Harmonized census geography and spatio-temporal analysis: gender equality and empowerment of women in Africa

Sula Sarkar, Lara Cleveland, Majory Silisyene, and Matthew Sobek — University of Minnesota

Introduction. Changes in administrative boundaries pose a major challenge for spatio-temporal population research. Researchers interested in change over time need to hold space constant to study contextual or spatial effects on behaviors and outcomes. Boundary changes risk polluting their analyses with artifacts that obscure real changes that may have occurred. This paper describes the method by which spatially consistent geographic units have been constructed in the IPUMS-International census data collection for dozens of countries over a fifty year period. Low-level geographic units are grouped into temporally compatible base units. Regionalization[A] (combining) techniques are applied to create spatio-temporally harmonized units that meet the 20,000 population threshold required for public dissemination of the data. The base units are then aggregated as necessary to define spatial units that are completely consistent across all census years. We illustrate the utility of the harmonized units by exploring progress toward UN Millennium Development Goals[B] in a number of African countries at the sub-national level: specifically the goals to "promote gender equality and empower women." The analysis shows generalized increases in the number of women completing secondary education and participating in the labor force, but the pattern of growth differs markedly both across and within countries. We show how the use of harmonized geographic units facilitates comparative metrics.

IPUMS Integrated Statistical Areas. The Integrated Public use Microdata Series, International (IPUMS)[C] is the world's largest publically accessible population database. It currently includes sample data for 258 censuses from 79 countries. IPUMS is composed of microdata, with each record representing a person for whom all individual census characteristics are known. IPUMS is designed for comparative research across time and countries. Most countries have multiple censuses in the database. Variables are harmonized across IPUMS samples, so the same codes apply in all times and places. The dissemination system allows users to build custom data extracts that pool data from as many samples as they request.

In the past, IPUMS performed only rudimentary harmonization of geographic variables, which present some of the most difficult challenges in the development of the data series. With the most recent data release in summer 2014, IPUMS has initiated a thorough overhaul of sub-national geography using GIS techniques. The GIS work depends on the acquisition or creation of historical boundary files for each of a country's censuses. For older samples, images from historical maps or census volumes are digitized and converted into digital files through a labor intensive process.

Once boundary files are available, the first stage of the geographic work involves confidentiality. For many developing countries IPUMS has detailed geographic information, but is obligated to only identify places with at least 20,000 population in the samples that are disseminated. Smaller places need to be combined to reach that threshold, which typically arises at the second administrative level within countries (e.g., counties in the U.S. or municipios in Mexico). Sub-20,000 units are grouped using a regionalization algorithm[A] that privileges contiguity and similarity in population density, and which prohibits combinations across first-level boundaries.

The confidentialized second-level units are developed for the most recent census year and applied backward in time. Where the borders of the modern units do not align with historical units, because of boundary changes, larger aggregated units are created. Any historical boundary changes happen within these aggregated units and no changes cross their perimeter. The goal is to define the smallest possible unit that shares the exact same footprint in all census years. The first administrative level units must also be harmonized over time, aggregating as necessary to account for their boundary changes. This is less challenging than the lower level work.

In the end, users are presented with a set of Integrated Statistical Areas (ISAs): spatially consistent units at the first and second geographic level within countries. In addition, a globe-spanning variable combines all the first

level harmonized units from all countries as an aid to comparative research. The harmonized units are sometimes substantially larger than the places that can be identified in a single census year for a country, but they are stable over time. Their main purpose is for research over time. In order not to disadvantage users interested in only a specific sample, the contemporary units and associated GIS files applying to each census are also available.

Millennium Development Goals. Variables in IPUMS are coded consistently across time and country. These harmonized data can be used to measure progress over time and across space on several of the MDG indicators[D]. In this paper we focus on Goal 3, to promote gender equality and women's empowerment. To illustrate the use of ISA geographic variables, we map changes in outcomes for select African countries for two Goal 3 indicators: 3.1) gender disparity in primary and secondary education; and 3.2) share of women in wage employment in the non-agricultural sector. We measure progress on the MDGs at the national and first level of geography for all the selected countries. In the process, we demonstrate the need for a spatially consistent geographical footprint and also indicate when year-specific census geography should be used in conjunction with the spatially consistent ISA geographies. We also visualize progress on the goals at the lower (second) level of geography using spatially consistent geographic units for a single country: Malawi. For the purposes of this abstract, we focus only on goal 3.2 and visualize Mali and Malawi. Our final paper studies all parts to MDG goal 3, gender equality and women's empowerment and visualizes results from a number of African countries.

Table 1 presents national-level statistics on the share of women in non-agricultural wage work in a number of African countries. The measures are calculated from the IPUMS microdata and are limited to those countries with at least two censuses containing the requisite variables. The Millennium Goals were stipulated in 2000, and census years differ from one country to the next. We have categorized the data sources by time periods according to whether the data was collected prior to the implementation of the Millennium Development Goals, in the first 5 years following implementation, or more than 5 years after implementation when effects might begin to show in the data.

Table 1. Percent female of non-agricultural wage employment (MDG goal 3.2)

Country	1988 to 2000 Pre-MDG	2001 to 2006 Early MDG	2007 to 2011 Late MDG
Egypt	18.9 (1996)	21.2 (2006)	--
Ghana	34.4 (2000)	--	33.7 (2010)
Malawi	19.4 (1998)	--	24.6 (2008)
Mali	25.3 (1998)	--	46.2 (2009)
Morocco	23.7 (1994)	24.4 (2004)	--
South Africa	--	44.9 (2001)	43.7 (2007)
Zambia	23.9 (2000)	--	28.0 (2010)

Table 1 demonstrates significant variation across countries in progress toward gender equality in employment. Malawi and Mali show notable improvement, while the other countries have made modest progress or even lost ground with respect to the MDG indicator. These national figures, however, can hide significant variation in levels and rates of change at the subnational level.

To explore MDG progress within countries, we focus on the two in this abstract that displayed greatest improvement overall: Mali and Malawi. In both cases we use visual representations of performance toward the gender employment goal, initially at the first geographic level. In Figure 1, for Mali, all regions show considerable progress, but the central area has the highest measures and the west lags behind the rest of the country. Malawi (Figure 2, Map A and B) presents a more variegated pattern of achievement. Much of the

progress was concentrated in the northern districts, which helped drive up the national figures. The far south was largely stagnant.

For the spatial visualization discussed above, we used the Integrated Statistical Areas to hold boundaries constant over time. While that enables an apples-to-apples temporal comparison of places, the nature of the ISAs is to merge census units to encapsulate any boundary changes that occurred between censuses. In the process, some detail that might be useful for the analysis gets lost. Figure 2, Map C illustrates this point. In it, we map the original census units from 2008: Lilongwe city, Balaka, and Zomba city are identified as separate units not observable in the spatially consistent 1998 and 2008 maps (Figure 2, Maps A and B). Each of them has greater female wage employment rates than their surrounding areas. In fact, they demonstrate that much of the apparent progress in their regions was more localized in urban places.

Figure 1. Female non-agricultural wage employment, Mali Regions 1998-2009

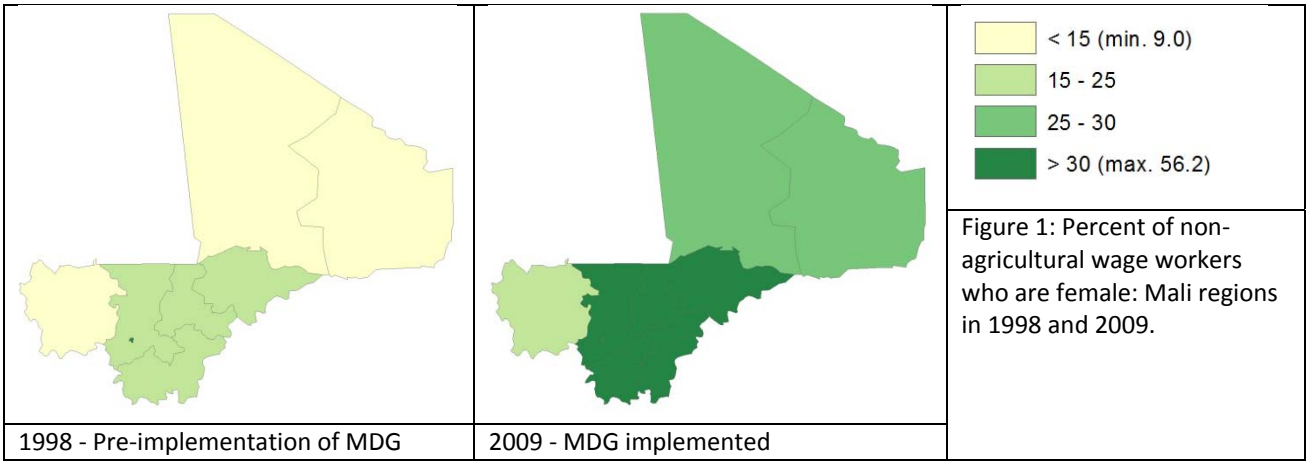
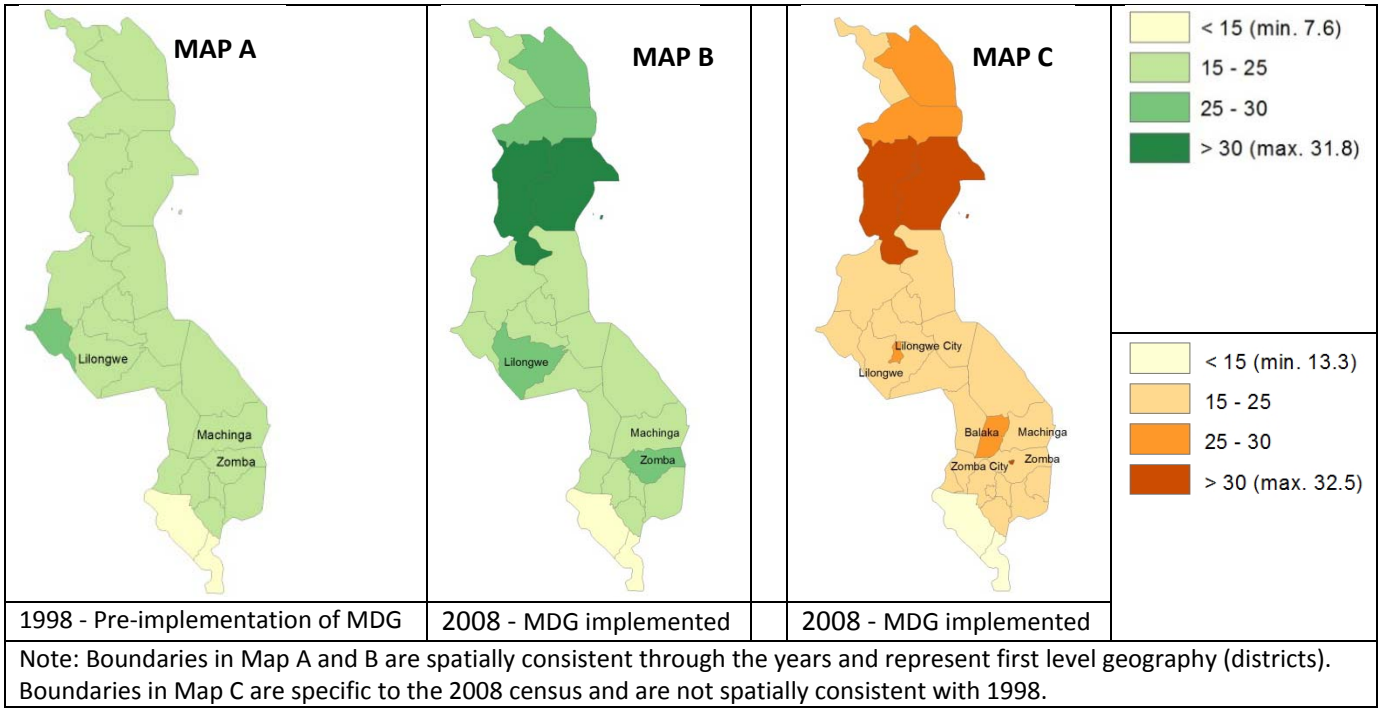


Figure 1: Percent of non-agricultural wage workers who are female: Mali regions in 1998 and 2009.

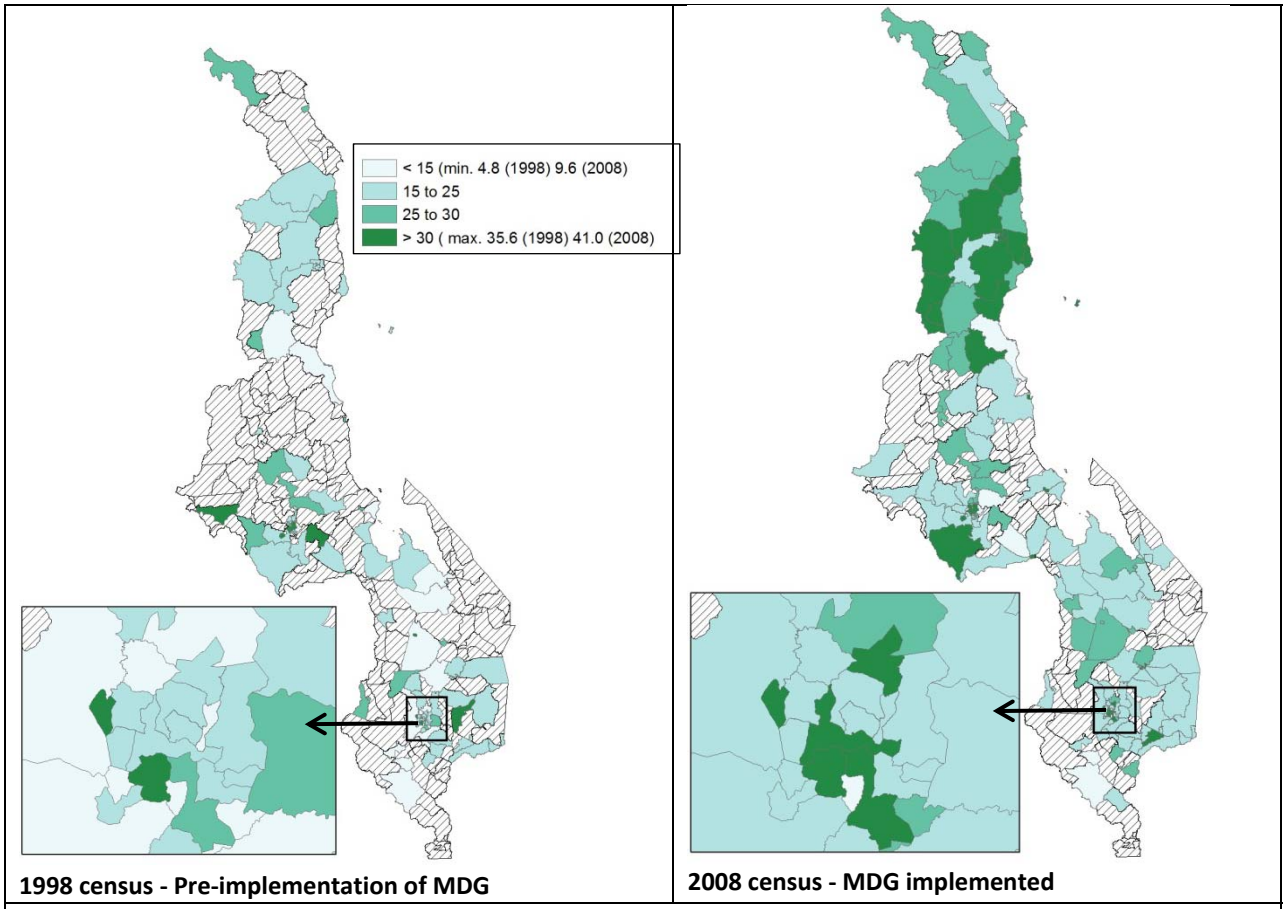
Figure 2. Female non-agricultural wage employment, Malawi Districts 1998-2008



Note: Boundaries in Map A and B are spatially consistent through the years and represent first level geography (districts). Boundaries in Map C are specific to the 2008 census and are not spatially consistent with 1998.

Figure 3 presents the percent share of female non-agricultural wage employment in the Traditional Areas (TA) of Malawi. TAs are the second-level geographic units in Malawi, and Figure 3 employs the spatially consistent variant of them to enable direct comparison across censuses. At this scale one gets the benefit of harmonized geography without some of the cost described at the higher geographic level in Figure 2 above. The TAs show regions of little or no gain in the MDG indicator -- patterns that were not observable at the larger scale. The blow-up of the urban area of Blantyre shows distinctions at a near neighborhood level where population densities are sufficient to overcome confidentiality constraints. Even though from Figure 2 we see limited progress in the Blantyre area, there is significant progress towards goal 3.2 in some of its constituent parts. The limitations of sample data are evident in Figure 3, however: cases are too sparse to calculate reliable non-agricultural statistics in many Traditional Areas.

Figure 3. Female non-agricultural wage employment, Malawi Traditional Areas 1998-2008



Note: The boundaries of the Traditional Areas are spatially consistent through the two census time periods. The inset map shows the urban area of Blantyre.
Note: The non-colored hatched TA boundaries represent very low (n<20) female non-agricultural wage earners in the sample data.

Summary. Consistent geographic units are necessary for accurate measures of change over time that involves contextual or spatial elements. The IPUMS integrated spatial units impose some costs in terms of lost precision, which can be explored using census-specific geography. These costs are mitigated at the second geographic level, although certain measures may prove hard to calculate at that scale. At this time, IPUMS is working to make the second-level geography available for as many countries as possible.

References.

[A] Guo, D. (2008). "Regionalization with Dynamically Constrained Agglomerative Clustering and Partitioning (REDCAP)".*International Journal of Geographical Information Science*. 22(7), pp. 801-823.
[B] United Nations. Millennium Development Goals. http://www.un.org/millenniumgoals
[C] Minnesota Population Center. *Integrated Public Use Microdata Series, International: Version 6.3* [Machine-readable database]. Minneapolis: University of Minnesota, 2014.
[D] Cuesta, A. & Lovaton, R. (2014) "Millennium Development Goals (MDGs): Measuring Within-Country Inequalities for Selected Indicators for South America using IPUMS-International Data (1990-2010)", presented in Session 59, Global Perspectives on Demography and Gender Inequality, Population Association of America (PAA) Conference 2014