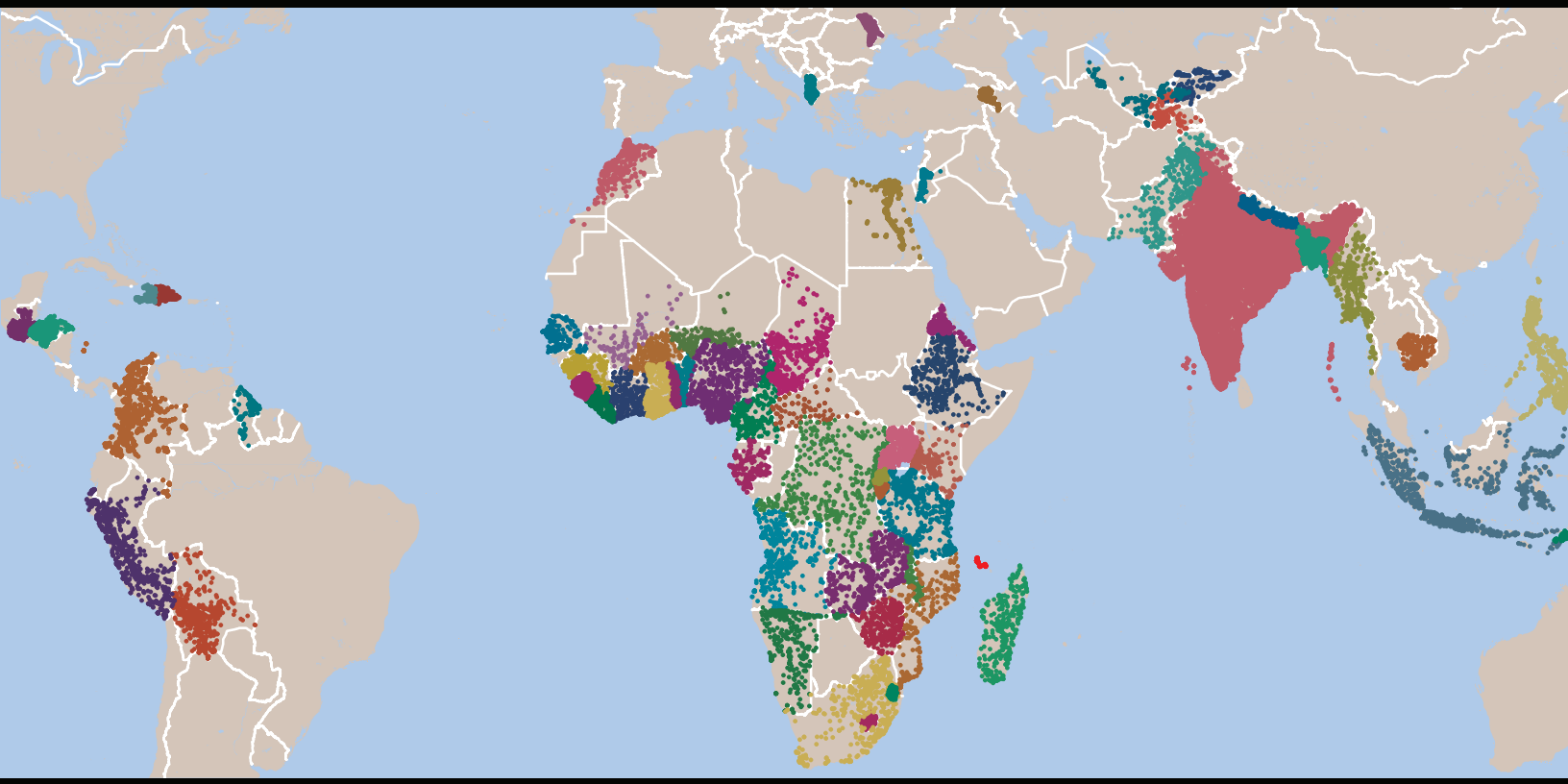
USAID
FROM THE AMERICAN PEOPLE

INTERPOLATION OF DHS SURVEY DATA AT SUBNATIONAL ADMINISTRATIVE LEVEL 2

DHS SPATIAL ANALYSIS REPORTS 17



September 2019

This publication was produced for review by the United States Agency for International Development (USAID). The report was prepared by Benjamin K. Mayala, Trinadh Dontamsetti, Thomas D. Fish, and Trevor N. Croft.

DHS Spatial Analysis Reports No. 17

Interpolation of DHS Survey Data at Subnational Administrative Level 2

Benjamin K. Mayala
Trinadh Dontamsetti
Thomas D. Fish
Trevor N. Croft

The DHS Program
ICF
Rockville, Maryland, USA

September 2019

Corresponding author: Benjamin K. Mayala, The DHS Program, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: 301-572-0507; fax: 301-407-6501; email: Ben.Mayala@icf.com

Acknowledgments: The authors wish to thank Jon Wakefield for his review of this report.

Editor: Diane Stoy
Document Production: Natalie Shattuck

This study was implemented with support provided by the United States Agency for International Development (USAID) through The DHS Program (#720-OAA-18C-00083). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

The DHS Program assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. Additional information about The DHS Program can be obtained from ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850 USA; telephone: +1 301-407-6500, fax: +1-301-407-6501, email: info@DHSprogram.com, internet: www.DHSprogram.com.

Recommended citation:

Mayala, Benjamin. K., Trinadh Dontamsetti, Thomas. D. Fish, and Trevor N. Croft. 2019. *Interpolation of DHS Survey Data at Subnational Administrative Level 2*. DHS Spatial Analysis Reports No. 17. Rockville, Maryland, USA: ICF.

CONTENTS

TABLES

FIGURES

PREFACE

The Demographic and Health Surveys (DHS) Program is one of the principal sources of international data on fertility, family planning, maternal and child health, nutrition, mortality, environmental health, HIV/AIDS, malaria, and provision of health services.

The DHS Spatial Analysis Reports supplement the other series of DHS reports that respond to the increasing interest in a spatial perspective on demographic and health data. The principal objectives of all the DHS report series are to provide information for policy formulation at the international level and to examine individual country results in an international context.

The topics in this series are selected by The DHS Program in consultation with the U.S. Agency for International Development. A range of methodologies are used, including geostatistical and multivariate statistical techniques.

It is hoped that the DHS Spatial Analysis Reports series will be useful to researchers, policymakers, and survey specialists, particularly those engaged in work in low- and middle-income countries, and will be used to enhance the quality and analysis of survey data.

Sunita Kishor
Director, The DHS Program

ABSTRACT

Over the last several years and within the framework of the Sustainable Development Goals, there has been a need to improve the measurement and understanding of local geographic patterns to support more decentralized decision-making and more efficient program implementation. This requires more disaggregated data that are not currently available in a nationally representative household survey. The spatial modeling techniques that leverage existing survey data, spatial relationships between survey clusters, and relationships with geospatial covariates have become increasingly popular for mapping key development indicators at high spatial resolution.

This study explores the potential of model-based geostatistics methodology to model DHS survey indicators. We implement a stacked ensemble modeling approach that combines multiple model algorithmic methods to increase predictive validity relative to a single modeling. The approach captures potentially complex interactions and non-linear effects among the geospatial covariates. Three submodels are fitted to six DHS indicator survey data using the geospatial covariates as exploratory predictors. The model prediction surfaces generated from the submodels are used as covariates in the final Bayesian geostatistical model, which is implemented through a stochastic partial differential equation approach in the integrated nested Laplace approximations. To explore the ability of our modeling approach to estimate indicators below the first subnational level, pixel-level estimates generated from the Bayesian model were aggregated to the second subnational level by using the population-weighted average within the administrative boundary.

Results of the individual submodels vary spatially, which is explained by the uncertainties in the individual model algorithm. The use of an ensemble model approach seems more adequate than relying on predictions from any single modeling method. We demonstrate the predictive ability of the model at the second administrative level using cross-validation. The results indicate good predictive performance.

The proposed approach can help to inform the allocation of resources and program implementation in areas that need more attention. Countries can use this approach to model other DHS survey indicators at much smaller spatial scales.

ACRONYMS AND ABBREVIATIONS

ADMIN 1	first subnational administrative level
ADMIN 2	second subnational administrative level
AIS	AIDS Indicator Survey
ANC	antenatal care
DHS	Demographic and Health Survey
DPT3	diphtheria-tetanus-pertussis
EVI	enhanced vegetation index
GAM	generalized additive model
GBM	gradient boosted trees
GPS	global positioning system
HMIS	health management information system
INLA	integrated nested Laplace approximation
LASSO	least absolute shrinkage and selection operator
LST	land surface temperature
MAE	mean absolute error
MBG	model-based geostatistics
MCMC	Markov Chain Monte Carlo
ME	mean error
MIS	Malaria Indicator Survey
PET	potential evapotranspiration
RMSE	root, mean, squared error
SDG	Sustainable Development Goal
SPDE	stochastic partial differential equations
WHO	World Health Organization
WHZ	weight-for-height z score

1 BACKGROUND AND OBJECTIVES

1.1 Background

The Demographic and Health Surveys (DHS) Program has been a leader in collecting and providing cluster-randomized survey data on various development and health indicators. In addition to the standard open-source data files in which household and individual survey results can be tabulated by first-order subnational units (states/provinces or regions) and urban/rural strata, more surveys are now providing georeferenced data for individual clusters. The availability of the Global Positioning System (GPS) coordinates for DHS, the Malaria Indicator Survey (MIS), and the AIDS Indicator Survey (AIS) clusters provides highly local scale information that can be linked with survey outputs for quantifying demographic and health status heterogeneities and inequities.

During the last several years and within the framework of the Sustainable Development Goals (SDGs), there has been an expressed need to improve the measurement and understanding of local geographic patterns in order to support more decentralized decision-making and more efficient program implementation (United Nations General Assembly 2015). This requires additional disaggregated data that are not currently available in a nationally representative household survey.

Analyses of the DHS survey indicators are conducted primarily at the national level, but also at the first subnational administrative level (ADMIN 1). Since estimates produced at the national level are more useful for making comparisons between nations and aggregating across large world regions, their natural audience includes international policymakers and donors (Li et al. 2019). However, the ADMIN 1 analysis does not provide comprehensive estimates at lower levels, such as the second subnational administrative level (ADMIN 2), where health programs are designed and implemented.

To better address the need for fine spatial and lower level estimates, there are three possible options:

(1) Scaling-up the nationally representative survey data collection process by increasing the sample size, survey costs, and survey time to create a representative sample at the desired administrative level.

(2) Using data derived from routine health management information systems (HMIS) from health facilities, communities, census, or other household surveys, such as data that can determine the vaccination coverage in a district.

(3) Creating spatially interpolated maps that use modeling techniques to predict values at non-surveyed locations.

Increasing DHS survey sample size to enable increased geographic disaggregation is both time consuming and expensive. Thus, the first option may not be feasible in an increasingly resource-constrained environment. With the second option, HMIS data quality is not always reliable, and the data are not easily accessed. The third option, which uses spatial modeling techniques that leverage existing survey data, spatial relationships between survey clusters, and relationships with geospatial covariates, has become

increasingly popular in mapping key development indicators at high spatial resolution (Utazi et al. 2018; Gething and Burgert-Brucker 2017).

The Bayesian spatial approach is increasingly recognized as an excellent geostatistical analysis method for addressing uncertainty in the model estimates (predictions) and for being flexible and capable of handling missing data (Cressie and Wikle 2011). This approach has been widely used to predict and map various indicators such as those described in SAR 11 (Gething et al. 2015), poverty (Steele et al. 2017), and malaria (Gething et al. 2011; Gosoniu et al. 2006; Gosoniu et al. 2012; Gosoniu, Veta, and Vounatsou 2011; Kazembe et al. 2006; Raso et al. 2012, Riedel et al. 2011; Hay et al. 2009). In these studies, environmental data layers that are thought to influence the indicators are used to explain some of the variations in prevalence across different areas. This can aid our understanding of the relationships between the indicator and the influence of climatic/environmental and socioeconomic factors (Noor et al. 2009).

The Markov Chain Monte Carlo (MCMC) algorithms have been the most common method for making Bayesian statistical inferences with generalized linear geostatistical models (GLGM) (Gilks, Richardson and Spiegelhalter 1996). The MCMC has been developed for model estimation, but can be computationally expensive, especially with big data. There has been a recent increase in the application of integrated nested Laplace approximation (INLA) methodology and software (http://www.rinla.org) in Bayesian spatial models (Rue, Martino, and Chopin. 2009). The choice of this method over MCMC is based on the speed of calculation and the ease with which model comparison can be performed (Rue, Martino, and Chopin. 2009).

1.2 Objectives

In this report, we explore the potential of model-based geostatistics (MBG) methodology (described in the methods section) to model DHS survey indicators. More specifically, we use the stacking and ensemble model approaches to predict the indicators at a high resolution gridded pixel level, produce estimates at the ADMIN 2 level, and measure the uncertainty of estimates. The INLA methodology is used to create a model for predicting the indicators based on the different geospatial covariates and spatially correlated random effects, and to produce prediction maps. The report will develop R code structure and workflow for routine interpolation of survey data at the second subnational administrative level (ADMIN 2).

2 SELECTION OF DHS INDICATOR

The six indicators in our analysis were extracted from two national DHS surveys, the Kenya 2014 DHS and Ethiopia 2016 DHS. We considered data only from those surveys that had described the total number of individuals examined, the proportion of positive cases, and the coordinates of their geographical locations. From these surveys, we obtained 1,583 and 620 point clusters for Kenya and Ethiopia, respectively. Table 1 describes the indicators we modeled.

Table 1 Description of DHS indicators used in the study

Indicator	Definition
Antenatal visits for pregnancy: 4+ visits	Percentage of women who had a live birth in the 5 years before the survey who had 4+ antenatal care visits
Stunting in children	Percentage of children under age 5 stunted (below -2 SD of height-for-age according to the WHO standard)
Wasting in children	Percentage of under 5 children with a weight-for-height z-score (WHZ) more than two SD below the median WHO growth standards
Population living in household with an improved water source	Percentage of the de jure population living in households whose main source of drinking water is an improved source
Women age 15-49 with any anemia**	Percentage of women classified as having any anemia (<12.0 g/dl for non-pregnant women and <11.0 g/dl for pregnant women)
Diphtheria-tetanus-pertussis (DPT3) received	Percentage of children age 12-23 months who had received a third DPT dose

** Data not collected for the Kenya DHS 2014.

3 METHODS

A modeling framework for generating standardized modeled surfaces using DHS survey data has been described in SAR 11 (Gething et al. 2015) and SAR 14 (Burgert 2014). In this analysis, we employed a new geospatial modeling approach similar to that used in recent mapping of child growth failure (Osgood-Zimmerman et al. 2018), education attainment (Graetz et al. 2018), vaccine coverage (Mosser et al. 2019), HIV (Dwyer-Lindgren et al. 2019), exclusive breastfeeding (Bhattacharjee et al. 2019), and childhood diarrheal diseases (Reiner et al. 2018). We adopted this method because it has been shown to improve the prediction accuracy based on the stacked generalization that allows for multiple, non-linear algorithmic mean functions to be embedded within a Gaussian process framework (Bhatt et al. 2017). We detail this approach in the next sections.

3.1 Geospatial Covariates

To model the DHS indicators, we assembled environmental and socioeconomic geospatial covariate data layers, which were obtained from publicly available remote sensing sources. These data included access (travel time to nearest settlement), aridity, diurnal temperature range, precipitation, potential evapo-transpiration (PET), daily maximum temperature, elevation, enhanced vegetation index (EVI), daytime land surface temperature, diurnal difference in land surface temperature, night land surface temperature (LST), and population categories (children under age 5, women age 15 to 49, and total population). Further description of each covariate can be obtained from the DHS Geospatial Covariate Report (Mayala et al. 2018).

The geospatial covariates were selected for their potential to predict DHS indicators, and they have previously been shown to correlate with the development of indicators in different settings (Alegana et al. 2015, Gething et al. 2015). Table 2 describes the spatial and temporal resolution of each geospatial covariate and the sources.

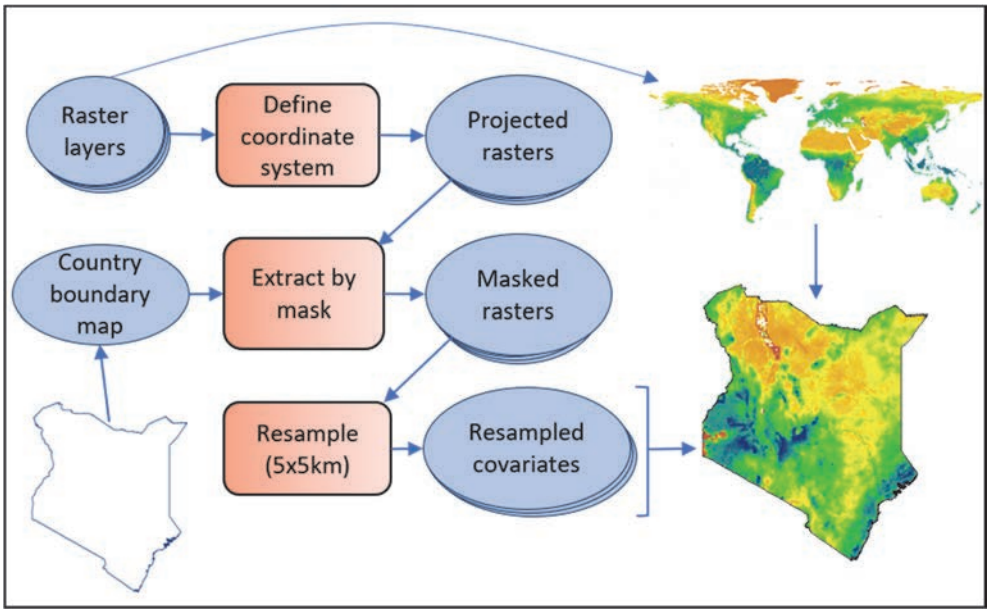
Table 2 Geospatial covariates used to develop the models in this study

Covariates	Spatial resolution	Temporal resolution	Source
Travel time to nearest settlement >50,000 inhabitants	5x5 km	Static	https://map.ox.ac.uk/research-project/accessibility_to_cities/
Aridity	10x10 km	Annual	http://wps-web1.ceda.ac.uk/ui/home
Diurnal temperature range	10x10 km	Annual	http://wps-web1.ceda.ac.uk/ui/home
Precipitation	10x10 km	Annual	http://wps-web1.ceda.ac.uk/ui/home
Potential evapotranspiration (PET)	10x10 km	Annual	http://wps-web1.ceda.ac.uk/ui/home
Daily maximum temperature	10x10 km	Annual	http://wps-web1.ceda.ac.uk/ui/home
Elevation	1x1 km	Static	http://webmap. ornl.gov
Enhanced vegetation index (EVI)	5x5 km	Annual	https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod13a1
Daytime land surface temperature (LST)	5x5 km	Annual	http://wps-web1.ceda.ac.uk/ui/home
Diurnal difference in LST	5x5 km	Annual	http://wps-web1.ceda.ac.uk/ui/home
Nighttime LST	5x5 km	Annual	http://wps-web1.ceda.ac.uk/ui/home
Population distribution	1x1 km	Annual	http://www.worldpop.org.uk/data/get_data/

3.1.1 Geospatial covariates processing

The geospatial covariate data layers used in this analysis were acquired from a myriad of data sources, and therefore have different spatial references, projections, extents, and dimensions. For example, gridded population data and elevation had a 1 x 1 km spatial resolution, EVI was at 5 x 5 km, and temperature range at 10 x 10 km resolution. We used the 'raster' and 'shapefiles' packages in the R software (R Core Team 2018) to (1) re-project to the same coordinate reference system (the standard-based World Geodetic System 1984), (2) crop and mask to an extent encompassing the boundaries of the study area and (3) resample with bilinear interpolation to the same spatial resolution used in the modeling, 5 x 5 km (Figure 1).

Figure 1 The conceptual framework of processing the geospatial covariates, using Kenya as an example



3.2 Geostatistical Model

3.2.1 Overview of the modeling approach

Figure 2 depicts the geospatial modeling framework used for modeling DHS indicators and the underlying covariates and producing the gridded pixel and subnational level estimates. The approach involved the following steps:

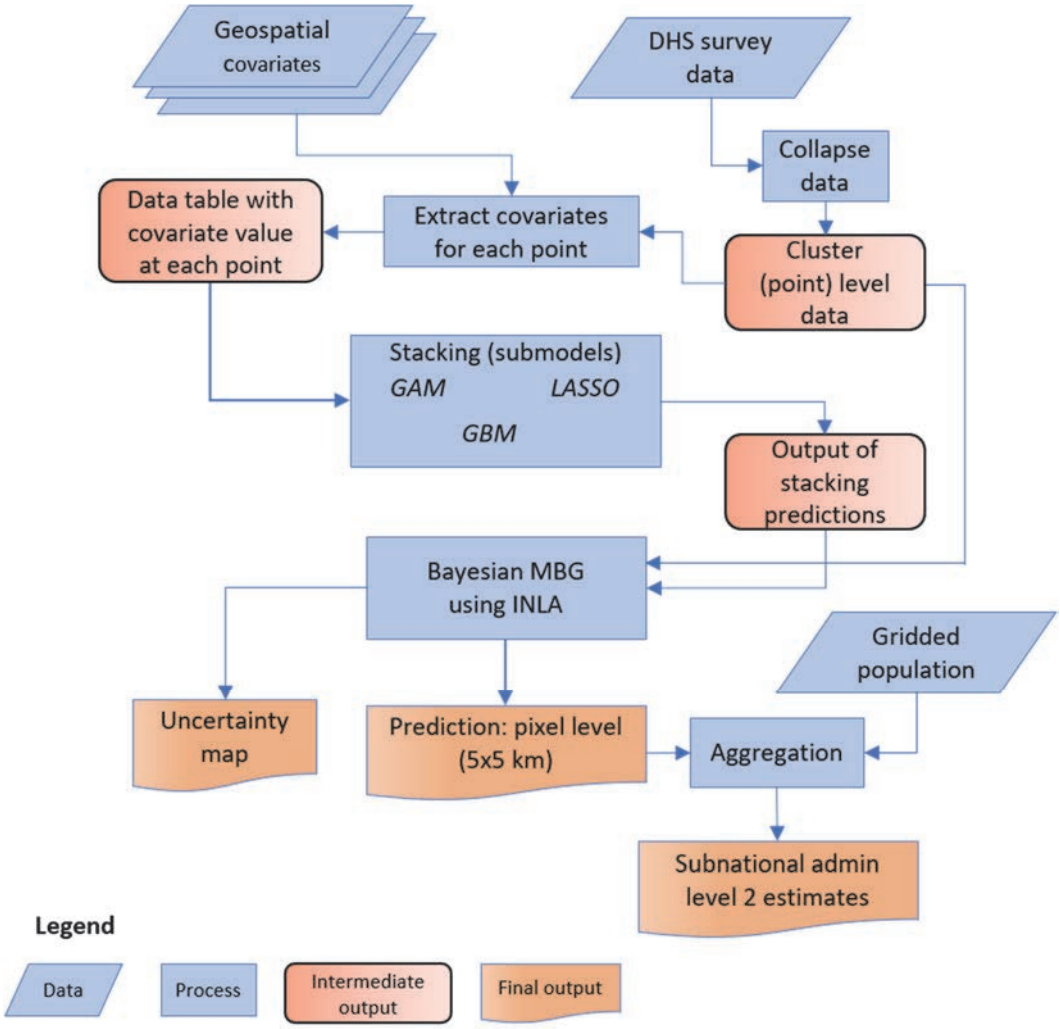
Step 1 - We summarized the individual-level DHS survey data to the finest spatial resolution (latitude and longitude) that represented the location of the survey cluster.

Step 2 - The processed geospatial covariates (from the previous section) and the cluster (point) level data were imported into the R environment for statistical computing (R Core Team. 2018). We then applied the 'raster' package to extract the corresponding covariate pixel values at each survey cluster point.

Step 3 - The point level data (from Step 2) and their associated geospatial covariates were used in the stacked (submodels) generalization ensemble model (described in Section 3.2.2). The prediction surfaces generated from the stacked ensemble models were then used as covariates in the final geospatial (MBG) model. The outputs of the final model are pixel-level mean estimates with associated uncertainty at the 5 x 5 km resolution.

Step 4 - We aggregated the prediction output from the final model (Step 3) to the ADMIN 2.

Figure 2 Geospatial modeling flowchart



Note: Figure modified from Mosser et al. 2019 and Dwyer-Lindgren et al. 2019.

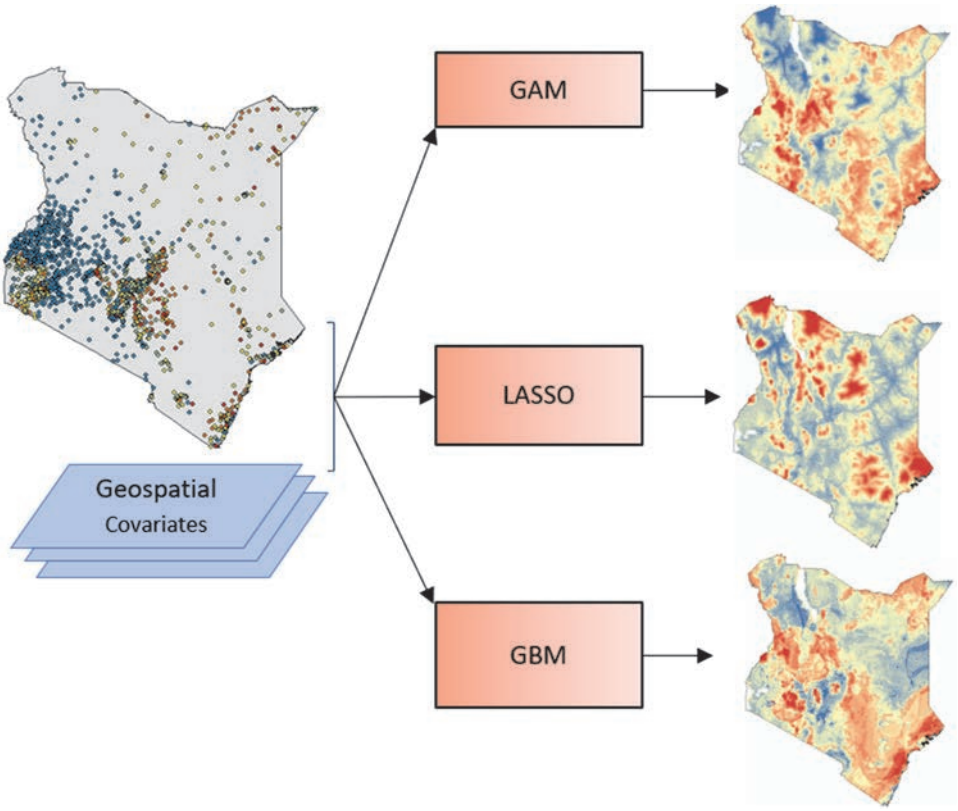
3.2.2 Covariate ensemble modeling using stacked generalization

Stacking (also called stacked generalization/regression) is an ensemble modeling approach that combines multiple model algorithmic methods to increase predictive validity relative to a single modeling approach. We employed this approach to capture the potential complex interactions and non-linear effects among the geospatial covariates (Figure 3). The ensemble approach has been shown to improve the predictive accuracy of the geostatistical models, as compared to prediction from any single method (Bhatt et al. 2017). Numerous recent studies have implemented the stacking approach to derive continuous estimated surfaces of indicators of interest from DHS household surveys. These include mapping of HIV prevalence (Dwyer-Lindgren et al. 2019), vaccine coverage (Mosser et al. 2019), exclusive breastfeeding (Bhattacharjee et al. 2019), child growth failure (Osgood-Zimmerman et al. 2018), education attainment (Graetz et al. 2018), and childhood diarrheal diseases (Reiner et al. 2018).

In our analysis, we fitted three submodels to each set of the selected DHS indicator survey data using the geospatial covariates (described in Table 2) as exploratory predictors. These include (1) GAM: generalized additive model (Wood 2017), (2) LASSO: least absolute shrinkage and selection operator regression (Zou and Hastie 2005) and (3) GBM: gradient - boosted trees (Friedman 2001). The submodels were implemented in R statistical for the computing environment using packages 'caret', 'mgcv', 'xgboost', and 'glmnet'. We selected these model algorithms because they have demonstrated high predictive accuracy in previous studies (Barbet-Massin et al. 2012; Elith and Graham 2009; Elith et al. 2006; Franklin 2009; Giovanelli et al. 2010; Lobo, Jiménez-Valverde and Hortal 2010; Mateo et al 2010; Pearson et al. 2007; Peterson et al. 2011; Phillips, Anderson, and Schapire 2006; Wisz and Guisan 2009).

To make better predictions and avoid overfitting, each submodel was fit using five-fold cross–validation, which generated the out-of-sample predictions that were included as exploratory geospatial covariates when fitting the geostatistical model. In addition, each submodel was fit with a full dataset, which produced the in-sample predictions that were then used as covariates when generating predictions from the full geostatistical model. A logit transformation of the predictions was used to place the out-of-sample and in-sample predictions on the same scale as the linear predictor in the geostatistical model. This process has been described in detail (Bhatt et al. 2017; Dwyer-Lindgren et al. 2019). We illustrate this process in Figure 3, which shows the predicted in-sample layers generated from the submodels using stunting in the Kenya DHS 2014 as an example.

Figure 3 **Example of in-sample predictions from the three submodels (generalized additive model, lasso regression, and gradient - boosted trees) fitted using the geospatial covariates and DHS data. These predicted surfaces are then used as covariates in the full geostatistical model**



3.2.3 Model specification and development

As described in the previous section, the ensemble modeling approach allows for non-linear relationships and interactions between the geospatial covariates to better predict the DHS indicators. Since the approach does not explicitly account for spatial patterns in the data, we used the Bayesian geostatistical modeling framework in our analysis to account for the spatial dependence.

For each indicator of interest, we modeled Y_i, the number of 'positive' individuals among those sampled at cluster location $s_i, i = 1, \ldots n$, using a binomial spatial regression with a logit link function (Banerjee, Carlin, and Gelfand 2014; Diggle and Giorgi 2019). Letting N_i be the total number of individuals sampled at cluster s_i, the model can be written as:

$$Y_i \sim Binomial(N_i, p_i)$$

$$logit(p_i) = \beta_0 + \beta \mathrm{X}_i + \omega_i + \varepsilon_i$$

$$\omega_i \sim GP(0, \Sigma)$$

Where:

- β_0 denotes the intercept,
- p_i is the probability, representing the underlying prevalence at cluster s_i,
- $X_i = \left(X_{i1}, X_{i2}, \ldots X_{im}\right)$ is the vector of logit-transformed covariates for location s_i obtained from the submodels (*GAM, LASSO, and GBM*), generated from the stacked ensemble covariate modeling (as described in Section 3.2.2),
- $\beta = (\beta_1, \beta_2, \ldots \beta_m)$ vector of regression coefficients on the submodels represent their respective predictive weighting and are constrained to sum to one (Bhatt et al. 2017),
- ω_i is a correlated spatial error term, accounting for spatial autocorrelation between data points, and
- $\varepsilon_i \sim \mathrm{N}\left(0, \sigma^2_{nug}\right)$ is an independent error term known as nugget effect.

The spatial error term ω_i is modeled as Gaussian process with a zero-mean and spatially structured covariance matrix $\sum$.

The spatial covariance $\sum$ was modeled using a stationary and isotropic Matérn function (Banerjee et al. 2014), given by:

$$\sum\left(s_i, s_j\right) = \frac{\sigma^2}{\Gamma(\lambda)2^{\lambda-1}}\left(\kappa d\left(s_i, s_j\right)^{\lambda} K_{\lambda}\left(\kappa d\left(s_i, s_j\right)\right)\right)$$

Where $d\left(s_i, s_j\right)$ is the distance between the two locations and $\sigma^2$is the spatial process variance. The term K_{λ} denotes the modified Bessel function of second kind and order λ, which measures the degree of smoothness. Conversely, κ is a scaling parameter related to the range r, that is the distance at which the

spatial correlation becomes almost null (i.e., smaller than 10%), and the definition for the range is given in equation below. See example by Lindgren et.al (2011) for detail description.

$$r = \frac{\sqrt{8\lambda}}{\kappa}$$

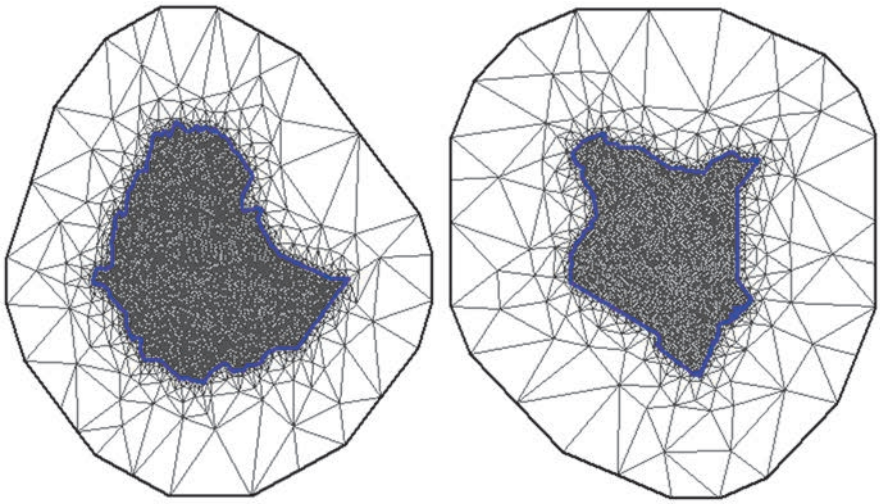
The Bayesian geostatistical model analysis was implemented through a stochastic partial differential equations (SPDE) approach in the recently developed INLA algorithm as applied in the R-INLA package (Rue, Martino, and Chopin 2009). This algorithm provides an effective estimation and spatial prediction strategy for spatial data by specifying a spatial data process as well as a spatial covariance function depending on locations and time points at which infection and covariate data are collected (Rue, Martino, and Chopin 2009). The INLA approach offers an advantage of providing accurate and fast results as compared to the MCMC algorithms, which are known to have problems of convergence and dense covariate matrices that increase the computational time. Thus, for large datasets, spatial and spatiotemporal estimation could lead to several days of computing time (Blangiardo and Cameletti 2015; Cameletti et al. 2013; Rue, Martino, and Chopin 2009).

The SPDE allow us to define a grid on spatial data by creating a constrained refined Delaunay triangulation (usually called mesh) over the study region. With this approach, observations are treated as initial vertices for the triangulation. Further vertices are then added or removed to satisfy triangulation quality constraints defined by three parameters: (1) cutoff, (2) offset and (3) maximum edge (Blangiardo and Cameletti 2015; Lindgren, Rue, and Lindstrom 2011).

A cutoff value was specified to avoid building too many small triangles around the clustered data locations. An offset value defined how far the mesh should be extended in the inner part (within areas where predictions are required) and the outer part (outside the area where predictions are required). The maximum edge value specified the maximum allowed edge length of the triangle in the inner domain and the outer extension. The inner maximum edge value was small enough to allow the triangulation to support functions with small enough features, and typically smaller than the spatial correlation range of the model (Lindgren, Rue, and Lindstrom 2011). Figure 4 provides an example of the finite mesh used for modeling.

As opposed to the regular grid, this approach is more dense in regions where there are more observations and consequently generates more information. Another advantage is that this approach saves computing time because prediction locations are typically much lower in number than those in a regular grid.

Figure 4 **INLA mesh triangulation for Ethiopia (left) and Kenya (right). The larger triangles show the buffer region surrounding the modeling region (maximum triangle edge length of 5.0 degrees, while the finer inner mesh overlays the modeling region (maximum triangle edge length of 0.1 degrees). The simplified polygon used to define the modeling county boundary is shown in blue.**



3.2.4 Pixel level model estimates

The prediction surfaces generated from the ensemble submodels (described in Section 3.2.2) were used as input covariates in the geostatistical models implemented in INLA. The final estimates (and uncertainty) for each indicator were generated by taking $k = 1,\ldots 1000$ samples from the posterior predictive distribution. Pixel level estimates that covered the modeling country (Kenya and Ethiopia) were produced at a high spatial resolution of 5 x 5 km.

3.2.5 Model estimates at admin level 2

In addition to the 5 x 5 km pixel level estimates, we overlaid the prediction prevalence surfaces (from Section 3.2.4) with the relevant population layer (children under age 5, women age 15 to 49, and total population) for each indicator we modeled. We then constructed estimates of each indicator at the second subnational administrative level by calculating population-weighted averages of prevalence for all grid cells within a given administrative boundary. The procedure was performed for each of the 1,000 posterior predictive samples with final point estimates derived from the mean of these draws and uncertainty intervals from the 2.5 and 97.5 percentiles.

3.2.6 Model validation

For each of the indicator model outputs, we implemented a validation procedure and calculated a set of performance statistics. This involved using an out-of-sample cross-validation with a five-fold hold-out procedure and a comparison of the predicted values at the locations of the hold-out data with their observed values. This procedure was repeated five times without replacement so that every data point was omitted one time across the five validation runs. Standard validation statistics were then computed as measures of the predictive accuracy of the modeled estimates. This included mean absolute error (MAE), mean error (ME) or bias; root - mean - squared - error (RMSE, which summarizes the total variance); and 50%, 80%, and 95% coverage of our predictive intervals aggregated to the spatial holdout level. Each predictive metric was calculated by first simulating predictive draws using a binomial distribution. The predictive metric of

interest was then calculated as a sample-size-weighted mean over the second administrative levels (Mosser et al. 2019). To complement the out-of-sample predictive validity metrics, we also calculated in-sample predictive validity metrics that used the same process but matched each data point to predictions from a model fitted with all data.

4 RESULTS

4.1 Stacking Results

Here we present results obtained from the individual submodels according to the environmental and socioeconomic predictor variables in our model. The submodels revealed that the prediction of the DHS indicators varies spatially in the different areas in the country. Figure 5 and Figure 6 represent in-sample prediction areas with high and low prevalence of stunting for the Kenya DHS 2014 and Ethiopia DHS 2016 surveys that we generated from the three submodels.

Figure 5 Predicted surfaces for stunting in Kenya generated from the three submodels (GAM, GBM, and LASSO)

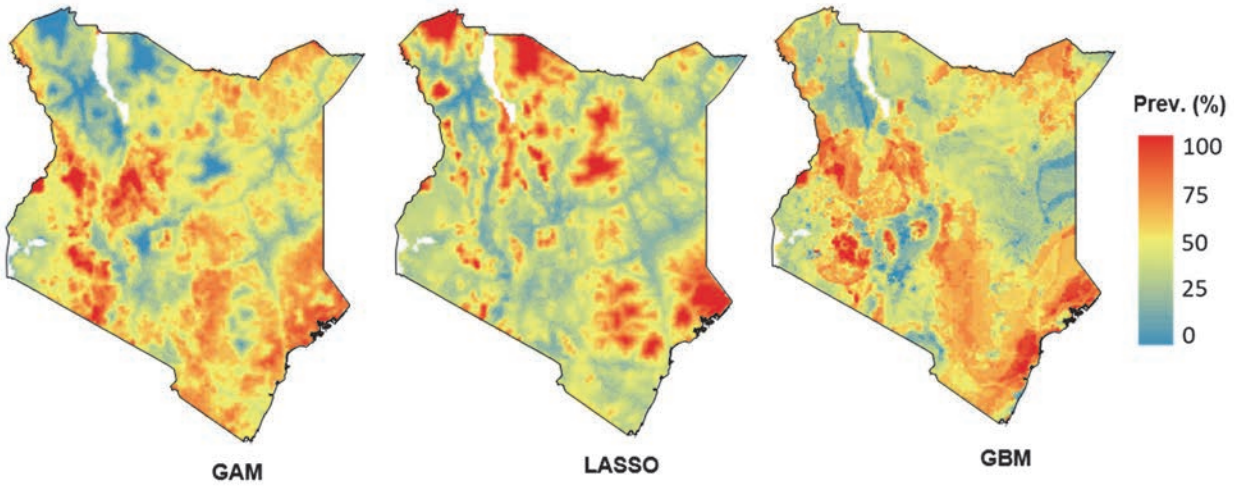
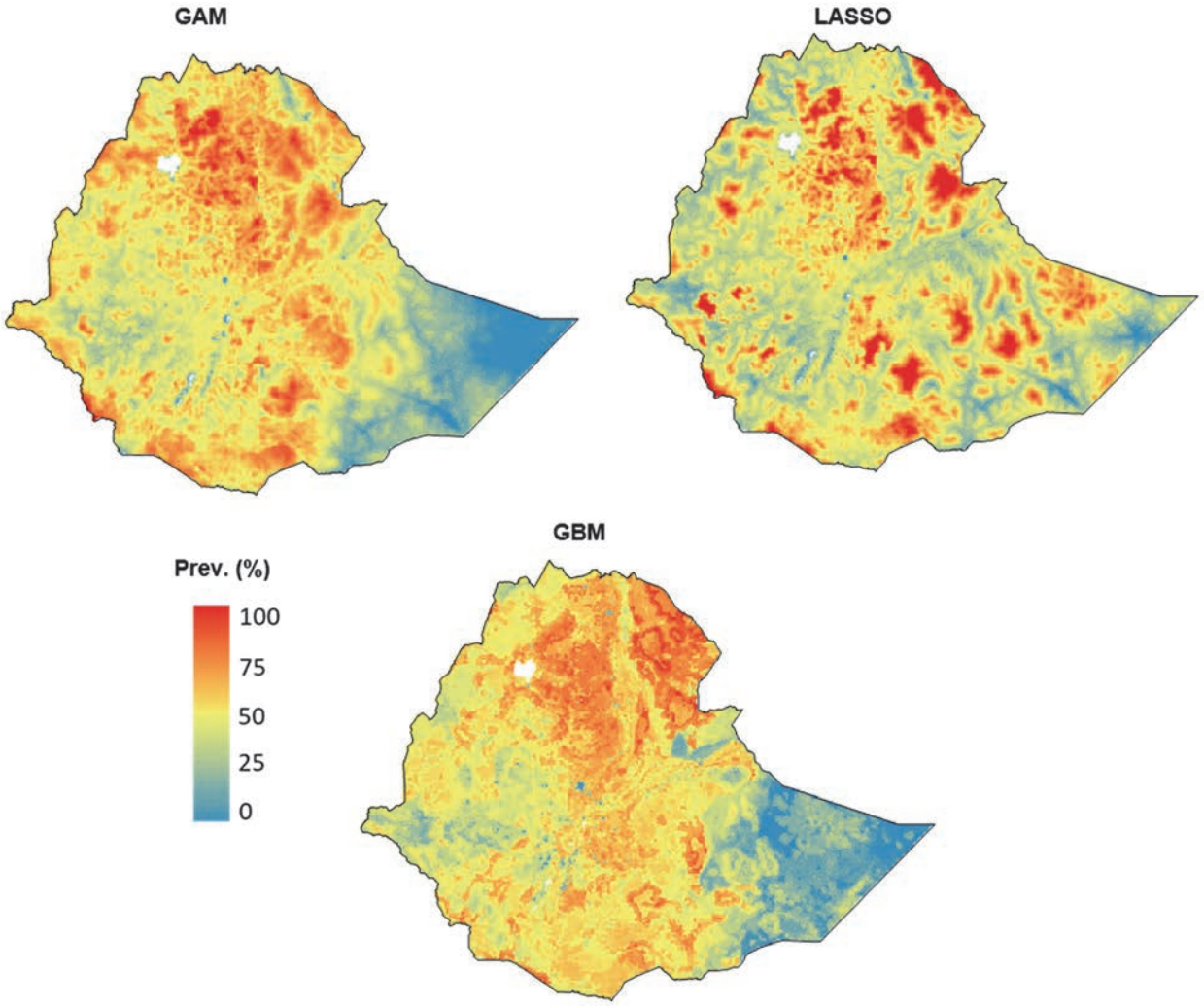


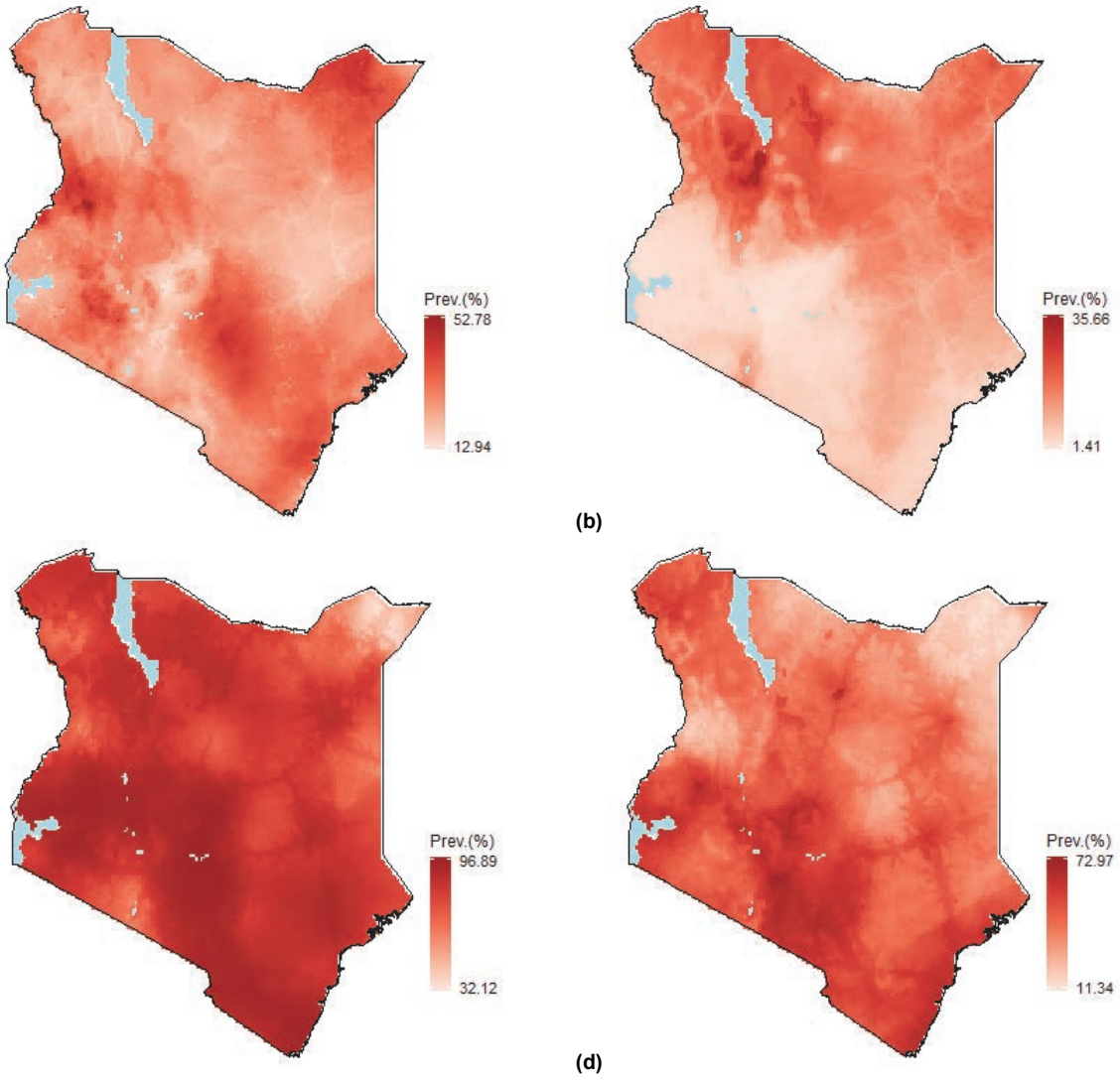
Figure 6 **Predicted surfaces for stunting in Ethiopia generated from the three submodels (GAM, GBM, and LASSO)**



4.2 Prediction Maps

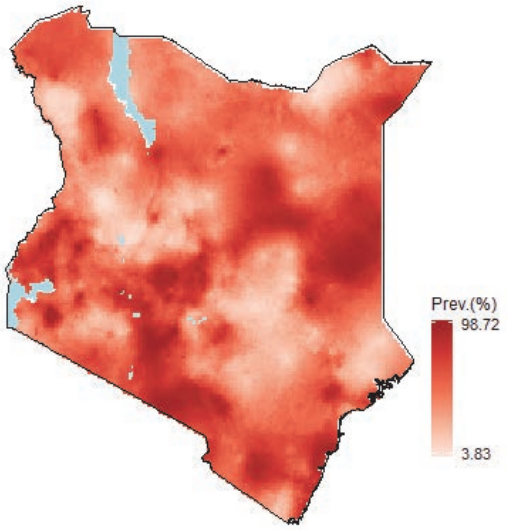
Prediction prevalence maps for each indicator were created using the full geospatial Bayesian model. Figure 7 and Figure 8 show the pixel level prediction surface maps for Kenya and Ethiopia, respectively. Areas with high and low estimated prevalence of each indicator can be seen clearly across all maps.

Figure 7 Pixel level prediction of prevalence of the indicators modeled by using the Kenya DHS 2014: (a) Stunting, (b) Wasting, (c) Vaccine DPT3, (d) ANC visits, and (e) Water sources



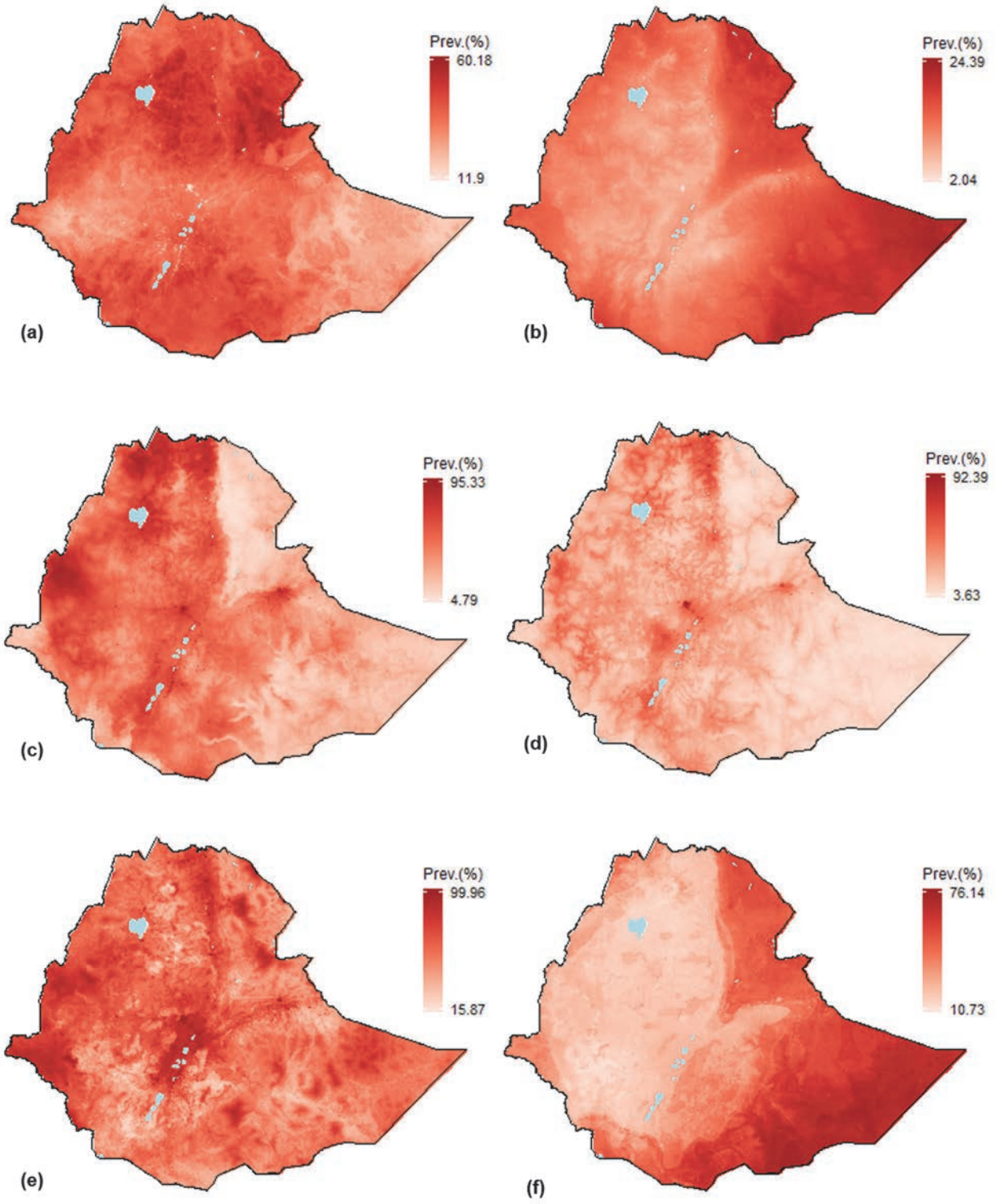
Continued…

Figure 7—*Continued*



(e)

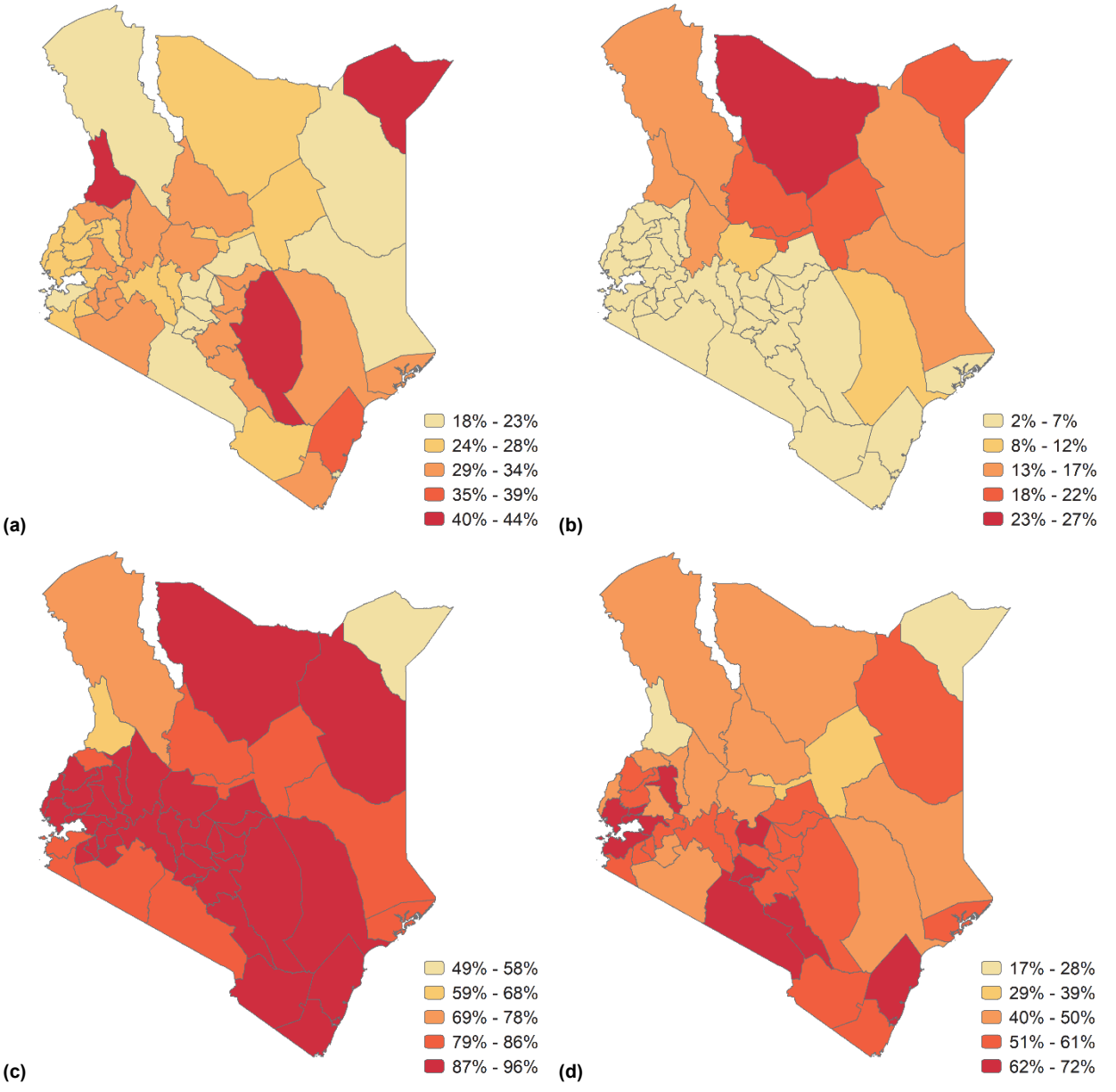
Figure 8 **Pixel level prediction of prevalence of the indicators modeled by using the Ethiopia DHS 2016: (a) Stunting, (b) Wasting, (c) Vaccine DPT3, (d) ANC visits, (e) Water sources, and (f) Women's anemia**



4.3 Admin Level 2 Estimates

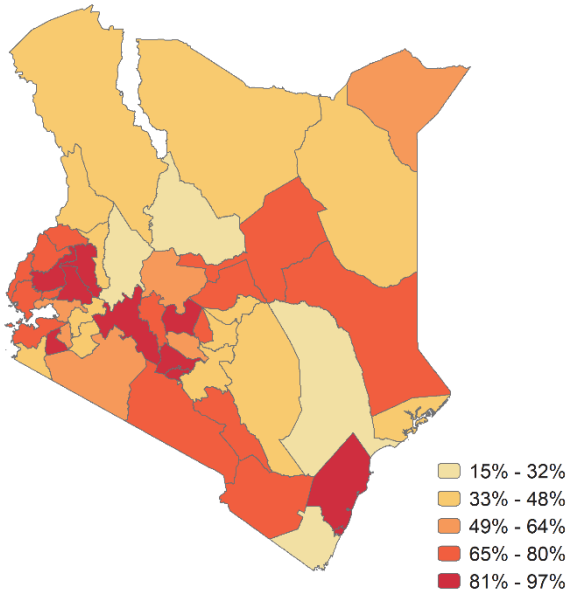
Figure 9 and Figure 10 show the second subnational administrative level estimates that highlight areas with high and low prevalence of each indicator we modeled. Additional results that show the mean prediction and the uncertainty intervals from the 2.5 and 97.5 percentiles are described in the Appendix Table A1 (Kenya) and Appendix Table A2 (Ethiopia).

Figure 9 **Second subnational administrative level estimates for Kenya DHS 2014: (a) Stunting, (b) Wasting, (c) Vaccine DPT3, (d) ANC visits, and (e) Water sources**



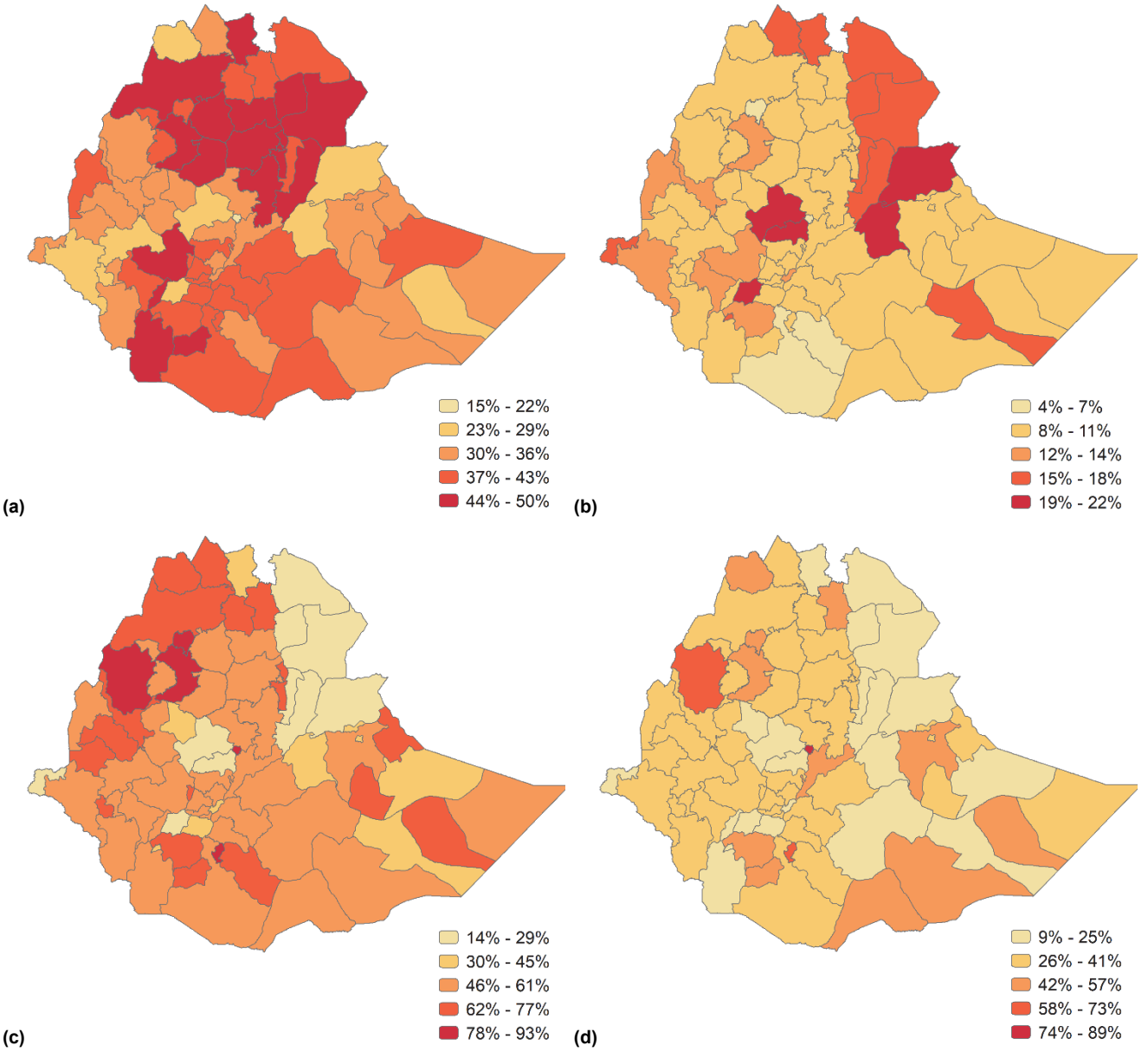
Continued…

Figure 9—*Continued*



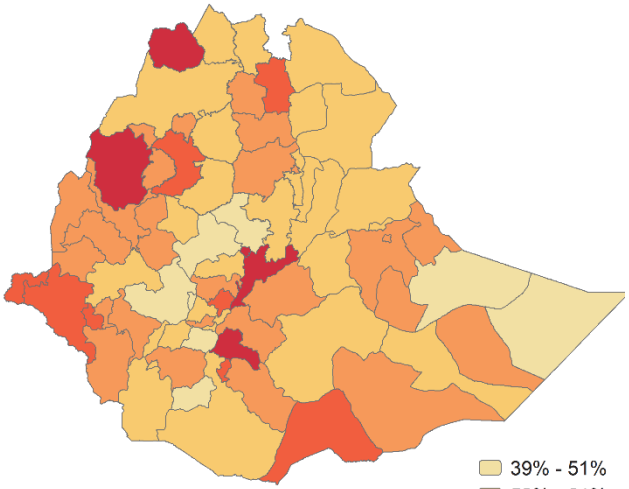
(e)

Figure 10 **Second subnational administrative level estimates for Ethiopia DHS 2016: (a) Stunting, (b) Wasting, (c) Vaccine DPT3, (d) ANC visits, (e) Water sources, and (f) Women's anemia**

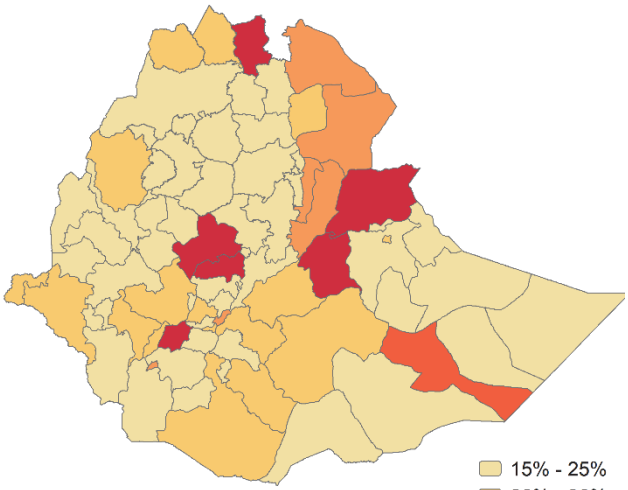


Continued…

Figure 10—*Continued*



(e)



(f)

4.4 Model Validation Metrics

Model validation was performed by calculating bias (mean error); mean absolute error (MAE); variance (RMSE); 50%, 80%, and 95% data coverage within prediction intervals; and the correlation between observed data and predictions.

Results from the validation indicated the best performance for each indicator, where correlation increased with decreased MAE and RMSE values. The coverage values for some indicators (vaccine) were too high, which was most likely a result of high uncertainties that arise from the small sample at the cluster locations (Table 3 and Table 4).

Table 3 Predictive metrics for each indicator aggregated at admin 2 (Kenya)

Indicator		ME	MAE	RMSE	50% Cov	80% Cov	95% Cov	Correlation
Stunting	In-sample	0.0019	0.0150	0.0183	0.7012	0.9272	0.9938	0.9742
	Out-of-sample	0.0025	0.0207	0.0254	0.6470	0.8819	0.9816	0.9501
Wasting	In-sample	-0.0003	0.0065	0.0083	0.8651	0.9680	0.9950	0.9849
	Out-of-sample	0.0002	0.0083	0.0104	0.8389	0.9460	0.9826	0.9769
DPT3	In-sample	0.0026	0.0217	0.0260	0.8930	0.9792	0.9975	0.9624
	Out-of-sample	-0.0016	0.0276	0.0341	0.8538	0.9461	0.9911	0.9288
4+ ANC Visits	In-sample	-0.0001	0.0149	0.0199	0.7129	0.9229	0.9863	0.9844
	Out-of-sample	-0.0004	0.0209	0.0288	0.6575	0.8838	0.9701	0.9643
Water Sources	In-sample	-0.0099	0.0321	0.0399	0.7566	0.9541	0.9947	0.9766
	Out-of-sample	-0.0148	0.0453	0.0549	0.5953	0.8731	0.9644	0.9582

Table 4 Predictive metrics for each indicator aggregated at admin 2 (Ethiopia)

Indicator		ME	MAE	RMSE	50% Cov	80% Cov	95% Cov	Correlation
Stunting	In-sample	0.0010	0.0257	0.0346	0.7608	0.9513	0.9939	0.9505
	Out-of-sample	0.0024	0.0403	0.0513	0.6262	0.8787	0.9777	0.8808
Wasting	In-sample	0.0018	0.0247	0.0327	0.7268	0.9221	0.9863	0.8010
	Out-of-sample	0.0009	0.0242	0.0325	0.7242	0.9221	0.9854	0.7981
DPT3	In-sample	-0.0015	0.0628	0.0898	0.9229	0.9963	1.0000	0.9545
	Out-of-sample	-0.0076	0.0706	0.1023	0.8935	0.9786	1.0000	0.9364
4+ ANC Visits	In-sample	-0.0054	0.0299	0.0441	0.8339	0.9705	0.9978	0.9823
	Out-of-sample	-0.0078	0.0493	0.0764	0.6513	0.9002	0.9884	0.9358
Water Sources	In-sample	-0.0217	0.0510	0.0734	0.7663	0.9607	0.9955	0.9505
	Out-of-sample	-0.0318	0.0767	0.1156	0.6518	0.8773	0.9724	0.8535
Women's Anemia	In-sample	0.0002	0.0271	0.0381	0.6792	0.9155	0.9944	0.9569
	Out-of-sample	0.0016	0.0385	0.0564	0.4965	0.7694	0.9095	0.9087

4.4.1 Comparison of model estimates versus DHS estimates

Figures 11 to 14 show the comparison estimates for each indicator produced by the models in our analysis and the equivalent estimates from the observed DHS survey data. The results indicate a high correlation between MBG and DHS estimates for most indicators.

Figure 11 Comparison of in-sample predictions for each indicator, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level for Kenya

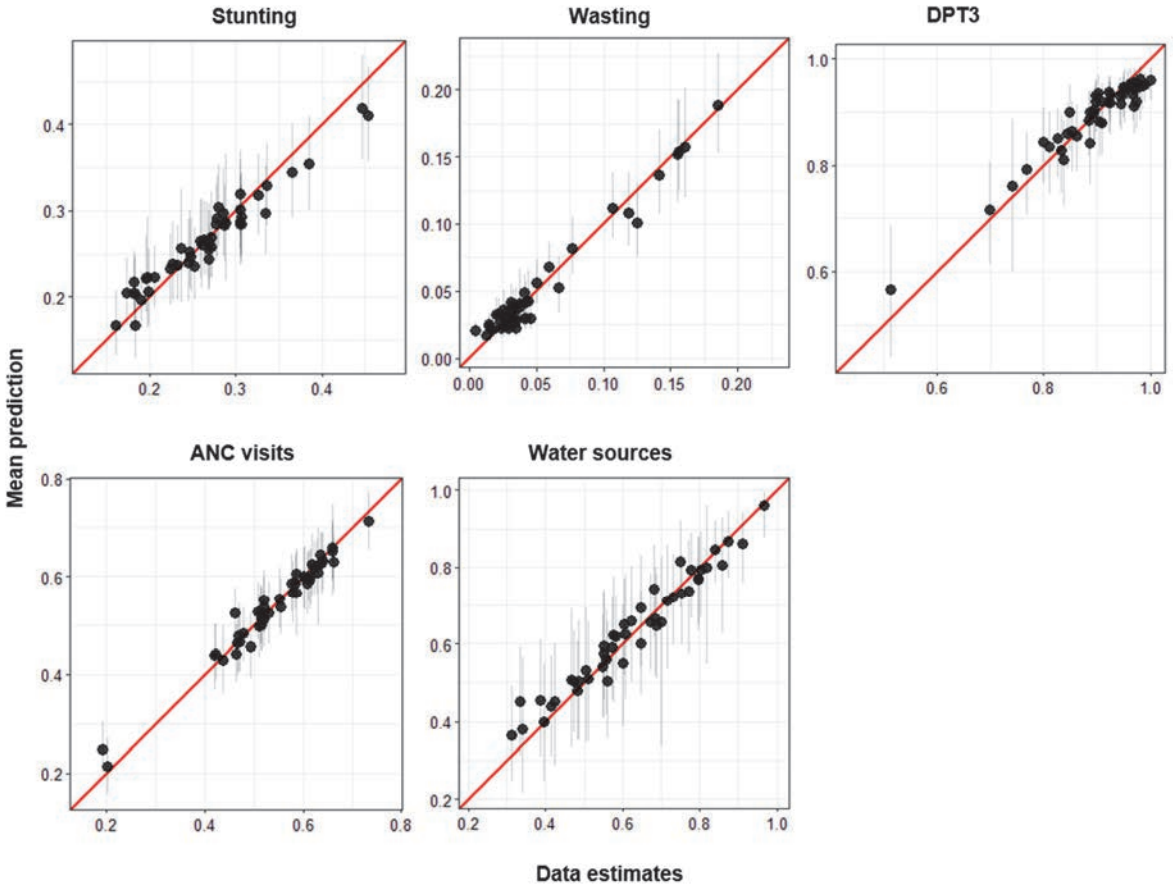


Figure 12 **Comparison of in-sample predictions for each indicator, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations from the same area aggregated to the second subnational administrative level for Ethiopia**

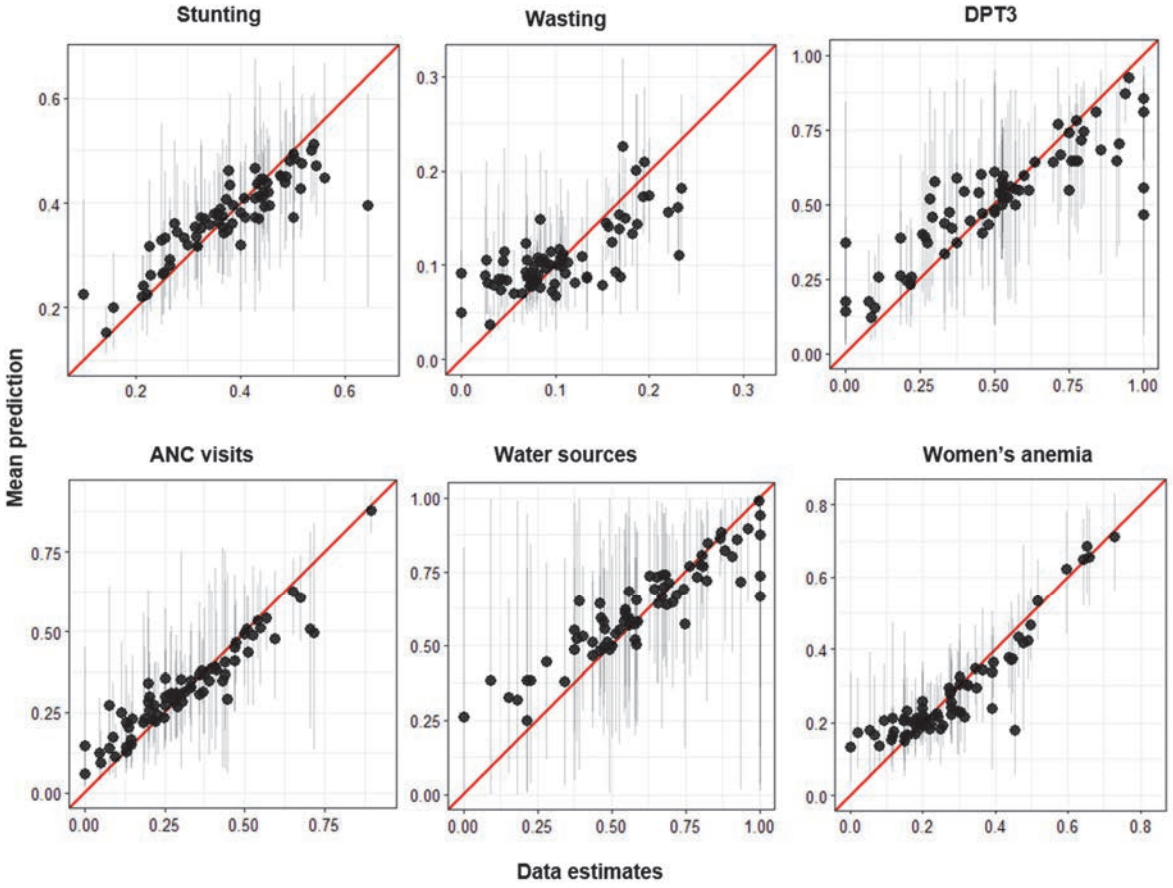


Figure 13 **Comparison of out-of-sample predictions for each indicator, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations for Kenya**

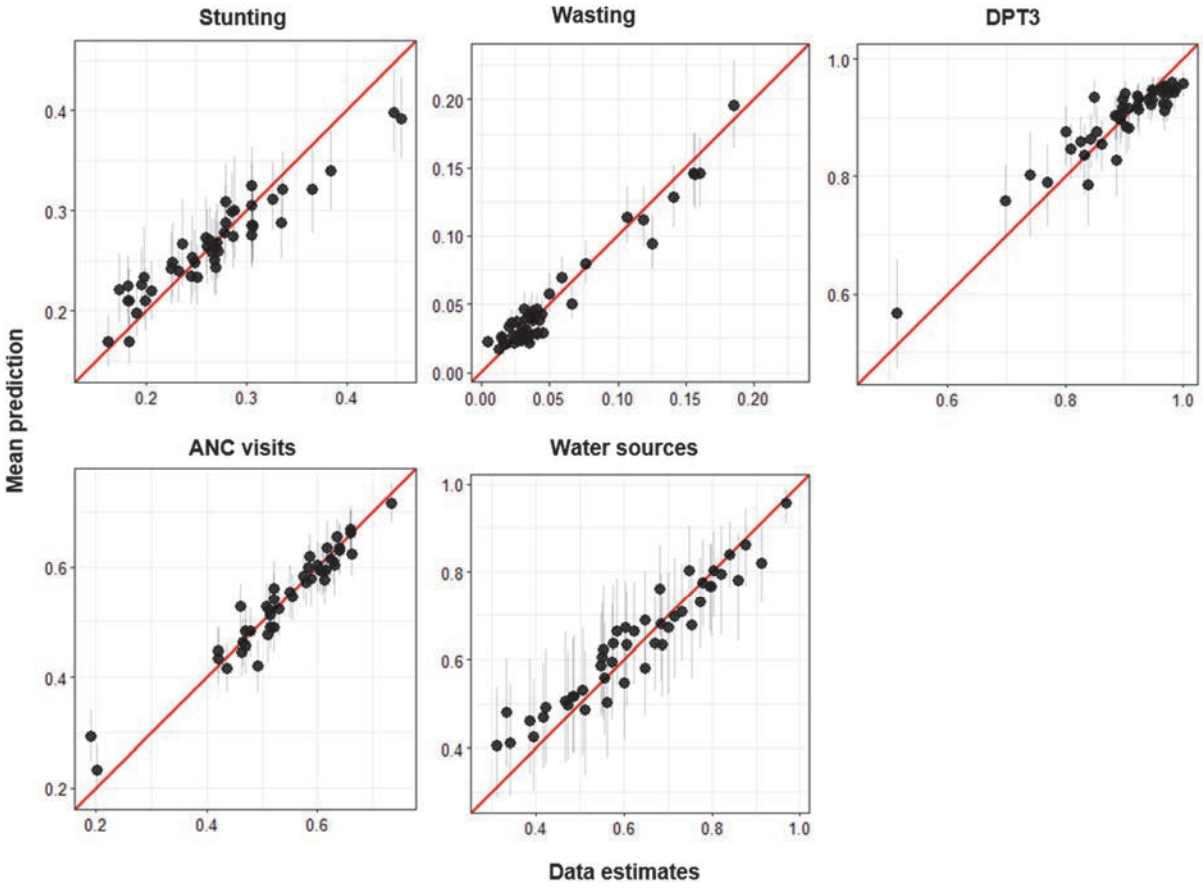
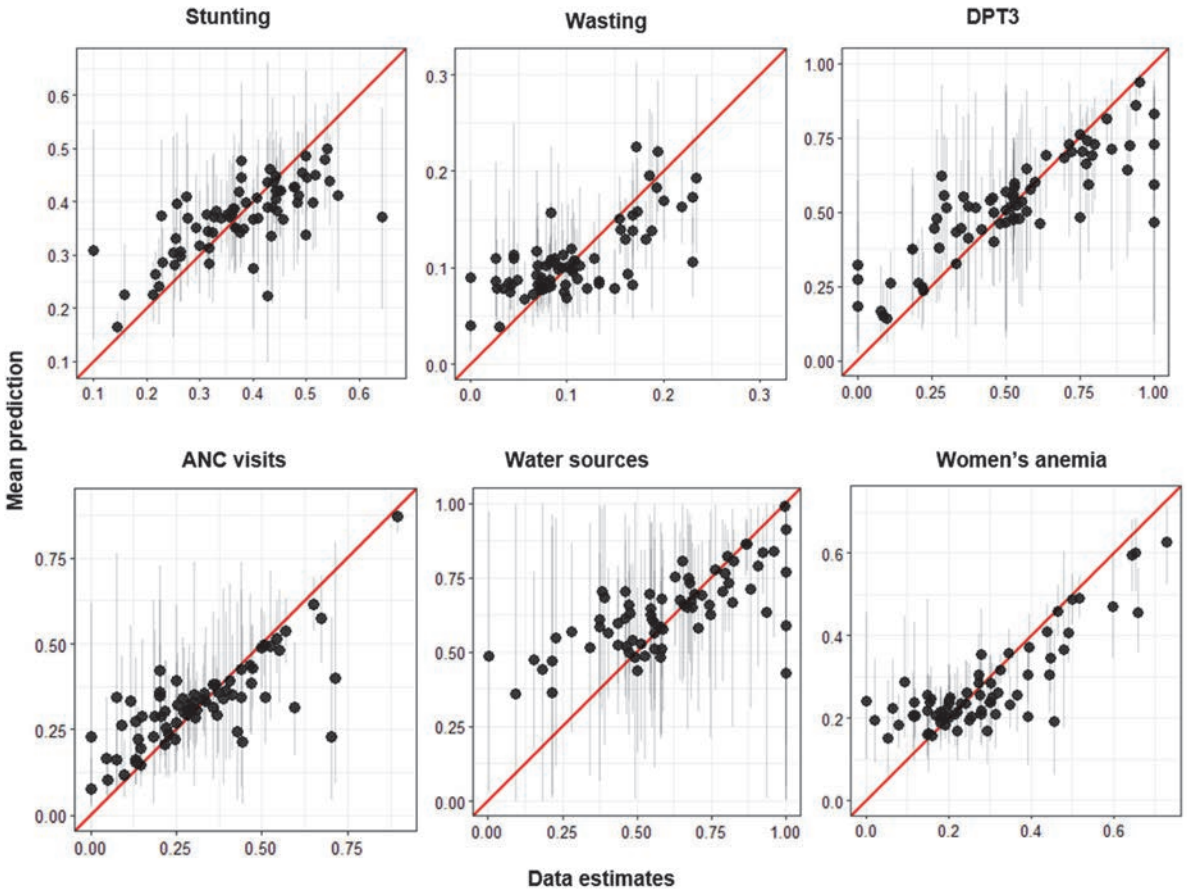


Figure 14 **Comparison of out-of-sample predictions for each indicator, aggregated to the second subnational administrative level with 95% uncertainty intervals, plotted against data observations for Ethiopia**



5 DISCUSSION AND CONCLUSION

In recent years, there has been a need for district (ADMIN 2) estimates currently not available in a DHS survey. In an increasingly resource-constrained environment, high resolution maps of key health indicators and development derived from cluster point data through spatial interpolation methods offer an attractive solution.

In this analysis, we developed a methodological framework for estimating DHS indicators at the ADMIN 2 level. We took advantage of the advancement in geospatial technologies, availability of free and open source spatial data, and geospatial software tools relevant for spatial modeling. This framework used an ensemble modeling approach that combines multiple model algorithmic methods to increase predictive validity relative to a single modeling approach. We employed this approach to capture potential complex interactions and non-linear effects among the geospatial covariates. We fitted three submodels (GAM, LASSO, and GBM) to each of the selected DHS survey indicator data using the geospatial covariates as exploratory predictors. The submodels were selected because they were available in standalone packages that required minimal data preparation after the predictor variables had been produced; they are particularly useful in cases where presence-only data are available (Peterson, Papes, and Eaton 2007; Phillips and Dudik 2008; Elith et al. 2006) and have been successfully used in previous analyses that have implemented the stacking approach to derive continuous estimated surfaces of indicators of interest from the DHS household surveys. These include mapping of malaria (Bhatt et al. 2017), HIV prevalence (Dwyer-Lindgren et al. 2019), vaccine coverage (Mosser et al. 2019), exclusive breastfeeding (Bhattacharjee et al. 2019), child growth failure (Osgood-Zimmerman et al. 2018), education attainment (Graetz et al. 2018), and childhood diarrheal diseases (Reiner et al. 2018).

We found variability in the individual model prediction outputs. For example, the LASSO model output indicated a low prediction of stunting in the northeast areas of Kenya (Figure 5) as compared to the other algorithms. This could be explained by the prediction uncertainties in the individual submodels (Elith et al. 2006; Peterson, Papes, and Eaton 2007). Using different models and combining them in an ensemble model could improve these uncertainties (Arauj and New 2007; Marmion et al. 2009; Jones-Farrand et al. 2011). Our findings demonstrated that the predictions from the stacking ensemble model approach were more accurate than those from the individual model algorithms. The results suggest that use of an ensemble model approach is more adequate than predictions from any single modeling methods. These findings are consistent with other studies (Marmion et al. 2009; Bhatt et al. 2017), which showed the ensemble model approach to be the best. By developing the methodological framework within the R statistical computing environment, we have created a tool that can be used to model other health indicators.

The results from the ADMIN 2 level, generated from the full geostatistical model, show that the estimated prevalence of each indicator varied across the different areas in the country. We found that the proportion of ADMIN 2 with uncertain or very uncertain estimates is high. The reason could be that DHS survey data are designed to be representative at the national and first administrative levels, but not at the ADMIN 2 level. The accuracy of the estimates could be improved by computing them at a lower resolution (if data are available) or by increasing the sample size of the DHS survey. Both approaches increase the number of observations per administrative unit. However, increasing the sampling size of a survey is costly, especially in resource–constrained countries (Utazi et al. 2018; Gething and Burgert-Brucker 2017).

Although we have estimated the prevalence for each indicator at the pixel-level and the ADMIN 2 level, our study has some limitations. Our analysis used a suite of standard geospatial covariates that included those that are not directly related to the indicators we modeled. To improve predictions, further studies should restrict the model input data to those covariates associated with the indicators of interest. Due to computational limitations, we did not quantify uncertainty in the covariates and submodel estimates. Thus further analysis should develop methods that are capable of propagating uncertainty in both the covariates and submodel estimates (Dwyer-Lindgren et al. 2019; Wakefield et al. 2018).

We generated maps showing estimates of high risk areas for each indicator we modeled. Our approach in this analysis can help inform the allocation of resources and program implementation in areas that need more attention. Interventions and programs that can be implemented and directed at much smaller spatial scales using the MBG estimates such as the one described in our analysis could enable better programmatic decisions.

REFERENCES

Alegana, V. A., P. M. Atkinson, C. Pezzulo, A. Sorichetta, D. Weiss, T. Bird, E. Erbach-Schoenberg, and A. J. Tatem. 2015. "Fine Resolution Mapping of Population Age-structures for Health and Development Applications." *Journal of the Royal Society Interface* 12 (105): 20150073. https://doi.org/10.1098/rsif.2015.0073.

Araujo, M. B. and M. New. 2007. "Ensemble Forecasting of Species Distributions." *Trends in Ecology and Evolution* 22 (1): 42-47. https://doi.org/10.1016/j.tree.2006.09.010.

Banerjee, S., B. P. Carlin, and A. E. Gelfand. 2014. *Hierarchical Modeling and Analysis for Spatial Data*. Second Edition. Boca Raton, Florida, USA: Chapman and Hall/CRC.

Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. "Selecting Pseudo-absences for Species Distribution Models: How, Where and How Many?" *Methods in Ecology and Evolution* 3 (2): 327-338. https://doi.org/10.1111/j.2041-210X.2011.00172.x.

Bhatt, S., E. Cameron, S. R. Flaxman, D. J. Weiss, D. L. Smith, and P. W. Gething. 2017. "Improved Prediction Accuracy for Disease Risk Mapping Using Gaussian Process Stacked Generalization." *Journal of the Royal Society Interface* 14 (134): p ii: 20170520. https://doi.org/10.1098/rsif.2017.0520.

Bhattacharjee, N. V., L. E. Schaeffer, L. B. Marczak, J. M. Ross, S. J. Swartz, J. Albright, W. M. Gardner, et al. 2019. "Mapping Exclusive Breastfeeding in Africa between 2000 and 2017." *Nature Medicine* 25: 1205-1212. https://doi.org/10.1038/s41591-019-0525-0.

Blangiardo, M., and M. Cameletti. 2015. *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Chichester, United Kingdom: Wiley.

Burgert, C. R. 2014. *Spatial Interpolation with Demographic and Health Survey Data: Key Considerations.* DHS Spatial Analysis Reports. Rockville, Maryland, USA: *ICF International*. http://dhsprogram.com/pubs/pdf/SAR9/SAR9.

Cameletti, M., F. Lindgren, D. Simpson, and H. Rue. 2013. "Spatio-temporal Modeling of Particulate Matter Concentration through the SPDE Approach." *Advances in Statistical Analysis* 97(2): 109-131. http://dx.doi.org/10.1007/s10182-012-0196-3.

Cressie, N. and C. K. Wikle. 2011. *Statistics for Spatio-Temporal Data*. Indianapolis, Indiana, USA: John Wiley & Son, Inc.

Diggle, P. J., and E. Giorgi. 2019. *Model-based Geostatistics for Global Public Health: Methods and Applications.* New York, USA: Chapman and Hall/CRC.

Dwyer-Lindgren, L., M. A. Cork, A. Sligar, K. M. Steuben, K. F. Wilson, N. R. Provost, B. K. Mayala, et al. 2019. "Mapping HIV Prevalence in Sub-Saharan Africa between 2000 and 2017." *Nature* 570 (7760): 189-193. https://doi.org/10.1038/s41586-019-1200-9.

Elith, J., and C. H. Graham. 2009. “Do They? How Do They? WHY Do They Differ? On Finding Reasons for Differing Performances of Species Distribution Models.” *Ecography* 32 (1): 66-77. https://doi.org/10.1111/j.1600-0587.2008.05505.x.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, et al. 2006. “Novel Methods Improve Prediction of Species’ Distributions from Occurrence Data.” *Ecography* 29 (2): 129-151. https://doi.org/10.1111/j.2006.0906-7590.04596.x.

Franklin, J. 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge, United Kingdom: Cambridge University Press.

Friedman, J. H. 2001. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics* 29 (5):1189-1232. https://doi.org/10.1214/aos/1013203451.

Gething, P. W., A. P. Patil, D. L. Smith, C. A. Guerra, I. R. F. Elyazar, G. L. Johnston, A. J. Tatem, and S. Hay. 2011. “A New World Malaria Map: *Plasmodium Falciparum* Endemicity in 2010.” *Malaria Journal* 10 (378). https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-10-378.

Gething, P., A. Tatem, T. Bird, and C. Burgert-Brucker. 2015. *Creating Spatial Interpolation Surfaces with DHS Data.* ICF International Spatial Analysis Reports No. 11. Rockville, Maryland, USA: ICF International. https://dhsprogram.com/pubs/pdf/SAR11/SAR11.pdf.

Gething, P. W., and C. Burgert-Brucker. 2017. *The DHS Program Modeled Map Surfaces: Understanding the Utility of Spatial Interpolation for Generating Indicators at Subnational Administrative Levels*. DHS Spatial Analysis Reports No. 15. Rockville, Maryland, USA: ICF International. http://dhsprogram.com/pubs/pdf/SAR15/SAR15.pdf.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1995. *Markov Chain Monte Carlo in Practice*. Boca Raton, Florida, USA: London: Chapman & Hall/CRC.

Giovanelli, J. G. R., M. F. D. Siqueira, C. F. B Haddad, and J. Alexandrino. 2010. “Modeling a Spatially Restricted Distribution in the Neotropics: How the Size of Calibration Area Affects the Performance of Five Presence-only Methods.” *Ecological Modelling* 221 (2): 215-224. https://econpapers.repec.org/RePEc:eee:ecomod:v:221:y:2010:i:2:p:215-224.

Gosoniu, L., A. Msengwa, C. Lengeler, and P. Vounatsou. 2012. “Spatially Explicit Burden Estimates of Malaria in Tanzania: Bayesian Geostatistical Modeling of the Malaria Indicator Survey Data.” *PLoS One* 7(5): e23966. https://doi.org/10.1371/journal.pone.0023966.

Gosoniu, L., A. M. Veta, and P. Vounatsou, 2011. “Bayesian Geostatistical Modeling of Malaria Indicator Survey Data in Angola.” *PLoS One* 5: e9322. https://dx.plos.org/10.1371/journal.pone.0009322.

Gosoniu, L., P. Vounatsou, N. Sogoba, and T. Smith. 2006. “Bayesian Modelling of Geostatistical Malaria Risk Data.” *Geospatial Health* 1 (1): 127-139. https://doi.org/10.4081/gh.2006.287.

Graetz, N., J. Friedman, A. Osgood-Zimmerman, R. Burstein, M. H. Biehl, C. Shields, J. F. Mosser, et al. 2018. “Mapping Local Variation in Educational Attainment across Africa.” *Nature* 555: 48-53. http://dx.doi.org/10.1038/nature25761.

Hay, S. I., C. A. Guerra, P. W. Gething, A. P. Patil, A. J. Tatem, A. M. Noor, , C. W. Kabaria, et al. “A World Malaria Map: *Plasmodium Falciparum* Endemicity in 2007.” *PLoS Med* 6: e1000048. https://doi.org/10.1371/journal.pmed.1000048.

Jones-Farrand, D. T., T. M. Fearer, W. E. Thoghmartin, F. R Thompson, M. D. Nelson, and J. M Tirpak. 2011. “Comparison of Statistical and Theoretical Habitat Models for Conservation Planning: The Benefit of Ensemble Prediction.” *Ecological Applications* 21(6): 2269-2282. https://doi.org/10.1890/10-1047.1.

Kazembe, L., I. Kleinschmidt, T. Holtz, and B. Sharp. 2006. “Spatial Analysis and Mapping of Malaria Risk in Malawi Using Point-referenced Prevalence of Infection Data.” *International Journal of Health Geographics* 5 (1): 41. https://doi.org/10.1186/1476-072X-5-41.

Li, Z., Y. Hsiao, J. Godwin, B. D. Martin, J. Wakefield, and S. J. Clark with support from the United Nations Inter-agency Group for Child Mortality Estimation and its technical advisory group. 2019. “Changes in the Spatial Distribution of the Under-five Mortality Rate: Small-area Analysis of 122 DHS Surveys in 262 Subregions of 35 Countries in Africa.” *PloS one Public Library of Science* 14 (1): e0210645. https://doi.org/10.1371/journal.pone.0210645.

Lindgren, F., H. Rue, and J. Lindstrom. 2011. “An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach.” *Journal of Royal Statistical Society Series B* 73 (4): 423-498. https://doi.org/10.1111/j.1467-9868.2011.00777.x.

Lobo, J. M., A. Jiménez-Valverde, and J. Hortal. 2010. “The Uncertain Nature of Absences and Their Importance in Species Distribution Modelling.” *Ecography* 33 (1): 103-114. https://doi.org/10.1111/j.1600-0587.2009.06039.x.

Marmion, M., M. Parviainen, M. Luoto, R. K Heikkinen, and W. Thuiller. 2009. “Evaluation of Consensus Methods in Predictive Species Distribution Modelling.” *Diversity and Distributions* 15 (1): 59-69. https://doi.org/10.1111/j.1472-4642.2008.00491.x.

Mateo, R. G., T. B. Croat, A. M. Felicısimo, and J. Munoz. 2010. “Profile or Group Discriminative Techniques? Generating Reliable Species Distribution Models Using Pseudo-absences and Target-group Absences from Natural History Collections.” *Diversity and Distributions* 16: 84-94. https://doi.org/10.1111/j.1472-4642.2009.00617.x.

Mayala, B. K., T. D. Fish, D. Eitelberg, and T. Dontamsetti. 2018. *The DHS Program Geospatial Covariate Datasets Manual*. 2nd Edition. Rockville, Maryland, USA: ICF International.

Mosser, J. F., W. Gagne-Maynard, P. C. Rao, A. Osgood-Zimmerman, N. Fullman, N. Graetz, R. Burstein, et al. 2019. “Mapping Diphtheria-pertussis-tetanus Vaccine Coverage in Africa, 2000–2016: A Spatial and Temporal Modelling Study.” *The Lancet* 393 (10183): 1843-1855. https://doi.org/10.1016/S0140-6736(19)30226-0.

Noor, A. M., P. W. Gething, V. A. Alegana, A. P. Patil, S. I. Hay, E. Muchiri, E. Juma, and R. W. Snow. 2009. "The Risks of Malaria Infection in Kenya in 2009." *BMC Infectious Diseases* 9 (180). https://doi.org/10.1186/1471-2334-9-180.

Osgood-Zimmerman, A., A. I. Millear, R. W. Stubbs, C. Shields, B. V. Pickering, L. Earl, N. Graetz, et al. 2018. "Mapping Child Growth Failure in Africa between 2000 and 2015." *Nature* 555 (7684): 41-47. http://dx.doi.org/10.1038/nature25760.

Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson. 2007. "Predicting Species' Distributions from Small Numbers of Occurrence Records: A Test Case Using Cryptic Geckos in Madagascar." *Journal of Biogeography* 34: 102-117. https://doi.org/10.1111/j.1365-2699.2006.01594.x

Peterson, A. T., J. Soberon, R. G. Pearson, R. P. Anderson, E. Martinez-Meyer, M. Nakamura, and M. B. Araujo. 2011. *Ecological Niches and Geographic Distributions*, Princeton, New Jersey, USA: Princeton University Press.

Peterson, A. T., M. Papes, and M. Eaton. 2007. "Transferability and Model Evaluation in Ecological Niche Modeling: A Comparison of Garp and Maxent." *Ecography* 30 (4): 550-560. https://doi.org/10.1111/j.0906-7590.2007.05102.x.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum Entropy Modeling of Species Geographic Distributions." *Ecological Modelling* 190 (3-4): 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Phillips, S. J., and M. Dudik. 2008. "Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation." *Ecography* 31: 161-175.

R Core Team 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raso, G., N. Schur, J. Utzinger, B. Koudou, E. Tchicaya, F. Rohner, and E. N'Goran. 2012. "Mapping Malaria Risk among Children in Cote d'Ivoire using Bayesian Geo-statistical Models." *Malaria Journal* 11(1): 160. https://doi.org/10.1186/1475-2875-11-160.

Reiner, R. C., N. Graetz, D. C. Casey, C. Troeger, G. M. Garcia, J. F. Mosser, A. Deshpande, et al. 2018. "Variation in Childhood Diarrheal Morbidity and Mortality in Africa, 2000-2015." *New England Journal of Medicine* 379 (12): 1128-1138. https://www.nejm.org/doi/10.1056/NEJMoa1716766.

Riedel, N., P. Vounatsou, J. M. Miller, L. Gosoniu, E. Chizema-Kawesha, V. Mukonka, and R. W. Steketee. 2011. "Geographical Patterns and Predictors of Malaria Risk in Zambia: Bayesian Geostatistical Modelling of the 2006 Zambia National Malaria Indicator Survey (ZMIS)." *Malaria Journal* 9: 37. https://doi.org/10.1186/1475-2875-9-37.

Rue, H., S. Martino, and N. Chopin. 2009. "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations." *Journal of the Royal Statistical Society: Series B* 71 (2): 319. https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Steele, J. E., P. R Sundsøy, P. Carla, V. A. Alegana, T. J. Bird, B. Joshua, B. Johannes, et al. 2017. "Mapping Poverty using Mobile Phone and Satellite Data." *Journal of The Royal Society Interface Royal Society* 14 (127): 20160690. https://doi.org/10.1098/rsif.2016.0690.

United Nations General Assembly 2015. *Transforming our World: The 2030 Agenda for Sustainable Development*. New York, New York, USA: United Nations Population Fund. https://www.unfpa.org/resources/transforming-our-world-2030-agenda-sustainable-development.

Utazi, C. E., J. Thorley, V. A. Alegana, J. M. Ferrari, S. Takahashi, C. J. E. Metcalf, J. Lessler, and A. Tatem. 2018. "High Resolution Age-structured Mapping of Childhood Vaccination Coverage in Low and Middle Income Countries." *Vaccine* 36 (12): 1583-1591. https://doi.org/10.1016/j.vaccine.2018.02.020.

Wakefield, J., G. A. Fuglstad, A. Riebler, J. Godwin, K. Wilson, and S. Clark. 2018. "Estimating Under Five Mortality in Space and Time in a Developing World Context." *Statistical Methods in Medical Research*. https://doi.org/10.1177/0962280218767988.

Wisz, M., and A. Guisan. 2009. "Do Pseudo-Absence Selection Strategies Influence Species Distribution Models and Their Predictions? An Information-Theoretic Approach Based on Simulated Data. *BMC Ecology* 9 (1): 8. https://doi.org/10.1186/1472-6785-9-8.

Wood, S. N. 2017. *Generalized Additive Models: An Introduction with R.* 2nd Edition. New York, New York, USA: CRC Press.

Zou, T., and T. Hastie. 2005. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 67 (2): 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

APPENDICES

Appendix A

Appendix Table A1 Summarized admin 2 mean and 95% uncertainty interval for Kenya

ADM2_NAME	Mean stunting	Upper	Lower	Mean wasting	Upper	Lower	Mean DPT3	Upper	Lower	Mean ANC visits	Upper	Lower	Mean water sources	Upper	Lower
Kiambu	0.199	0.342	0.099	0.021	0.046	0.008	0.925	0.986	0.790	0.630	0.804	0.410	0.816	1.000	0.131
Kirinyaga	0.189	0.327	0.095	0.021	0.043	0.008	0.961	0.994	0.871	0.600	0.779	0.395	0.707	0.998	0.043
Muranga	0.194	0.338	0.098	0.022	0.047	0.009	0.934	0.988	0.801	0.593	0.771	0.393	0.613	0.997	0.019
Nyandarua	0.241	0.430	0.119	0.020	0.043	0.008	0.926	0.987	0.767	0.604	0.793	0.389	0.770	0.999	0.048
Nyeri	0.177	0.315	0.078	0.016	0.034	0.007	0.935	0.988	0.811	0.657	0.817	0.454	0.861	1.000	0.150
Kilifi	0.376	0.588	0.186	0.043	0.087	0.017	0.866	0.978	0.602	0.626	0.811	0.384	0.823	1.000	0.065
Kwale	0.327	0.529	0.169	0.042	0.086	0.017	0.947	0.992	0.828	0.579	0.779	0.365	0.293	0.979	0.001
Lamu	0.295	0.488	0.135	0.046	0.094	0.019	0.821	0.964	0.554	0.572	0.781	0.350	0.357	0.988	0.001
Mombasa	0.223	0.376	0.110	0.045	0.092	0.019	0.946	0.990	0.838	0.672	0.826	0.474	0.825	1.000	0.113
Taita Taveta	0.267	0.449	0.137	0.050	0.103	0.019	0.952	0.995	0.817	0.557	0.740	0.346	0.707	0.999	0.031
Tana River	0.323	0.528	0.156	0.092	0.179	0.038	0.879	0.987	0.609	0.428	0.653	0.216	0.282	0.986	0.000
Embu	0.317	0.518	0.170	0.033	0.068	0.013	0.960	0.994	0.872	0.537	0.736	0.326	0.377	0.982	0.003
Isiolo	0.246	0.420	0.118	0.180	0.332	0.080	0.794	0.964	0.430	0.326	0.535	0.157	0.797	1.000	0.057
Kitui	0.414	0.598	0.228	0.045	0.096	0.018	0.881	0.984	0.650	0.582	0.785	0.370	0.345	0.989	0.000
Machakos	0.317	0.512	0.168	0.027	0.056	0.010	0.959	0.994	0.866	0.583	0.781	0.382	0.475	0.991	0.005
Makueni	0.284	0.460	0.140	0.039	0.082	0.015	0.934	0.991	0.766	0.658	0.841	0.450	0.686	0.999	0.018
Marsabit	0.231	0.421	0.103	0.272	0.485	0.125	0.883	0.988	0.600	0.436	0.669	0.224	0.438	0.996	0.002
Meru	0.197	0.347	0.097	0.035	0.075	0.014	0.960	0.994	0.872	0.559	0.765	0.349	0.655	0.998	0.015
Tharaka-Nithi	0.307	0.501	0.160	0.033	0.070	0.013	0.956	0.993	0.850	0.535	0.730	0.335	0.361	0.981	0.002
Nairobi	0.180	0.319	0.082	0.024	0.051	0.010	0.919	0.983	0.785	0.720	0.862	0.536	0.965	1.000	0.658
Garissa	0.197	0.360	0.081	0.121	0.225	0.052	0.840	0.976	0.519	0.485	0.711	0.257	0.650	1.000	0.005
Mandera	0.430	0.633	0.234	0.192	0.334	0.083	0.492	0.849	0.149	0.167	0.357	0.070	0.563	0.998	0.004
Wajir	0.213	0.362	0.102	0.141	0.271	0.061	0.882	0.977	0.628	0.517	0.721	0.308	0.422	0.991	0.003
Kisii	0.232	0.392	0.121	0.025	0.053	0.010	0.921	0.985	0.777	0.521	0.715	0.327	0.929	1.000	0.369
Homa Bay	0.195	0.335	0.089	0.026	0.053	0.011	0.843	0.966	0.609	0.612	0.787	0.417	0.701	0.999	0.034
Kisumu	0.232	0.389	0.110	0.025	0.052	0.010	0.921	0.986	0.766	0.615	0.789	0.409	0.545	0.997	0.008
Migori	0.280	0.463	0.147	0.026	0.053	0.011	0.843	0.962	0.614	0.559	0.742	0.350	0.376	0.985	0.002
Nyamira	0.285	0.454	0.156	0.025	0.051	0.010	0.951	0.991	0.848	0.504	0.697	0.296	0.542	0.996	0.011
Siaya	0.244	0.414	0.118	0.018	0.039	0.007	0.939	0.990	0.812	0.624	0.797	0.425	0.684	0.998	0.030
Baringo	0.298	0.490	0.145	0.131	0.249	0.057	0.906	0.985	0.706	0.428	0.639	0.221	0.152	0.869	0.000
Bomet	0.312	0.494	0.168	0.027	0.055	0.011	0.948	0.991	0.829	0.428	0.631	0.244	0.455	0.989	0.004
Kajiado	0.229	0.402	0.105	0.044	0.084	0.018	0.848	0.976	0.561	0.649	0.829	0.427	0.794	1.000	0.064
Kericho	0.284	0.459	0.145	0.034	0.072	0.014	0.942	0.991	0.828	0.507	0.709	0.311	0.461	0.995	0.004
Laikipia	0.290	0.468	0.152	0.090	0.175	0.036	0.924	0.987	0.748	0.483	0.686	0.278	0.483	0.998	0.004
Elgeyo-Marakwet	0.337	0.538	0.180	0.058	0.118	0.025	0.919	0.985	0.762	0.442	0.646	0.253	0.359	0.983	0.002
Nakuru	0.251	0.424	0.129	0.026	0.054	0.010	0.890	0.982	0.667	0.589	0.784	0.379	0.802	1.000	0.070

Continued…

Appendix Table A1—*Continued*

ADM2_NAME	Mean stunting	Upper	Lower	Mean wasting	Upper	Lower	Mean DPT3	Upper	Lower	Mean ANC visits	Upper	Lower	Mean water sources	Upper	Lower
Narok	0.305	0.495	0.152	0.055	0.115	0.022	0.811	0.964	0.483	0.449	0.668	0.235	0.525	0.997	0.004
Samburu	0.308	0.496	0.155	0.211	0.373	0.092	0.781	0.950	0.464	0.451	0.675	0.245	0.210	0.963	0.000
Trans Nzoia	0.314	0.487	0.174	0.038	0.083	0.016	0.819	0.956	0.564	0.438	0.641	0.242	0.796	1.000	0.096
Turkana	0.204	0.375	0.093	0.166	0.312	0.073	0.766	0.951	0.428	0.445	0.671	0.238	0.346	0.987	0.001
Uasin Gishu	0.255	0.414	0.128	0.031	0.064	0.013	0.950	0.991	0.847	0.676	0.835	0.462	0.965	1.000	0.681
West Pokot	0.439	0.642	0.255	0.154	0.271	0.070	0.611	0.872	0.265	0.184	0.345	0.077	0.374	0.993	0.001
Nandi	0.283	0.465	0.153	0.028	0.058	0.012	0.881	0.977	0.677	0.491	0.692	0.292	0.831	1.000	0.114
Bungoma	0.238	0.400	0.122	0.021	0.045	0.008	0.922	0.986	0.789	0.571	0.763	0.357	0.698	0.999	0.032
Busia	0.266	0.444	0.134	0.023	0.047	0.010	0.914	0.983	0.752	0.490	0.693	0.290	0.782	0.999	0.098
Kakamega	0.247	0.403	0.124	0.023	0.049	0.010	0.944	0.989	0.826	0.603	0.788	0.410	0.865	1.000	0.200
Vihiga	0.252	0.408	0.122	0.024	0.049	0.010	0.957	0.993	0.871	0.635	0.818	0.426	0.777	0.999	0.086

Appendix Table A2 Summarized admin 2 mean and 95% uncertainty interval for Ethiopia

ADM2_NAME	Mean stunting	Upper	Lower	Mean wasting	Upper	Lower	Mean DPT3	Upper	Lower	Mean ANC visits	Upper	Lower	Mean water sources	Upper	Lower	Mean women's anemia	Upper	Lower
Region 14	0.146	0.195	0.108	0.035	0.053	0.023	0.928	0.970	0.854	0.887	0.931	0.822	0.991	0.999	0.944	0.164	0.229	0.112
Awi/Agew	0.406	0.473	0.345	0.091	0.109	0.075	0.565	0.732	0.385	0.311	0.430	0.211	0.726	0.841	0.592	0.200	0.238	0.162
No Name	0.392	0.527	0.272	0.070	0.112	0.045	0.798	0.933	0.544	0.449	0.669	0.214	0.661	0.819	0.409	0.157	0.274	0.084
North Gonder	0.451	0.512	0.395	0.106	0.121	0.091	0.660	0.763	0.517	0.365	0.451	0.288	0.585	0.691	0.467	0.174	0.208	0.142
South Wollo	0.446	0.513	0.386	0.081	0.096	0.068	0.584	0.717	0.424	0.367	0.457	0.283	0.630	0.731	0.521	0.205	0.246	0.168
Special Woreda	0.457	0.559	0.354	0.106	0.144	0.075	0.501	0.779	0.227	0.289	0.508	0.127	0.679	0.894	0.418	0.229	0.302	0.166
Kemashi	0.356	0.419	0.298	0.112	0.131	0.095	0.632	0.785	0.468	0.284	0.378	0.200	0.714	0.830	0.581	0.221	0.265	0.183
Agnuak	0.274	0.319	0.233	0.133	0.157	0.114	0.503	0.607	0.399	0.330	0.393	0.262	0.825	0.893	0.741	0.286	0.331	0.244
Majang	0.270	0.340	0.208	0.107	0.134	0.084	0.663	0.827	0.474	0.398	0.506	0.300	0.756	0.877	0.607	0.177	0.221	0.137
Nuer	0.358	0.425	0.295	0.148	0.177	0.123	0.177	0.291	0.093	0.140	0.200	0.091	0.778	0.872	0.668	0.351	0.412	0.292
Arsi	0.364	0.431	0.305	0.079	0.093	0.066	0.521	0.666	0.370	0.313	0.404	0.229	0.720	0.812	0.611	0.289	0.338	0.244
Bale	0.388	0.464	0.319	0.096	0.111	0.083	0.457	0.630	0.286	0.248	0.349	0.168	0.574	0.710	0.435	0.359	0.418	0.299
East Shewa	0.309	0.369	0.254	0.088	0.105	0.073	0.598	0.758	0.432	0.466	0.573	0.351	0.883	0.937	0.807	0.225	0.261	0.192
East Wellega	0.309	0.379	0.244	0.095	0.112	0.079	0.522	0.705	0.333	0.327	0.442	0.216	0.663	0.796	0.513	0.194	0.237	0.158
Horo Guduru	0.347	0.432	0.269	0.087	0.106	0.070	0.450	0.678	0.241	0.248	0.378	0.142	0.592	0.761	0.415	0.205	0.251	0.163
Ilubabor	0.276	0.343	0.214	0.095	0.113	0.082	0.571	0.731	0.396	0.336	0.447	0.237	0.594	0.744	0.441	0.191	0.232	0.154
Kelem Wellega	0.331	0.407	0.267	0.108	0.128	0.090	0.627	0.804	0.423	0.371	0.504	0.251	0.700	0.833	0.560	0.212	0.252	0.173
West Arsi	0.399	0.459	0.344	0.071	0.085	0.058	0.482	0.630	0.351	0.316	0.414	0.227	0.694	0.801	0.582	0.308	0.356	0.266
West Wellega	0.344	0.410	0.284	0.102	0.121	0.086	0.718	0.867	0.543	0.351	0.463	0.243	0.704	0.818	0.569	0.212	0.253	0.178
Bench Maji	0.352	0.431	0.277	0.097	0.118	0.079	0.492	0.678	0.320	0.354	0.475	0.235	0.615	0.769	0.454	0.161	0.204	0.124
Segen Peoples'	0.439	0.523	0.355	0.103	0.127	0.086	0.629	0.791	0.421	0.421	0.563	0.293	0.403	0.568	0.243	0.237	0.299	0.182
Southern	0.403	0.458	0.347	0.097	0.114	0.080	0.653	0.757	0.546	0.529	0.603	0.459	0.830	0.892	0.758	0.204	0.241	0.171
Zone 1	0.500	0.561	0.439	0.171	0.199	0.146	0.198	0.302	0.115	0.190	0.253	0.135	0.570	0.684	0.456	0.460	0.518	0.402
Zone 2	0.395	0.454	0.338	0.171	0.201	0.146	0.196	0.291	0.117	0.160	0.218	0.115	0.535	0.652	0.407	0.455	0.520	0.400
Zone 3	0.453	0.517	0.393	0.159	0.188	0.132	0.136	0.220	0.073	0.124	0.177	0.082	0.610	0.726	0.489	0.412	0.474	0.353
Zone 4	0.499	0.573	0.431	0.141	0.167	0.116	0.210	0.334	0.118	0.146	0.213	0.092	0.515	0.662	0.377	0.360	0.415	0.309
Zone 5	0.421	0.481	0.362	0.142	0.169	0.120	0.260	0.388	0.151	0.220	0.295	0.154	0.535	0.654	0.408	0.385	0.438	0.335
East Gojam	0.463	0.531	0.391	0.080	0.095	0.066	0.510	0.670	0.339	0.257	0.339	0.184	0.565	0.694	0.424	0.198	0.237	0.165
North Shewa(R3)	0.456	0.523	0.395	0.096	0.113	0.080	0.523	0.677	0.376	0.297	0.391	0.214	0.580	0.701	0.456	0.198	0.234	0.165
North Wollo	0.445	0.519	0.374	0.110	0.133	0.088	0.478	0.673	0.289	0.283	0.413	0.177	0.661	0.789	0.502	0.249	0.297	0.207
Oromia	0.495	0.564	0.428	0.088	0.105	0.073	0.653	0.787	0.506	0.307	0.406	0.220	0.532	0.672	0.385	0.176	0.212	0.146
South Gonder	0.490	0.562	0.416	0.110	0.127	0.091	0.481	0.666	0.304	0.254	0.353	0.164	0.580	0.729	0.426	0.186	0.222	0.154
Wag Himra	0.425	0.487	0.367	0.085	0.100	0.072	0.612	0.747	0.471	0.296	0.374	0.221	0.666	0.775	0.552	0.186	0.223	0.155

Continued…

Appendix Table A2—*Continued*

ADM2_NAME	Mean stunting	Upper	Lower	Mean wasting	Upper	Lower	Mean DPT3	Upper	Lower	Mean ANC visits	Upper	Lower	Mean water sources	Upper	Lower	Mean women's anemia	Upper	Lower
West Gojam	0.443	0.496	0.386	0.116	0.135	0.098	0.827	0.897	0.733	0.453	0.530	0.383	0.831	0.895	0.751	0.220	0.263	0.181
Asosa	0.399	0.454	0.347	0.114	0.131	0.097	0.555	0.673	0.428	0.295	0.361	0.231	0.744	0.833	0.645	0.197	0.237	0.162
Metekel	0.315	0.400	0.236	0.095	0.133	0.068	0.861	0.924	0.751	0.675	0.783	0.528	0.933	0.969	0.855	0.285	0.389	0.197
Dire Dawa	0.312	0.369	0.258	0.082	0.097	0.067	0.407	0.565	0.264	0.265	0.357	0.182	0.542	0.667	0.401	0.171	0.206	0.140
Hareri	0.347	0.414	0.287	0.106	0.125	0.088	0.430	0.590	0.282	0.293	0.398	0.201	0.642	0.768	0.510	0.276	0.324	0.230
Borena	0.388	0.464	0.321	0.070	0.090	0.054	0.571	0.737	0.405	0.285	0.408	0.182	0.609	0.782	0.404	0.267	0.325	0.213
East Harerge	0.346	0.404	0.291	0.076	0.092	0.061	0.594	0.727	0.449	0.450	0.556	0.344	0.656	0.768	0.528	0.216	0.255	0.180
Guji	0.345	0.399	0.293	0.066	0.080	0.053	0.684	0.797	0.541	0.308	0.392	0.222	0.736	0.829	0.637	0.281	0.329	0.235
Jimma	0.430	0.517	0.341	0.122	0.142	0.105	0.494	0.687	0.292	0.300	0.416	0.198	0.397	0.575	0.237	0.254	0.323	0.190
North Shewa(R4)	0.349	0.427	0.273	0.087	0.113	0.066	0.512	0.742	0.285	0.375	0.552	0.215	0.484	0.701	0.272	0.196	0.248	0.150
South West Shewa	0.346	0.418	0.280	0.193	0.228	0.163	0.237	0.379	0.135	0.128	0.181	0.084	0.511	0.638	0.370	0.674	0.743	0.598
West Harerge	0.244	0.303	0.191	0.186	0.217	0.159	0.325	0.476	0.187	0.099	0.145	0.065	0.610	0.725	0.483	0.576	0.642	0.510
West Shewa	0.284	0.342	0.221	0.198	0.230	0.169	0.226	0.364	0.119	0.103	0.155	0.065	0.469	0.602	0.333	0.665	0.732	0.595
Alaba	0.295	0.355	0.239	0.133	0.156	0.111	0.304	0.463	0.174	0.208	0.300	0.131	0.565	0.700	0.413	0.435	0.503	0.369
Basketo	0.403	0.459	0.348	0.162	0.186	0.137	0.335	0.472	0.215	0.208	0.282	0.152	0.545	0.673	0.411	0.457	0.506	0.407
Dawro	0.229	0.304	0.164	0.218	0.259	0.184	0.207	0.367	0.095	0.085	0.137	0.050	0.571	0.711	0.427	0.678	0.761	0.586
Gamo Gofa	0.406	0.459	0.358	0.117	0.137	0.098	0.734	0.820	0.626	0.518	0.599	0.434	0.747	0.831	0.652	0.201	0.236	0.168
Gedio	0.398	0.456	0.342	0.105	0.125	0.086	0.783	0.872	0.668	0.595	0.683	0.502	0.766	0.861	0.666	0.219	0.266	0.180
Gurage	0.418	0.479	0.351	0.088	0.104	0.074	0.519	0.675	0.368	0.299	0.387	0.217	0.683	0.790	0.568	0.217	0.254	0.184
Hadiya	0.425	0.490	0.363	0.108	0.124	0.094	0.481	0.638	0.339	0.292	0.379	0.219	0.523	0.649	0.394	0.360	0.423	0.292
Keffa	0.371	0.418	0.323	0.111	0.128	0.096	0.541	0.651	0.432	0.289	0.365	0.222	0.645	0.746	0.544	0.332	0.367	0.296
Konta	0.448	0.521	0.374	0.091	0.109	0.075	0.471	0.636	0.311	0.239	0.331	0.161	0.524	0.672	0.377	0.347	0.405	0.291
KT	0.412	0.481	0.348	0.079	0.094	0.065	0.573	0.723	0.414	0.347	0.447	0.255	0.723	0.831	0.602	0.218	0.252	0.186
Selti	0.331	0.401	0.268	0.077	0.092	0.062	0.605	0.769	0.419	0.346	0.472	0.240	0.759	0.869	0.621	0.219	0.256	0.184
Sheka	0.339	0.406	0.278	0.080	0.093	0.067	0.548	0.698	0.388	0.263	0.351	0.182	0.667	0.786	0.537	0.213	0.250	0.180
Sidama	0.379	0.451	0.314	0.081	0.106	0.060	0.582	0.761	0.390	0.408	0.566	0.267	0.885	0.964	0.768	0.235	0.292	0.184
South Omo	0.470	0.584	0.348	0.095	0.130	0.068	0.457	0.786	0.163	0.221	0.442	0.086	0.533	0.811	0.242	0.156	0.227	0.103
Wolayita	0.365	0.445	0.295	0.090	0.110	0.073	0.409	0.619	0.215	0.244	0.363	0.146	0.394	0.568	0.236	0.176	0.215	0.140
Yem	0.416	0.482	0.350	0.086	0.103	0.071	0.621	0.760	0.462	0.380	0.484	0.280	0.501	0.648	0.359	0.204	0.248	0.163
Afder	0.346	0.408	0.289	0.076	0.092	0.062	0.559	0.715	0.395	0.470	0.576	0.359	0.746	0.845	0.622	0.237	0.278	0.199
Doolo	0.308	0.368	0.252	0.087	0.104	0.072	0.556	0.701	0.385	0.317	0.417	0.228	0.455	0.595	0.321	0.145	0.178	0.117
Fafan	0.344	0.406	0.280	0.073	0.092	0.056	0.637	0.789	0.468	0.400	0.526	0.281	0.688	0.814	0.538	0.207	0.251	0.168
Jarar	0.361	0.442	0.287	0.096	0.116	0.077	0.418	0.637	0.198	0.216	0.336	0.120	0.391	0.568	0.220	0.169	0.213	0.131

Continued…

Appendix Table A2—*Continued*

ADM2_NAME	Mean stunting	Upper	Lower	Mean wasting	Upper	Lower	Mean DPT3	Upper	Lower	Mean ANC visits	Upper	Lower	Mean water sources	Upper	Lower	Mean women's anemia	Upper	Lower
Korahe	0.272	0.332	0.215	0.089	0.112	0.071	0.672	0.828	0.496	0.447	0.570	0.326	0.617	0.756	0.464	0.151	0.191	0.116
Liben	0.373	0.445	0.310	0.076	0.094	0.060	0.583	0.754	0.398	0.525	0.646	0.399	0.834	0.917	0.722	0.238	0.284	0.199
Nogob	0.323	0.383	0.270	0.079	0.098	0.064	0.626	0.759	0.468	0.359	0.467	0.257	0.706	0.821	0.570	0.199	0.237	0.160
Shabelle	0.292	0.382	0.210	0.166	0.196	0.140	0.352	0.602	0.138	0.163	0.269	0.086	0.541	0.726	0.350	0.520	0.596	0.444
Siti	0.262	0.328	0.205	0.196	0.226	0.167	0.247	0.388	0.142	0.085	0.133	0.052	0.575	0.702	0.440	0.664	0.733	0.590
Central	0.433	0.512	0.351	0.156	0.183	0.133	0.366	0.553	0.218	0.197	0.275	0.135	0.528	0.663	0.374	0.620	0.705	0.533
Eastern	0.375	0.434	0.316	0.137	0.161	0.115	0.775	0.859	0.670	0.425	0.511	0.334	0.603	0.723	0.476	0.233	0.282	0.193
North Western	0.356	0.424	0.288	0.144	0.171	0.118	0.769	0.884	0.626	0.402	0.517	0.298	0.572	0.728	0.409	0.281	0.353	0.218
Western	0.285	0.385	0.206	0.071	0.104	0.047	0.701	0.840	0.492	0.429	0.612	0.235	0.883	0.949	0.740	0.268	0.392	0.163