DEGURBA: Why Your Urban Definition Depends on Your Data

Google NotebookLM/WorldPop, 16 December 2025

Speaker 1: You'd think that defining what's urban versus what's rural would be pretty straightforward, right?

Speaker 2: You would, but it's surprisingly complex on a global scale.

Speaker 1: I mean, you have countries like Denmark that say a place is urban with just 200 people. Then you look at Japan, where the threshold is 50,000.

Speaker 2: And some countries even base it on agricultural employment. It makes any kind of global comparison incredibly difficult.

Speaker 1: Exactly.

Speaker 2: And before we get into this solution, a quick note for everyone listening. We aren't real people. We're AI voices derived from source material provided by WorldPop.

Speaker 1: And this entire Deep Dive has been edited, checked and validated by the human experts at WorldPop.

Speaker 2: Right. So, our mission today is to look at a tool designed to fix this exact problem of fragmentation. It's called the Degree of Urbanisation framework, or DEGURBA.

Speaker 1: OK, DEGURBA. What does it do differently.

Speaker 2: Well, it's been endorsed by the UN Statistical Commission, and it creates a single harmonised approach. It uses a 1 kilometre grid system to classify any area as either an urban centre, an urban cluster or a rural area.

Speaker 1: So, it gets around all those messy administrative boundaries and different national rules. And our source for this is a new paper led by WorldPop Research Fellow Dr Wenbin Zhang. It analyses the DEGURBA framework and gives some really important guidance on how to implement it. Which brings us to the core question, and it's a big one. How sensitive is this DEGURBA classification to the actual gridded population data set you feed into it?

Speaker 2: So, the study looked at 16 countries across Africa and the Caribbean, and it used four different major grid population data sets to see what would happen.

Speaker 1: And those were WorldPop, bespoke national constrained, WorldPop global, unconstrained, WorldPop global, and GHS POP.

Speaker 2: Yes, those four.

Speaker 1: So, let's get right to the headline finding. What was the level of disagreement?

Speaker 2: It was pretty significant. On average 27.5% of the population fell into what the study calls mixed classes.

Speaker 1: A quarter of the population. So that means for one in four people, their classification as urban or rural change depending on which data set you used?

Speaker 2: Exactly. And that's a huge deal for resource allocation. Are you building a clinic in a rural area or an urban cluster? The data could give you 2 different answers.

Speaker 1: And the study found that the biggest point of friction was in that urban cluster category, right?

Speaker 2: That's where the disagreement was most pronounced. The mixed rural and urban clusters category alone accounted for 17.5% of the disagreement.

Speaker 1: So why? What's going on with these models that makes them disagree so much about these specific areas?

Speaker 2: It all comes down to the modelling approach. You see, three of the data sets WPB, WPC and GHS PO are constrained.

Speaker 1: What does constrained mean in this context?

Speaker 2: It means they only place population estimates where satellites can actually detect settlements. Buildings, lights, things like that. And those models were very consistent with each other.

Speaker 1: OK. So, they're tied to physical evidence, but what about the outlier? The unconstrained WorldPop global or WPU? It sounds like it should be more comprehensive.

Speaker 2: You'd think so, but that's where the problem lies. The WPU model distributes population more broadly, even into areas with no detected buildings.

Speaker 1: And that causes issues for the classification.

Speaker 2: A huge issue. DEGURBA needs a certain population density to classify an area as an urban cluster. Because WPU spreads the population out so thinly, many of these grid cells just don't hit that density threshold.

Speaker 1: So, the legitimate urban cluster might get misclassified as rural just because of the way the model spreads people out.

Speaker 2: That's exactly what was happening. When the researchers removed the WPU data set from the comparison, the share of the population in those mixed classes dropped by 8.7 percentage points. The uncertainty just decreased significantly.

Speaker 1: And there's one country where this effect was really extreme, wasn't there?

Speaker 2: Yes, South Sudan, it was the most sensitive country in the study. When WPU was removed, the uncertainty in its population classifications fell by an incredible 92.5%, which is a stark contrast to places like Namibia or Mali, which were much more stable across all four data sets.

Speaker 1: So, the guidance from the paper for policymakers seems crystal clear.

Speaker 2: It is. if you're going to use DEGURBA, you should use modelled population grids that are constrained to settled areas. They are simply more robust and suitable for this kind of classification.

Speaker 1: At the end of the day, then, DEGURBA is a powerful tool.

Speaker 2: A very stable and valuable tool, yes. But its accuracy completely depends on the quality of the input data. You need that high quality, geographically specific foundation.

Speaker 1: Which leads to a final thought for any planners or policymakers listening. Given these sensitivities, especially in places where you don't have official gridded population data, maybe just picking one data set isn't enough.

Speaker 2: That's a very important point.

Speaker 1: Perhaps comparing multiple constrained gridded data sets should be a mandatory first step, a kind of due diligence to make sure the data you're using to make critical decisions is actually the most appropriate for your area.

Speaker 2: A crucial step, I'd say.

Speaker 1: To read the full journal article, follow the link below.