

Measuring the geographic distribution of maternal education in Africa

Nick Graetz

A thesis

submitted in partial fulfillment of the requirements for the degree of
Master of Public Health

University of Washington

2016

Committee:

Emmanuela Gakidou, Chair

Simon Hay

Abraham Flaxman

David Pigott

Program Authorized to Offer Degree:

Global Health

© Copyright 2016

Nick Graetz

University of Washington

Abstract

Measuring the geographic distribution of maternal education in Africa

Nick Graetz

Chair of the Supervisory Committee:

Professor Emmanuela Gakidou

Global Health

International agendas have increasingly focused on education as a powerful social determinant of child and maternal health outcomes. However, comparable indicators of educational attainment only exist at the national level, which may obscure subnational inequality in both levels and progress. The advent of increasingly granular geographic data in household surveys along with advances in the field of Bayesian model-based geostatistics (MBG) allows for precise, efficient estimation of basic educational indicators at a high spatial resolution. By applying these methods to three years of data from the Demographic Health Survey (DHS) in Uganda, Kenya, and Nigeria, this study reveals stark geographic inequality in the most basic indicators of educational attainment despite apparent national progress. Moving into the era of the Sustainable Development Goals (SDGs), this underscores the need for evaluating progress at a more local level to avoid patterns of funding that may boost national statistics but entrench existing subnational inequities.

I. Introduction

With recent research suggesting that increases in maternal education are associated with huge decreases in child mortality at the national level (Gakidou et al. 2010), international agendas have increasingly focused on education as a powerful social determinant of child and maternal health outcomes.¹² A recent UNICEF report remarked that “across much of South Asia and sub-Saharan Africa, children born to mothers with no education are almost three times more likely to die before they are five than those born to mothers with a secondary education.”³ However, the demonstrated relationship between level of education and child mortality is variable by country, and there is concern that large gains in national averages have not been equitably distributed.⁴⁵³ Increases in the national average for an indicator could be entirely driven by gains in a specific population or a specific region of the country. Because of the correlation in national average progress in increasing education with national average progress in reducing child mortality (and other health outcomes), this has a large implication. It’s possible that countries may have achieved MDG targets over the last period of international agenda-setting, but who actually saw that progress within the country? What spaces saw the most progress, and which stagnate or deteriorate over the same period? In an effort to target intervention efforts more effectively at the local level and avoid entrenching subnational inequities, it’s becoming increasingly important to understand the distribution of education and progress achieved. The Sustainable Development Goals (SDGs), adopted last year by 193 UN Member States, aim by 2030 to eradicate poverty and ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.² In their Education for All 2030 initiative, UNESCO refocuses on an agenda for reforming education access in developing countries centered around equity.¹ However, comparable indicators for education across countries and time only exist at the national aggregate level.

In considering subnational inequality, it is important to mark the distinction between exploring the distribution of people versus the spatial distribution of geographic areas. In an ideal modeling world, there would be no distinction - we would be modeling every individual in space and time, and could explore their distribution of values (inequality) or the spatial clustering of values (segregation). This is obviously not feasible given real-world data coverage, so household survey data like the DHS is usually aggregated with the sample design to produce an unbiased estimate of a population statistic, i.e. national average years of education. However, rather than only focusing on the national mean you could focus on a statistic that describes the distribution of people in that country, like the standard deviation or the Gini coefficient, a traditional metric for characterizing inequality in a representative sample of incomes. In examining a distributional statistic over time, you could explore whether the mean has been increasing over time while the spread has also been increasing; the initially well-off population is progressing quickly while the initially disadvantaged population is stagnating or getting worse.

This would answer an important question, and more research is certainly required to explore whether national distribution characteristics of education are more predictive for health outcomes than national means. Alternatively, little research has been done on fully leveraging the increasingly comprehensive spatial information in more recent cycles of DHS and other household survey data, particularly for modeling social exposures.⁶ Modeling average education indicators at a much finer resolution than national could provide valuable insight into the question of educational inequality and the spatial clustering of progress. There should also be strong consideration paid to how policy is being enacted - over specific populations or over specific geographic areas.

New methods for high-resolution geospatial modeling offer and exciting and efficient

new avenue for estimating a surface of educational attainment using geolocated DHS clusters. Bayesian model-based geostatistics (MBG), recently pioneered largely in the field of paristology to explore distributions of malaria indicators (and other infectious diseases), allows efficient and robust estimation of the spatial structure of data that is precisely located in space, i.e. latitude and longitude.⁷⁸⁹ The DHS Spatial Analysis Reports have called for more research applying these types of methods to the new abundance of geolocated cluster data.

The goal of the current study is to explore how national education levels and gains in several African countries have been distributed subnationally over time, and how that distribution varies by country. This will be a novel application of very recently developed and evolving methods for leveraging the spatial information in geolocated DHS clusters. Being able to formalize and visualize a predictive model will provide useful insights on measuring equity in maternal and child health progress as they relate to international goals, and will allow exploration of whether or not all national average progress is occurring in populous urban centers as well as informing more targeted intervention efforts across space. This analysis also has implications for future large-scale descriptive analyses of health outcomes, such as the Global Burden of Disease. As estimates of disease burden become more granular and given that the strong relationship between average education and health outcomes holds at more local levels, it will be important to have a robust, high-resolution suite of social covariates to inform out-of-sample mortality and disease prediction.

II. Data

Data used in this analysis came from the Demographic Health Survey (DHS) for three countries: Nigeria, Uganda, and Kenya. These were chosen as pilot countries for this analysis because each had three DHS with geolocated clusters defining at least a 10-year range. This will allow for a reasonable assessment of progress over time, and the years are equally spaced over the period allowing us to reasonably define the temporal component of our model. At each cluster we were interested in two indicators: mean years of education and proportion of women with 0 years of education. For the sake of comparability between countries and over time, all education data were mapped to standard single-years of education using the UNESCO standards.¹⁰ Both indicators were estimated as the averages for all women between ages 15 and 54, treating the cluster as a simple random sample in space/time for the 5*5 km grid cell that the cluster latitude/longitude fell within.

Country	Period	Clusters
Nigeria	2003	362
	2008	886
	2013	896
Uganda	2000	298
	2006	368
	2011	404
Kenya	2003	400
	2008	398
	2014	1594

We used several covariate 5*5 km surfaces in our model to better predict maternal education

indicators in areas with sparse data coverage: brightness of night-time lights from the Defense Meteorology Satellite Program as a proxy for urbanicity, total population, and proportion of land under irrigation as a proxy for agricultural productivity. The irrigation layer was temporally static while the other surfaces varied temporally, so the mean was estimated for each period in our data. These surfaces are meant to serve as rough proxies for omitted variables related to sociodemographic status to inform our predictions in areas that are spatially or temporally far from data.

III. Methods

Bayesian model-based geostatistics (MBG) is a method for finding a large sample of maps that can explain a dataset, in which maps that do a better job of explaining the data are more likely to be represented.⁸ This sample represents the knowledge that the analyst has gained from the data about the unknown true map. The MBG framework provides a conceptually simple way to convert these samples to predictions of features of the unknown true map, for example averages or probability densities. For the proportion of women with 0 years of education we employed a spatiotemporal hierarchical generalized linear regression model for binomial data using the logit link function (Model 1). Similarly, for mean years of education this model was adapted for bounded continuous data (0-18 years of education) by standardizing and also using the logit link function (Model 2).

$$\begin{array}{ll}
 (1) & (2) \\
 N_i^+ \sim \text{Binomial}(\rho_i, N_i) & \mu_i^+ \sim \text{Gaussian}(\mu_i, \sigma_i^2) \\
 \log\left(\frac{\rho_i}{1-\rho_i}\right) = \alpha + X_i\beta + \epsilon_i & \log\left(\frac{\mu_i}{1-\mu_i}\right) = \alpha + X_i\beta + \epsilon_i \\
 \epsilon \sim GP(0, K_{space} \otimes K_{time}) & \epsilon \sim GP(0, K_{space} \otimes K_{time}) \\
 K_{space} = \tau (2^{\nu-1}\Gamma(\nu)) (\kappa D)^\nu K_\nu(\kappa D) & K_{space} = \tau (2^{\nu-1}\Gamma(\nu)) (\kappa D)^\nu K_\nu(\kappa D) \\
 K_{timeij} = \begin{cases} (1 + \rho^2) & \text{if } |t_i - t_j| = 0, \\ -\rho & \text{if } |t_i - t_j| = 1, \\ 0 & \text{otherwise} \end{cases} & K_{timeij} = \begin{cases} (1 + \rho^2) & \text{if } |t_i - t_j| = 0, \\ -\rho & \text{if } |t_i - t_j| = 1, \\ 0 & \text{otherwise} \end{cases}
 \end{array}$$

N_i represents the number of women within a cluster while N_i^+ represents the number of women with 0 years of education observed in a cluster. The probability of a given woman having 0 years of education, represented by p_i , is modelled as a logit-linear function of the global intercept α , the vector of cluster-level covariate values X_i and regression coefficients β , and spatio-temporally-correlated residuals ϵ_i drawn from a three-dimensional, zero-mean Gaussian process (GP) with covariance matrix constructed as the Kronecker product of a spatial covariance matrix K_{space} and temporal covariance matrix K_{time} . This method of parameterizing the Matérn covariance function corresponds to the stationary solution of the stochastic partial differential equation (SPDE) representation of the stationary Matérn family.¹¹ This representation of the Gaussian Markov random field (GMRF) allows us to make use of recent software in R-INLA to approximate the resulting sparse precision matrices extremely efficiently, and this approach has been shown to have extremely high accuracy when compared to other methods for estimating continuous GMRFs that would be computationally prohibitive, such as MCMC¹¹. This methodology is thoroughly covered

in Appendix I. A very similar approach was used for mean years of education, representing the parameter of interest through a Gaussian process with a logit link (years of education bounded 0-18 and standardized to avoid implausible prediction out-of-sample).

IV. Results

Results are presented in summary maps for Uganda (results for Kenya and Nigeria can be found in Appendix II). We sample from the posterior of the full model 100 times, yielding 100 credible realizations of the true space/time surface given our data (i.e. 100 maps that could credibly explain the data, or “credible maps”). Figure 1 summarizes the full posterior from our model by taking the average value at each 5x5 km cell (“pixel”) across these 100 credible maps. This is a useful and familiar tool for visualizing the predictions, but the spatial correlation *within* draw is broken when summarizing *between* draws and this can affect interpretability of results, especially for unstable models with a high degree of short-range variability.⁸ In the most extreme case of short-range variability, in half of our sample of credible maps we might estimate one pixel to be 0 years while the estimate for an adjacent pixel is 18 years. In the other half we may estimate the opposite, but in the mean surface this will look like both adjacent pixels have a very similar mean estimate of ~9 years. This is a case where the resulting summary surface, where both pixels are represented by their mean estimate of 9, will itself not be a credible sample of our model/data. There is no credible sample where our model estimates both these pixels to be 9 - the spatial correlation within draw has been broken by aggregating across draws. The extreme variability across credible maps in this specific scenario is unlikely, but in dense urban areas there may be true, large variation between adjacent pixels that can be obscured by summarizing with the mean across credible maps. This visualization of the mean also provides no inherent clarity on imprecision in the model; a pixel with an average of 9 years of education may be the result of estimates ranging from 0 to 18 across all 100 credible maps, or estimates ranging from 8 to 10. Despite these limitations, the mean surface is a useful tool for visualizing broad spatial and temporal trends in levels of education, and can also be helpfully visualized as the population-weighted aggregate at the admin2-level in Figure 2. This can be useful for benchmarking (and external data/model validation), as consumers of these maps may have strong priors on the level of certain administrative units relative to others. This method of aggregating estimates also takes into account how many people are actually living in each pixel within the administrative unit, unlike the continuous mean surface where an estimate over a large sparsely-populated area can be visually misleading.

An alternative to summarizing the model with the pixel-level average across the 100 credible maps is to use the full set to make an inference about the true surface relative to a fixed threshold. This is the most accurate depiction of the full posterior from the model, and this visualization illuminates informative trends while also forcing the consumer to reckon with the inherent uncertainty given our data and model in a way that the mean surface does not. The spatial and temporal correlation within each credible map is also not broken by this summary, as we are simply creating a density of draws for each pixel and estimating the proportion above or below a fixed threshold. Figure 3 shows the continuous probability of a pixel being above a threshold (as is the case of “goals” for mean years of education achieved) or below a threshold (as is the case of goals for proportion of women with 0 years of education). This visualizes precision in our estimates; for the pixel with a mean of 9 years that ranges from 0-18 years across all credible maps, we would not be able to infer that the pixel’s true value is above 6. However, for the pixel with a mean of 9

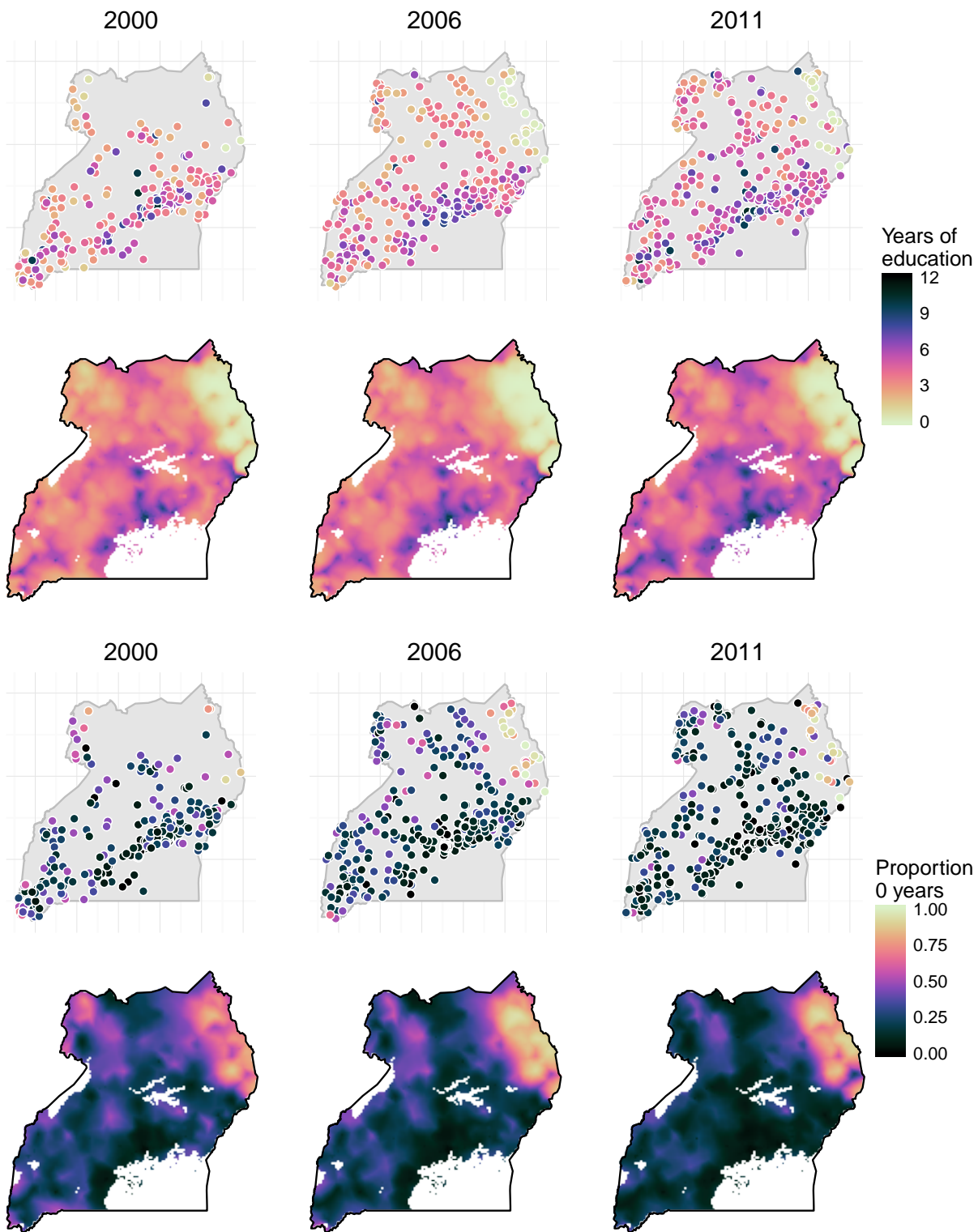


Figure 1: Data coverage and posterior mean for Uganda.

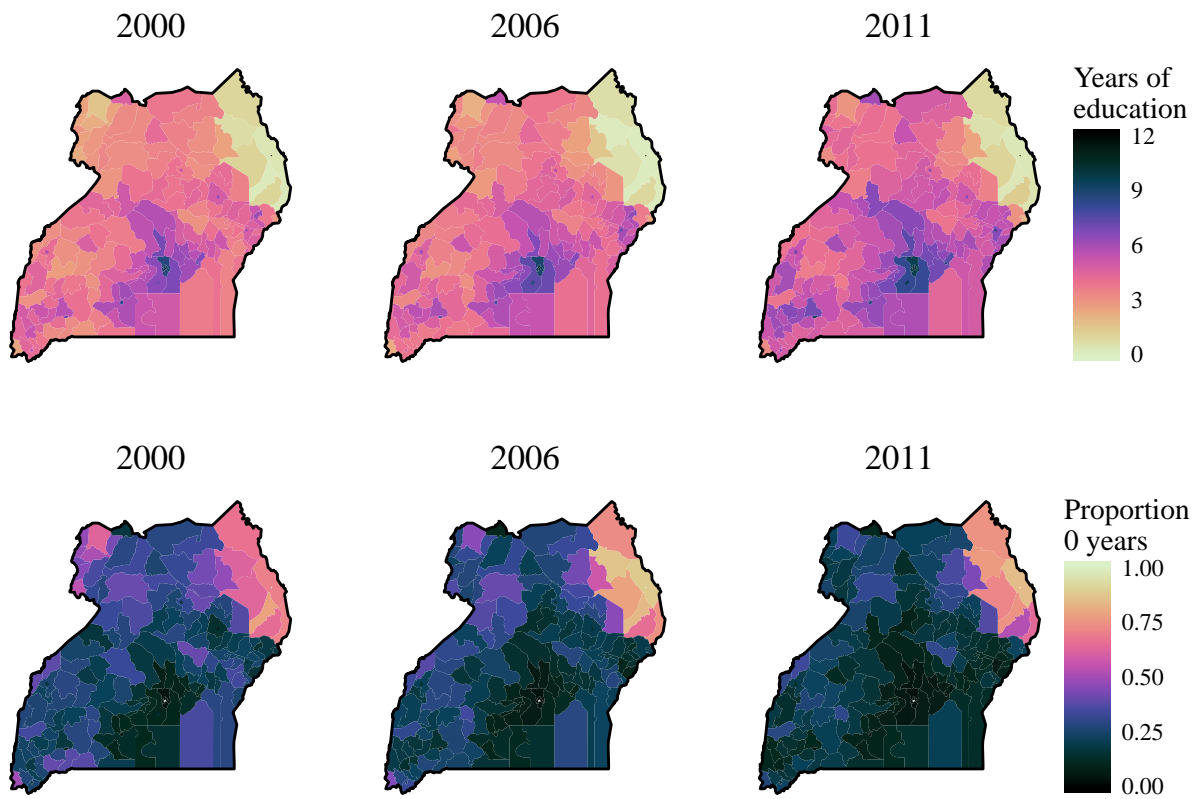


Figure 2: Population-weighted average by admin2 in Uganda.

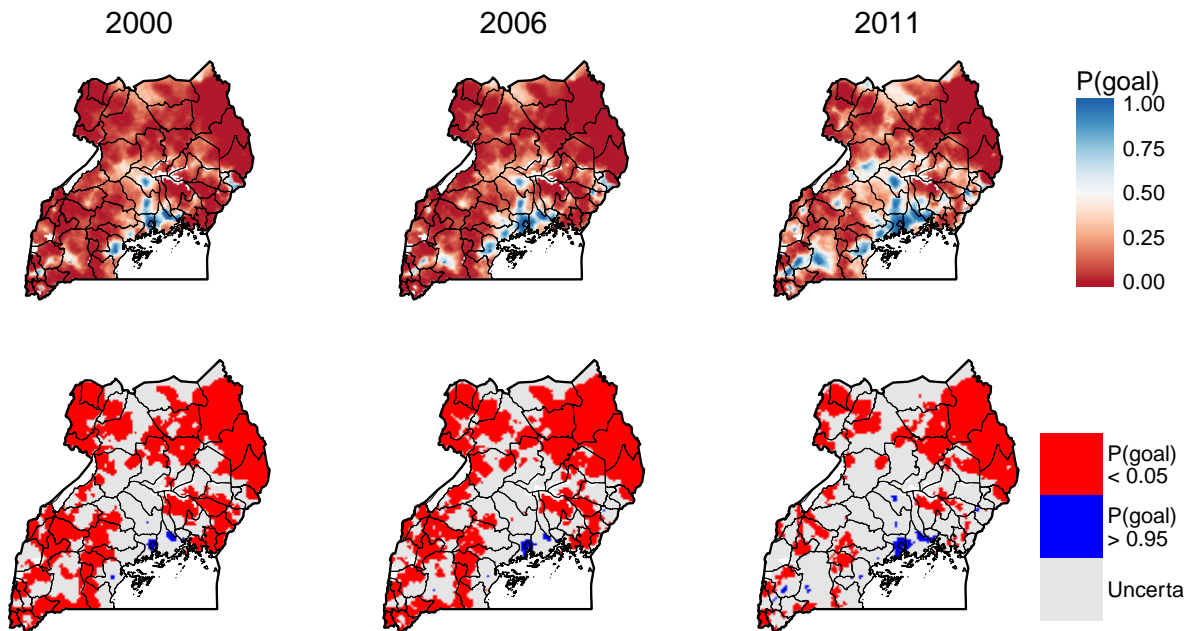
years that ranges from 8-10 across all credible maps we would be able to reasonably infer that the true value for that pixel is above 6 years. Where these pixels would both be visualized as 9 years in Figure 1, Figure 3 would show the first pixel as very uncertain and the second pixel as significantly above the threshold. The second row of Figure 3 is masked to only pixels where 95% of the draws are above or below the threshold.

This method of engaging with MBG models lends itself to political priority-setting, popularized on an international scale by the MDGs and now SDGs - i.e. “Achieve less than 20% of women with 0 years of education everywhere.” Given our data and model, where are we certain that goal has been met? Where can we credibly infer it has not been met? And perhaps most importantly, where is our inferential knowledge limited given the paucity of quality data, or by extreme small-scale variation? This method of visual inference can be applied to concrete targets within country or to externally imposed targets, such as the SDGs. Evaluating external goals that are the same across countries can be effective for cross-country comparison and benchmarking. Reports by the UN and other international organizations tend to focus on global numbers: “Enrolment in primary education in developing countries has reached 91 per cent but 57 million children remain out of school.”² If this is true, what does the spatial distribution of that ~10% look like? It’s certainly not the case that no matter where you look in developing countries the breakdown of school enrollment will be about 90-10. Proportions within country will vary widely even if they average out to something like 90% enrollment regionally, nationally, or globally. In looking at Uganda (Figure 3) we can credibly infer that many areas have achieved <20% of women with 0 years of education in the most recent period, but large areas of the country also remain significantly and persistently behind on this goal. This specifically quantifies the danger vocalized by many education advocates that aggregate progress is simply driven by progress in select urban areas of the country that may already enjoy traditionally higher sociodemographic privileges, more support for accessing public education, etc. This sort of interrogation of global or national aggregates, which tend to be the focus of international agendas, will be critical in the SDG era.

In considering the spatial distribution of average education, it's important to consider geographic segregation not just in the cross-sectional outcome but also in progress over time. Figure 4 extends the probability threshold summarization to examine how progress relative to the national average progress is geographically distributed over time. If the progress observed nationally was equally distributed, every pixel would be white in Figure 4. However, we can infer that some areas of the country are progressing much faster than the national average while some are stagnating or moving in the wrong direction. In the case of mean years of education it seems progress is distributed rather equally in Uganda with the exception of the northeast region. However, progress in reducing the proportion of women with 0 years of education is much more heterogeneous across the country. This difference can highlight that in some areas where the mean is increasing over time but the proportion of women with 0 years is also increasing, it's possible that women already enrolled are staying in school longer but the population never attending school is simultaneously growing. If the resulting balance is that the average years of education in Uganda increases anyway, this national statistic obscures the nuance of who is actually experiencing those gains.

Progress being distributed unequally isn't necessarily a normatively undesirable consequence; this could be the result of progress being distributed *equitably*, where the poorest performing areas in the baseline period were more heavily targeted by interventions and saw faster progress than areas that were already high-performing. On the other hand, progress could be unequally distributed in such a way that urban areas that were already high-performing at baseline saw huge increases over the time period, boosting the national average despite larger areas that stagnated or deteriorated.

GOAL: mean years of education > 6



GOAL: < 20% of women with 0 years of education

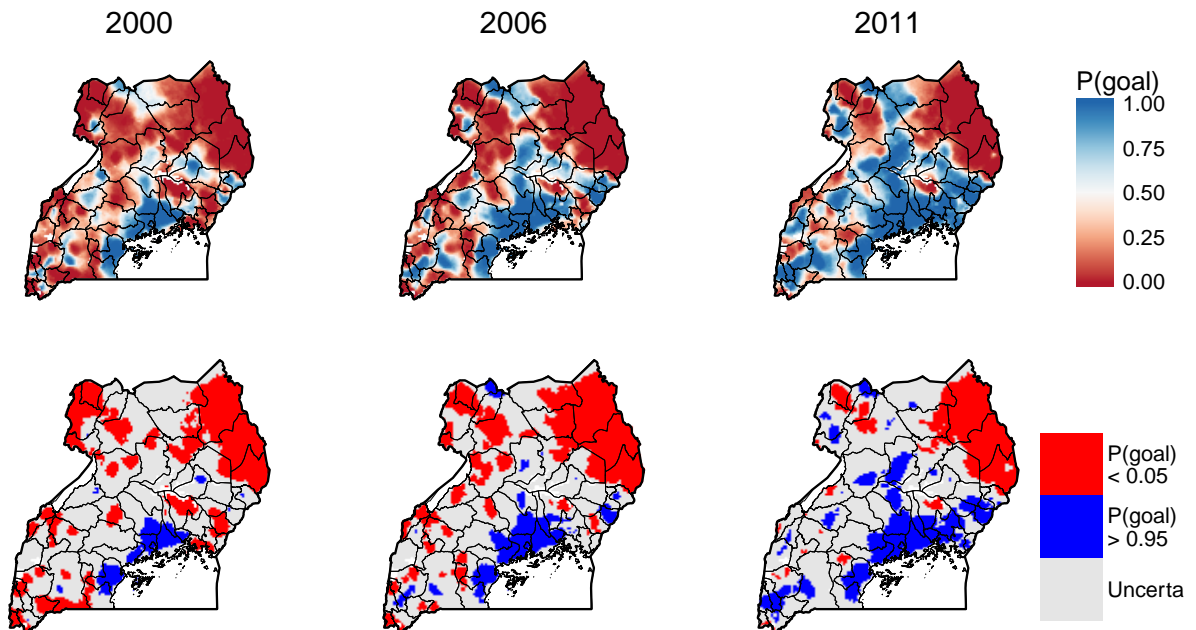


Figure 3: Probability threshold surface for Uganda relative to fixed goals. Top and bottoms rows show continuous probability and probability masked to $\alpha < 0.05$, respectively.

V. Discussion

This analysis provides a formal structure for evaluating the geographic segregation of levels of education and progress made in these three countries. Nigeria shows the most stark disparities in basic educational attainment between the north and south regions. Despite boasting the fourth highest rate of growth in GDP over the last decade, the EFA Global Monitoring Report (GMR) states that “corruption, conflict and a lack of investment has resulted in Nigeria having one of the worst education systems in the world.”¹² It is clear that certain areas in the urban south are making progress, which could inflate national average statistics by itself. However, a huge area in mostly northern Nigeria is being left behind, with levels either stagnating or decreasing. The Guardian writes “a lack of focus on the marginalized has left the poorest five times less likely to complete a full cycle of primary education than the richest and over a third of out-of-school children living in conflict affected zones.” The President of the Global University Network for Innovation (GUNI) Africa remarked “we have this rather poor culture of accurate data collection in a timely manner. On most global statistical annexes, Nigeria’s data are always outdated or not available.”¹² Moving into the SDG era of benchmarking and measurement, Nigeria represents the quintessential case where measuring progress towards goals with a national education statistic will not be sufficient to capture subnational inequity in both level and progress towards improving education.

In Kenya there is concern that the large population of nomadic children in the northeastern regions is being left behind.¹³ Programs targetting nomadic groups and other rural populations have been extremely limited in scale and underfunded. The other large vulnerable population is children living in urban slums; over 40% of the poorest children attend private schools in Kenyan slums. This year a Kenyan news outlet remarked that “it is a sad indictment that in these areas, NGOs are more visible than the government in providing access to basic education.”¹³

Government schools in Uganda, which primarily serve children from poor families, implemented a Universal Primary Education (UPE) scheme in 1997 as part of a national policy to provide free primary education for underprivileged children.⁵ The country was celebrated by global education advocates for achieving 90% of MDG2 (related to children completing primary school), but the UPE initiative has seen mixed results since. Dropout rates have increased as the financial burden has increasingly fallen on parents in poorer districts, where they are asked to provide fees for school supplies or lunches. This lack of basic supplies also disproportionately affects young girls in a culture that has a complicated relationship with menstruation. One NGO estimated that 30% of girls leave school when they start their periods, often because of a lack of sanitary pads.¹⁴ The present analysis corroborates the contention of many education advocates in Uganda that the northern and eastern regions suffer the most from this deterioration of basic funding coupled with concentrated rural poverty.⁵

This is the first analysis to apply Bayesian MBG methodology to geolocated cluster data to estimate maternal education indicators at a high degree of spatial granularity. This analysis leverages all available spatial and temporal information in the data structure rather than aggregating over this information. This method, coupled with effective and responsible visualization techniques, can allow for more sophisticated interrogation of national average social exposures. This will be critical moving into the new SDG era, as it is crucial that externally-imposed goals focus on the whole distribution within a country and don’t serve to inadvertently entrench existing inequities by incentivizing a certain pattern of investment that is effective in increasing the national average but may leave millions in the same place or worse.

As data and methods become increasingly sophisticated, this analysis also highlights the

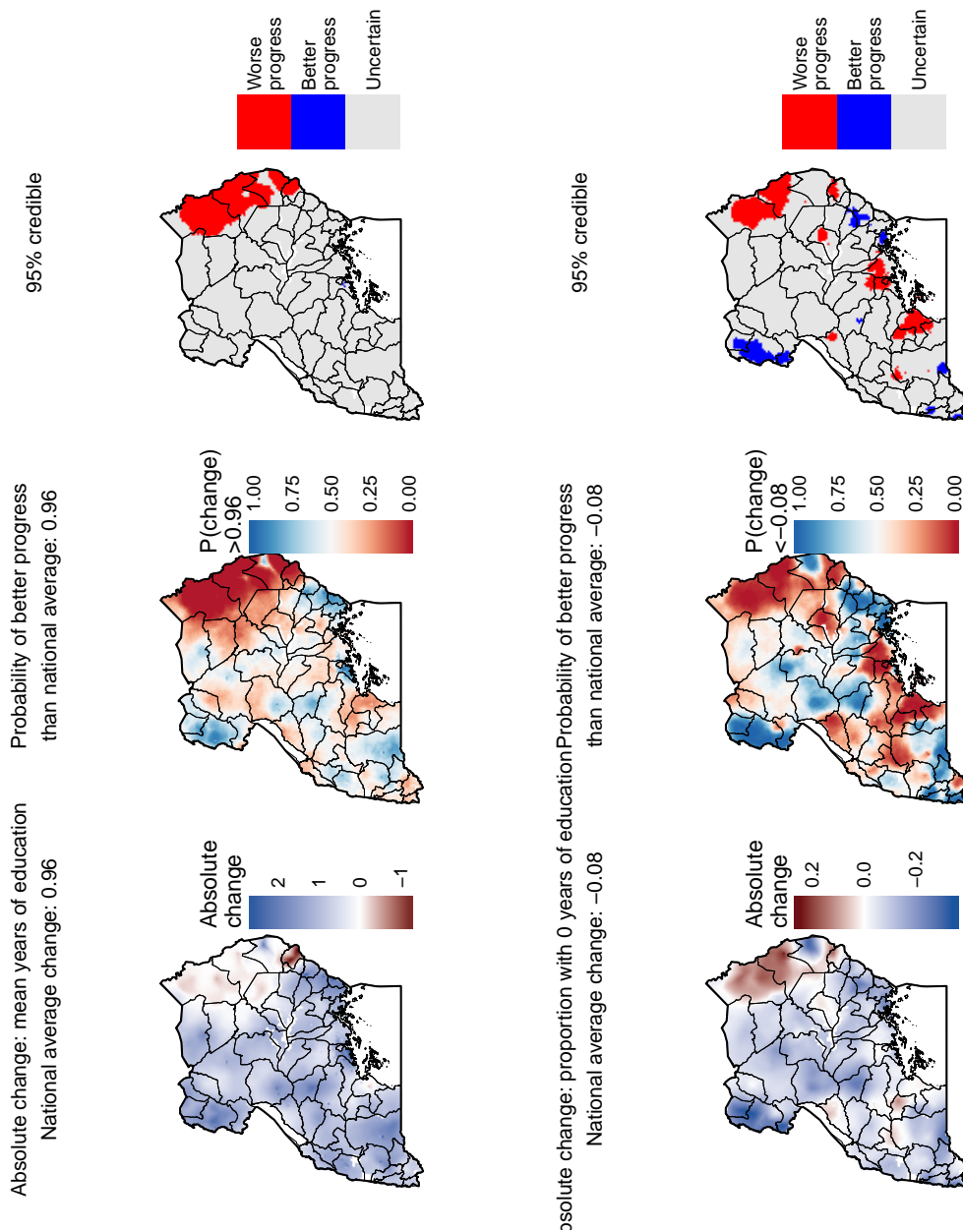


Figure 4: Probability threshold surface for Uganda relative to the absolute national population-weighted progress over the period. The top and bottom rows show progress in mean years of education and proportion of women with 0 years, respectively.

importance of interpreting full models with uncertainty; increasing the granularity of estimates allows us to make precise inferences about the true geographic distribution of education over certain localized areas, but huge subnational areas remain uncertain. In interpreting uncertainty in the probability threshold maps (Figures 3 and 4), it is important to consider the maps of data coverage as these maps may highlight the need for more data to make a conclusion relative to a goal. Though we can say with a high degree of certainty that areas in red have not met the goal and therefore may warrant a higher level of investment, it is important not to interpret the grey “uncertain” areas as adequate, i.e. performing somewhere *between* the blue and red areas. This may be the case, but it may be equally likely that some of these areas are the worst performing in the country and we simply don’t have the data required to make a reliable inference about the true level of education there. In this scenario, disinvesting in grey areas for the sake of investment in red areas may be misguided and further entrench existing social inequities (i.e. poor data coverage and thus poor precision in our framework may be spatially correlated with other sociodemographic disadvantage, lower national investment, etc.). The broadest recommendation to take away from these results is that red areas require increased investment in interventions to increase educational access and attainment, grey areas require increased data collection and surveillance efforts, and blue areas are largely high-performing relative to the rest of the country. However, this should be qualified by the fact that *within* a 5x5 km pixel, especially in dense urban settings, there may be a high degree of inequality even within blue pixels that is distributed across very localized space, ethnic/racial/religious demographics, etc. We are not capturing that in this analysis, which only seeks to inform policy decisions or benchmarking that happens across physical space or administrative unit.

Even without perfect data coverage, new advances in the field of MBG can offer a greater degree of granularity in assessing social exposures related to health and how patterns of progress over time are spatially distributed. Future research will focus on disentangling the nature of the relationship between health and maternal education below the level of national averages. An important topic will be exploring the best way to operationalize inequality spatially and its relationship with health. How does the relationship between spatial inequality and health differ from traditional non-spatial metrics, such as examining the Gini coefficient of all pixel means within a country? Does spatial segregation of mean educational attainment differ from person-level inequality? Formalizing and incorporating this field of metrics into international health goals will need to be thorough and intentional, both from a robust statistical perspective as well as reckoning with the various normative preferences inherent in selection of inequality criteria.¹⁵

Despite less than ideal coverage of geolocated cluster data in space and time, new advances in the field of MBG offer a robust, efficient modeling framework for estimating social exposures related to health and how patterns of progress over time are spatially distributed at a granular level. As it is recently developed, there are many exciting branches evolving in the MBG field. It will be worthwhile to explore and test all available model-fitting software for the current SPDE estimation besides R-INLA. Template Model Builder (TMB) is one alternative that needs to be further explored, as the flexibility it offers in explicitly specifying and constraining different likelihoods is appealing for non-traditional, complex datasets.¹⁶ This flexibility may also prove crucial in expanding these analyses to all of Africa and exploring solutions for incorporating not just geolocated clusters (“point data”) but also aggregates over administrative units (“polygon/aerial data”). This polygon data represents a huge amount of information that we are not leveraging under the current framework, particularly census data from the IPUMS survey family. This analysis could be expanded to all of Africa, using this much larger pool of data represented in Figure 5. This would also allow a more sophisticated structuring of the temporal component in the MBG framework, rather than the AR1

term being independent for each country as it is currently.

It is important to clarify this analysis and its results with several important limitations. The resulting summary maps for Nigeria demonstrate a huge discrepancy in level of education and progress over time between the north and south regions. This is explained at least in part by the heterogeneity in terms of *type* of education provided throughout Nigeria. The only way to measure and compare the *quantity* of education between countries and across time is to convert to the UNESCO single-year education standards, as we have done with all the DHS data used in this analysis. This counts responses associated with “religious education” as 0 years. In Nigeria, the North is largely Muslim and in wide areas girls may only attend religious school.⁴ Though Nigeria may be the most stark example of a divide on religious education, it is important to not attribute the immense education gap to the religious divide alone. There has been a long history of economic and political conflict that has rarely centered around religion - the South is much richer and has access to extensive oil reserves in the Niger Delta, while the economy of the North is one of the poorest performing in the world due to deindustrialization and lack of investment in agriculture and infrastructure.¹⁷

Related to type of education is the larger limitation of this analysis not being able to inform goals that are based largely on *quality* of education. This has been the focus on many broad UNESCO goals, especially in their Education for All initiative.¹ They discuss how more children may be going to school but what do levels of basic literacy look like? How are those levels distributed across space and populations? A recent report from UNICEF stresses that “almost two in five who do finish primary school have not learned how to read, write or do simple arithmetic.”³ An analysis based on quality of education would be a powerful aid to evaluating progress on these types of goals, but a quantitative analysis across time and space would be extremely difficult based on the a lack of comparability across the myriad indicators of quality. These sorts of analyses require reconciling differences in measurement and reporting across specific schools, and have been immensely varied and controversial even in the United States, despite the massive amount of reporting there relative to developing countries.

What we are trying to measure with the present analysis is *basic* access to education as a powerful predictor of child and maternal mortality. This pathway may not necessarily be through traditional measures of education quality like literacy and arithmetic scores. There is evidence that simply going to school and getting out of the home is protective, may increase age of marriage, increase age of first child, increase access to and knowledge of family planning resources, etc.¹⁸ It is a completely different question to say once a certain number of average years of education are attained in all areas within a country, what does the percent literate look like? This will vary widely across space and population, and those types of inequities will be important to interrogate. But they rely on a baseline, comparable index of basic educational access in geographies where data is sparse; this is the gap the present analysis seeks to improve as household data on basic social indicators becomes more granular in time and space.

This study highlights concerns for equity in international goals and the importance of leveraging all spatial information for social indicators that have been proven to be spatially correlated. If there is not a robust system for measuring indicators of education access as SDG4 campaigns gain steam, we risk perceiving progress at the national aggregate as an unmitigated achievement, when in reality such progress and investment could be extremely concentrated and further entrenching social inequities within a country.



Figure 5: Data coverage for all usable DHS and IPUMS education data in Africa.

References

- 1UNESCO. Education 2030. *World Education Forum 2015* 2015. <https://en.unesco.org/world-education-forum-2015/incheon-declaration>.
- 2United Nations. Goal 4: Ensure inclusive and quality education for all and promote lifelong learning. 2016 <http://www.un.org/sustainabledevelopment/education/>.
- 3UNICEF. The State of the World's Children 2015: Executive Summary. UNICEF, 2014 http://www.unicef.org/publications/index\protect\T1\textbraceleft/_\protect\T1\textbraceright77928.html.
- 4Nnebe M. Nigeria's Educational Imbalance... The North South Divide. *Nigeriaworld*. 2014. <http://nigeriaworld.com/feature/publication/nnebe/011414.html> (accessed July 5, 2016).
- 5Mwesigwa A. Uganda's success in universal primary education falling apart. *Kampala: The Gaurdian*, 2015 <https://www.theguardian.com/global-development/2015/apr/23/uganda-success-universal-primary-education-falling-apart-upe>.
- 6Gething P, Tatem A, Bird, Burgert-Brucker CR. Creating spatial interpolation surfaces with DHS data. *DHS Spatial Analysis Reports No 11* 2015. <http://dhsprogram.com/pubs/pdf/SAR11/SAR11.pdf>.
- 7Pigott DM, Howes RE, Wiebe A *et al.* Prioritising Infectious Disease Mapping. *PLOS Neglected Tropical Diseases* 2015; **9**: e0003756.
- 8Patil AP, Gething PW, Piel FB, Hay SI. Bayesian geostatistics in health cartography: The perspective of malaria. *Trends in Parasitology* 2011; **27**: 246–53.
- 9Hay SI, Battle KE, Pigott D *et al.* Global mapping of infectious disease. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 2013; **368**: 20120250.
- 10UNESCO. ISCED Mappings. 2016. <http://www.uis.unesco.org/Education/ISCEDMappings/Pages/default.aspx> (accessed Aug 2, 2016).
- 11Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology* 2013; **7**: 39–55.
- 12Eno-Abasi. Nigeria misses 2015 global education goals. *The gaurdian*. 2015. <http://guardian.ng/features/education/nigeria-misses-2015-global-education-goals/> (accessed Aug 1, 2016).
- 13Daily Nation. Kenya has made progress, but is yet to achieve goal of education for all. *Daily nation*. 2016. <http://www.nation.co.ke/oped/Opinion/Universal-Education-Goals-Kenya-Government/-/440808/2713344/-/gfflrf/-/index.html> (accessed Aug 1, 2016).
- 14Vision Reporter. 30% of girls leaving school for lack of sanitary pads. *NewVision*. 2013. http://www.newvision.co.ug/new\protect\T1\textbraceleft/_\protect\T1\textbracerightvision/news/1328198/-girls-leaving-school-lack-sanitary-pads (accessed Aug 1, 2016).
- 15Gakidou E, King G. International Journal for Equity in Health Measuring total health inequality: adding individual variation to group-level differences. *International Journal for Equity in Health World Health Organization* 2002.
- 16Kristensen K. Template Model Builder: A General Random Effect Tool Inspired by ADMB [R package TMB version 1.7.0].

17Campbell J. Why Nigeria's North South Distinction is Important. Huffington post. 2011. http://www.huffingtonpost.com/amb-john-campbell/why-nigerias-north-south-protect\T1\textbraceleft/_\protect\T1\textbracerightb\protect\T1\textbraceleft/_\protect\T1\textbraceright817734.html (accessed July 5, 2016).

18Cleland J. The benefits of educating women. *The lancet*. 2010; **376**: 933–4.