

Global high-resolution gridded population estimate datasets 2015-2030

Transcript

Professor Andy Tatem, Director, WorldPop - Human Planet Forum, 19 November 2025.

So, I'm going to talk a bit about what we do at WorldPop. We're an applied research group, around 40 people and primarily driven from the health side of things. The demand is coming for denominators, for populations at risk for planning intervention campaigns. And we're trying to tackle these types of problems of population data that comes primarily in low- and middle-income countries, where data may be available that we trust, but it's coarse resolution or it's outdated, incomplete, inaccurate and missing.

And so, we are kind of like grabbing bits of data from all over the place here - and this is where it's very exciting to hear about all these new data sets and trying to use them to better estimate populations. So that may be building data and the classification, and maybe new types of data coming on estimating building damage, which can be pretty useful in some settings, land use, land cover, infrastructure, images of the earth at night. Sometimes we're using mobility data coming from mobile phones. And the aim is to build up as detailed a picture of the landscape as possible, so that when we can't go and count everybody and enumerate people in censuses or surveys, we have a way of understanding what that landscape is like. And it looks more like a stack of data like this. And then it feeds into some form of population estimation model - whether it's population data from small areas, or a census settlement data to help us identify where people are probably not living. And then these geospatial data telling us something about conditions in those settled areas.

And then there's all kinds of models, some of which are just one variable, some of which are hundreds, to come up with these gridded population estimates. Sometimes with confidence intervals around them as well, depending on what we're doing.

And it breaks down, really, our work broadly into two types of approaches, whether we're working with an individual government Statistics Office, Ministry of Health and a single time point is needed for a single country, or what I'm going to talk about today is this new work on updating our global estimates for to try and get a consistent multiyear time series.

So, the national scale just briefly, this is over the last five years a big part of our work in collaboration with UNFPA, with UNICEF opening the doors to countries that need help with producing population estimates in the absence of a census. So, countries like South Sudan, Papua New Guinea. We've worked this co-development, like Martino said, "playing together as children" to try and estimate population numbers. And these are fed into national statistics; they feed into vaccination campaign planning. And they feed into things like catchment areas, school placements.

So that's a whole other talk, but the global side of things - this is what came out of those efforts 10 years ago, a set of global estimates for each year 2000 to 2020, broken down by age and sex, and brings together collaborations from lots of people in the room to produce that. It's since been used in a range of different ways and continues to be used around health, disease control

strategies, vaccination campaigns. And those were the ingredients are kind of around 2010 set of census data, around some old settlement maps that look rubbish compared to what we have now, and a set of covariates. But yeah, as I've said, these are now very much out of date, expired, and so this is what we've been working on the last two or three years, updating each component of this.

So, if I briefly go through this. So, a lot of work to try and assemble not only the 2010 round but the 2020 round of census data matched to administrative boundaries. It's a huge effort and ongoing, and this is certainly not a perfect complete data set, but we will soon be releasing that set and then working with our demographers at the School of Social Statistics and Demography to try and work out different methods for projecting between and past those data sets to get a smooth 2010 to 2030 time series. Then, thanks to Patrick again sharing the early outputs from population prospects so that we have a national numbers to anchor those to.

And then just, yeah, interesting insights that before we do any kind of gridding of what these show in terms of the trends in populations at subnational scales in this time period. Of populations going from younger to working age to elderly in different parts of the world. So that's another entire talk.

Then the settlement and covariate data sets. Again, thanks to everybody in the room here, again, we're bringing together both these global multiyear data sets and GHSL and the World Settlement Footprint but supplementing those with things like the building footprint data sets to identify areas that have been missed, or that can help us get more precision in some areas. This then builds into a settlement growth models to try and fill those gaps between those points where we don't have data to give us these time series. And again, notice we're kind of sidestepping classifying those into urban and rural. We are always working on a continuous scale here, don't want to get into that those challenges.

And then there is these covariates. So again, these are an open set of 100 metre gridded data sets that are available for anyone to use that cover things like built surface distance to coastlines, annual temperatures, all of which have been identified in the past as quite strong predictors in some cases of population distributions.

Then there's the modelling. So, we're going from administrative count like this, using this set of data of things like built settlement, building footprint counts, land cover classifications, images of the Earth at night to learn relationships between population densities and these set of data to go from an administrative counts to a gridded estimate of those population distributions. Trying to account for some of those factors that may exist within and between settlements. Again, these are all things that are openly available and there are plugins and in QGIS as well as new codes available. Happy to share those.

And it really also gives us some interesting scientific insights in terms of what are the things that are predicting and explaining population densities that we see across these areas, as well as highlighting where we're projecting from the last census and there is a real uncertainty in being able to explain those population distributions.

So finally, these estimates come out - and just to show the ability to improve things by bringing in all of these new data sets. So, this was Global 1 for a rural area, this is Global 2. The ability to identify individual buildings and settlements.

This was Global 1 and our predictions of population. You see just one tiny grid square at the bottom there - which is clearly wrong because there's lots of settlement there - which we've missed. But by being able to bring in GHSL World settlement footprint, the building footprint data sets, we're now capturing those areas much better.

This was absolute rubbish in Global 1 this was this was Niamey in Niger. We didn't have good data then, but because of the improvements in all of these data sets, we're capturing not only the extent better but the within settlement variations that exist.

And also, the types of data that are being produced through when we're bringing in Open Street Map data, when we're bringing in GHSL classifications, we can identify different types of land use and account for those in the model. So Global 1, we're just predicting every population pretty much everywhere here in Southampton. Now we're able to put populations in the right place and not predict people in the middle of airports or industrial zones.

An important part of bringing in the two time points of census data as well. In Global 1 we have a look here - this is 2015 and 2020. We see this massive growth in in these in two settlements here. And that's really because we were just relying on one time point of census and an assessment of how populations would change in the future. Now with Global 2, we have these anchors, these two time points, that enables us to adjust those population counts accordingly.

So, these are the outputs that are then produced as a time series that doesn't look so rubbish as before.

But yeah, just to bring in the Limitations. These are top down desegregations, so whatever goes in feeds through to the outputs. Projections can be highly uncertain. We've heard about that as well and there's a release statement talks a lot about these things.

Just to finish off - all of these data can be accessed in multiple different ways.

And a few things of what's coming next. We're going to be working on not only population totals, but densities, summaries by administrative units, degree of urbanisation. Hopefully there's a lot of things we can collaborate on here, I think. And to think forward to the next steps of bringing in some of the data sets we've heard today bringing in our bottom up estimates where we're working with countries to integrate those in. And also, the government and UN co-development, there's a huge amount of demand from statistics offices - so more than we can cope with.

So, this is something I hope we can discuss together where maybe we can join forces. We have a lot of training materials as a result of that.

And then just sticking this in at the end, because we've talked a lot about this, but some work funded by the Wellcome Trust to support us working with the integrated assessment modelling community, the CMIP community, to produce estimates that go into their modelling scenarios. So, these are just very early outputs. And new versions are coming, but hopefully this adds to what we've been talking about - the SSP scenario data sets.

So, I will finish there. And more information is on the screen. Thank you.

Lewis Dijkstra, JRC (Session Chair):

Thank you very much, Andy. I mean, amazing work that the WorldPop team is doing and really helping to make a big impact there as well on the policy front.