause challenges for the provision of drinking water, food, and energy in the face of climate change. The simulations also show that the size of the urban extents will likely more than triple to accommodate both the additional population and the continuing trend of moving from rural to urban areas.

Keywords: Geosimulation, Population Projection, Urban Expansion

1 Introduction

The global population has crossed the 7 billion mark in 2011. The United Nations Department of Economic and Social Affairs' (UN DESA) median scenario predicts that we will reach 10 billion by 2060, and 11 billion by the end of the century (Secretariat, Population Division of the Department of Economic and Social Affairs of the United Nations, 2015). Anticipating *where exactly* those people will be living is a key requirement to be able to assess future local requirements for food, water, and energy, as well as impacts of climate change. Likewise, potential measures to address the corresponding problems depend largely on the location of these population groups. As a starting point for a better understanding of those issues, a series of global geosimulations (Benenson & Torrens, 2004) on a collection of 1x1km raster layers has been implemented, which provide a first approximation of the geographic distribution of the global population in the future.

The question of where the population will be living is especially pressing for countries that are expected to see large increases in population in the coming decades, such as those in South-East Asia and, later in the century, Sub-Saharan Africa. In those countries, assessing the future expansion of urban areas is of particular importance, as the already warm climate, together with the anticipated increase in temperature due to climate change and potential future heat waves will be further elevated by the urban heat island effect (de Sherbinin et al., 2007). A realistic assessment of the population in urban areas is therefore a key requirement for an estimate of the number of people who will be affected by extreme heat. The simulations presented here hence also model the expansion of urban and suburban areas, starting from the status quo derived from satellite imagery.

The remainder of the paper will introduce the datasets used, describe the implementation of the simulations, present initial results, and conclude with an outlook on future work.

2 Datasets and Implementation

A number of datasets from existing sources have been used as a starting point for our simulations. UN DESA provides a dataset with global population prospects up to the year 2100, from which we are using the median scenario (Secretariat, Population Division of the Department of Economic and Social Affairs of the United Nations, 2015). It distinguishes urban and rural population per country. The current (i.e., 2000) spatial distribution is provided by a dataset from the Global Rural Urban Mapping Project (GRUMP, Center for International Earth Science Information Network et al, 2011; Balk, Deichmann & Yetman *et al.,* 2006). It comes as a raster file at 30 arc second (about 1km) resolution which contains the estimated number of people per 1sqkm raster cell. GRUMP also provides a raster file for urban extents; however, its urban extents are known to overestimate the size of urban extents by a wide margin, since they have been derived from light emitted at night (Schneider et al., 2009). We have therefore combined the GRUMP raster with the European Space Agency's Global Land Cover Map (GlobCover; Bontemps, Defourny & Bogaert *et al.,* 2011), which captures the urban extents much more accurately. It is based on data from the MERIS sensor on board the ENVISAT satellite mission and comes at 300m resolution. We have used those cells classified as *artificial surfaces and associated urban areas*, downsampled them to the GRUMP resolution, and then copied them into the GRUMP raster. From this point, we considered the remaining GRUMP cells marked as urban which were outside of those copied from GlobCover as suburban areas. In addition to those core layers, the Admin 0 layer from Natural Earth[1] was rasterised at the GRUMP resolution to give us an up-to-date assignment of each cell to a country (the GRUMP nations grid is outdated in some areas where new countries have formed, such as former Yugoslavia).

Each of those three input layers – population, urban/suburban/rural, and countries – spans the globe between 57° South and 84° North. Each layer consists of approximately 730 million cells and occupies just over 3 GB on disk in uncompressed format. In order to keep the simulation runnable in a reasonable amount of time, each layer was split up into a series of smaller rasters, each containing a subset of the dataset corresponding to the bounding box of one of the countries. The simulation can then be run separately for every country, and the output is stitched back together when complete. The process has been implemented in a way so that the simulation automatically uses all *n* processor cores available on a machine to run the simulations for *n* countries in parallel.

The whole process – breaking apart the dataset, the simulations as such, and the merging at the end – was implemented in Python. After first attempts in desktop GIS systems and using different libraries developed for handling geographic data, we turned to Numpy.[2] Its extremely efficient implementations of n-dimensional arrays and operations on them have enabled a "fast-enough" implementation of the simulation. As an example, the following code snippet shows how to select cells in the population layer that match criteria across all three layers (urban cells in the Netherlands – country ID 528 – with more than 1000 people) *without* iterating over those layers:

```
a = countriesLayer  == 528
b = urbanRuralLayer == urbanCell
c = populationLayer  > 1000

matchingCells = populationLayer[
                numpy.all((a, b, c), axis=0)]
```

Most modifications of the layers can be implemented similarly using Numpy, such as adding a fixed number of people to a randomly selected subset of the cells selected above.

3 Simulation Approach

Using those building blocks, the following simulation steps were conducted for each country.

1. Make a copy of the population and urban/suburban/rural layer from the last iteration.
2. Look up the projected urban and rural population numbers from the UN DESA dataset.
3. For both urban and rural areas in the country, compare the numbers in the population layer to the UN DESA number and randomly add or remove people from the urban and rural areas to make the numbers match.
4. Push population from cells where the population numbers become too high into neighbouring cells. The decision is based on a country-specific population threshold, which is currently defined as 95% of the average of the 50 cells with the highest population numbers. This step not only prevents unrealistically dense cells, but also simulates the effect that the population density in the densest urban areas is decreasing in many countries (Angel et al., 2010).
5. Turn rural cells into suburban cells and suburban cells into urban cells if a country-specific threshold for each class is exceeded. The thresholds for these transformations have been set to the mean between the rural median and the suburban median for the rural-suburban transformation, and the mean between the suburban median and urban median for the suburban-urban transformation.
6. Repeat step 3, since the numbers for urban, suburban, and rural do not match the UN DESA numbers any longer after reclassifying some of the cells.
7. Start over at step 1 for the next iteration until the final year. Currently, the simulation runs in 10-year steps up to 2100.

Using these rules, the simulations were completed for every country up to the year 2100, and the output was stitched back together to create one global GeoTIFF with population per cell and one with the urban/suburban/rural classification for every 10 year increment. The countries layer remains unmodified in our simulation.

4 Initial Results

Figures 1 and 2 show examples of the output of the simulations for the region around Kolkata, India. As the example shows, population growth and urban expansions are projected to remain high in South-East Asia throughout the first half of the century, and then slow down. These projected developments vary greatly by region. While South-East Asia and, later, Sub-Saharan Africa are projected to experience massive urban expansion, Europe and North America are projected to experience little change. Countries that UN DESA projects to go through a massive population decline towards the end of the century, such as Japan, are still challenging for our model. The current design of the rules seems to produce a behaviour that empties out city cores and predicts that the

---

[1] See http://www.naturalearthdata.com/downloads/10m-cultural-vectors/.

[2] See http://www.numpy.org.

Figure 1: Projected urban expansion of Kolkata, India.

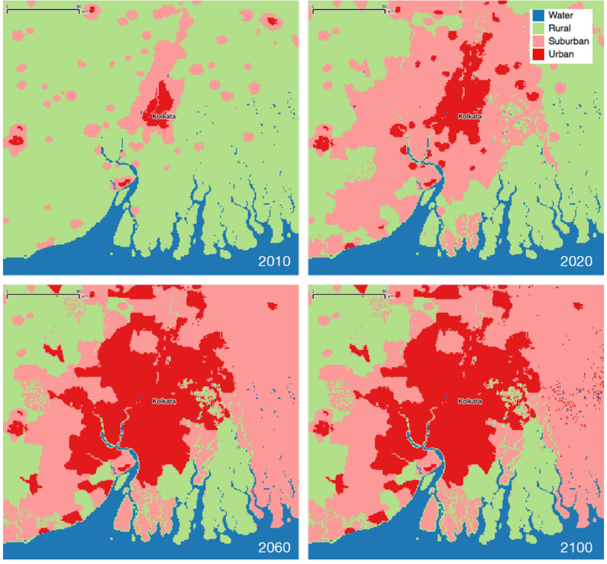


Figure 2: Projected population development of Kolkata, India.

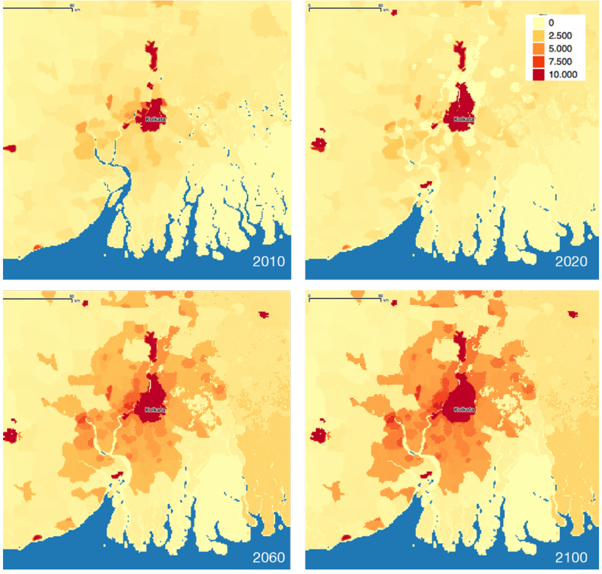
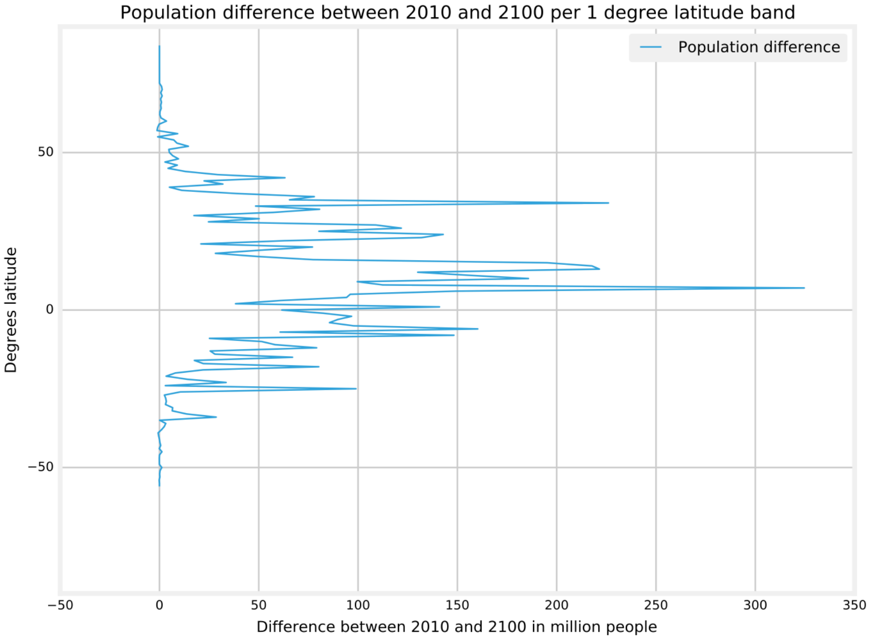


Figure 3: Projected population difference by latitude between 2010 and 2100.



remaining population will concentrate in the areas that are currently suburbs. It is still too early to say whether this is realistic or not, since no countries have gone through such a massive population decrease so far. While it seems likely that certain areas and maybe even entire cities cannot be sustained by the remaining population and have to be abandoned, how this phenomenon unfolds on a larger scale is subject to speculation.

While the local effects clearly depend on details in the model rules, the global behaviour seems to be less sensitive to a tweaking of the rules. The distribution of population, for example, largely hinges on the countries, and is therefore only subject to relocation within a country based on our simulation rules. The global distribution of population therefore changes only marginally when the rule base changes.

Figure 3 shows one of the insights gained from our simulations, namely the projected change in population per one degree latitude band. While there are only minor fluctuations in the areas North of 40°N and South of 25°S, the largest increases in population are projected for the area around the equator. This is problematic for several reasons, such as potential conflicts with the high biodiversity in this area (due to the *latitudinal diversity gradient*; see Hillebrand, 2004). Moreover – and maybe most importantly – we will see the largest increase in population in areas that are *already* hot today, and will become hotter with climate change. This will bring about new challenges concerning the provision of drinking water, food, and energy required for air conditioning. While we are still working on detailed numbers, we can already say that according to our model, combined with the different representative concentration pathway scenarios (Moss et al., 2008), up to hundreds of millions of people in those areas will be affected by prolonged periods of extreme heat in the second half of the century.

This problem will be exacerbated by the urban heat island effect (de Sherbinin et al., 2007). As such, the expansion of densely populated urban areas in these parts of the world will contribute to these challenges. Globally, the number of suburban cells is projected to increase by ~40%, from 5.394.931 in 2010 to 7.566.822 in 2100. The number of urban cells, however, more than triples, from 483.591 in 2010 to 1.675.945 in 2100.

5 Conclusion and Outlook

To the best of our knowledge, the research presented in this paper marks the first attempt to create a global geosimulation for population numbers. Our current efforts focus on using it as an explorative tool for different scenarios, namely different shared socioeconomic pathways (Riahi et al., 2017) and different representative concentration pathway scenarios (Moss et al., 2008).

While the rules used in the current version of the simulation are still relatively simple, they already produce realistic scenarios for urban expansion as well the future spatial distribution of the global population. To make these simulations more realistic at a local scale, a more complex rule base using statistical components (e.g., Arsanjani et al., 2013) will be required. The complexity of those models, however, currently prohibits their execution on off-the-shelf hardware in acceptable processing time. To make such an approach feasible for a dataset of the size given here, advances are required in the areas of parallel processing of big geospatial data, cross-scale geospatial analytics, cellular automata and agent-based modelling, as well as integration of vector- and raster based process models. As such, this research also presents some conceptual and computational challenges for geographic information science.

References

Angel, S., Parent, J., Civco, D.L. and Blei, A.M. (2010) The persistent decline in urban densities: Global and historical evidence of sprawl. *Lincoln Institute of Land Policy Working Paper*.

Arsanjani, J. J., Helbich, M., Kainz, W., & Boloorani, A. D. (2013). Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *International Journal of Applied Earth Observation and Geoinformation*, *21*, 265–275.

Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., et al. (2006) Determining Global Population Distribution: Methods, Applications and Data. *Advances in Parasitology*m 62:119–156.

Benenson, I. and Torrens, P.M. (2004) *Geosimulation: Automata-based modeling of urban phenomena*. John Wiley & Sons.

Bontemps, S., Defourny, P., Bogaert, E.V., Arino, O., et al. (2011) *GLOBCOVER 2009 – Products description and validation report*.

Center for International Earth Science Information Network (CIESIN) Columbia University, International Food Policy Research Institute (IFPRI), The World Bank, Centro Internacional de Agricultura Tropical (CIAT) (2011) *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Count Grid*. [Online; Accessed February 1, 2017]. Available from: http://dx.doi.org/10.7927/H4VT1Q1H.

Hillebrand, H. (2004) On the generality of the latitudinal diversity gradient. *The American Naturalist*, *163*(2), pp.192–211.

Moss, R., Babiker, W., Brinkman, S., Calvo, E., Carter, T., Edmonds, J., Elgizouli, I., Emori, S., Erda, L., Hibbard, K. and Jones, R.N. (2008) Towards New Scenarios for the Analysis of Emissions: Climate Change, Impacts and Response Strategies.

Riahi, K. et al., 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42, pp.153–168.

Secretariat, Population Division of the Department of Economic and Social Affairs of the United Nations (2015) *World Population Prospects, the 2015 Revision*. [Online; Accessed February 1, 2017]. Available from: https://esa.un.org/unpd/wpp/.

de Sherbinin, A., Schiller, A. & Pulsipher, A. (2007) The vulnerability of global cities to climate hazards, *Urbanization and Environment* 19(1): 39–64

Schneider, A., Friedl, M.A. & Potere, D., 2009. A new map of global urban extent from MODIS satellite data. *Environmental Research Letters*, 4(4), p.44003.