Mapping access to basic hygiene services in low- and middle-income countries: A cross-sectional case study of geospatial disparities

Weiyu Yu[a,b], Robert E. S. Bain[c], Jie Yu[d], Victor Alegana[a,e], Winfred Dotse-Gborgbortsi[a], Yi Lin[b,f,*], Jim A. Wright[a]

[a] School of Geography and Environmental Science, University of Southampton, Building 44, Highfield, Southampton SO17 1BJ, UK
[b] Research Center of Remote Sensing & Spatial Information Technology, Shanghai 200092, China
[c] Division of Data, Analysis, Planning and Monitoring, United Nations Children's Fund (UNICEF), 3 United Nations Plaza, New York, NY 10017, United States
[d] College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China
[e] Population Health Unit, Kenya Medical Research Institute (KEMRI) – Wellcome Trust Research Programme, Nairobi, Kenya
[f] College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China

* Corresponding authors:
Yi Lin (Email: linyi@tongji.edu.cn)
College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China

Abstract

Handwashing with water and soap is among the most a cost-effective interventions to improve public health. Yet billions of people globally lacking handwashing facilities with water and soap on premises, with gaps particularly found in low- and middle-income countries. Targeted efforts to expand access to basic hygiene services require data at geospatially explicit scales. Drawing on country-specific cross-sectional Demographic and Health Surveys with georeferenced hygiene data, we developed an ensemble machine learning model to predict the prevalence of basic hygiene facilities in Malawi, Nepal, Nigeria, Pakistan and Uganda. The ensemble model was based on a multiple-level stacking structure, where four predictive modelling algorithms were used to produce sub-models, and a random forest model was used to generalise the final predictions. An inverse distance weighted interpolation was incorporated in the random forest model to account for spatial autocorrelation. Local coverage and a local dissimilarity index were calculated to examine the geographic disparities in access. Our methodology produced robust outputs, as evidenced by performance evaluations (all R^2 were above 0.8). Among the five study countries, Pakistan had the highest overall coverage, whilst Malawi had the poorest coverage. Apparent disparities in basic hygiene services measured by local coverage were found across geographic locations and between urban and rural settings. Nigeria had the highest level of inequalities in basic hygiene services measured by a dissimilarity index, whilst Malawi showed the least segregation between populations with and without basic hygiene services. Both educational attainment and wealth were important predictors of the geospatial distribution of basic hygiene services. By producing geospatially explicit estimates of the prevalence of handwashing facilities with water and soap, this study provides a means of identifying geographical disparities in basic hygiene services. The method and outputs can be useful tools to identify areas of low coverage and to support efficient and precise targeting of efforts to scale up access to handwashing facilities and shift social and cultural norms on handwashing.

Keywords

Basic hygiene; handwashing; water and soap; WASH; ensemble model; machine learning

Introduction

Hand hygiene is a measure of personal hygiene and a cost-effective non-pharmaceutical intervention to improve public health by preventing the transmission of infectious diseases (Loughnan, Ram, & Luyendijk, 2015; Warren-gash, Fragaszy, & Hayward, 2012). In light of disease and epidemic persistence throughout history, practicing of good hand hygiene is often recommended in public health guidelines and has remained a key component of personal level protection strategy during the recent pandemic events (WHO & UNICEF, 2020; World Health Organization, 2009, 2020). The World Health Organization (WHO)/United Nations Children's Fund (UNICEF) Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP) identified handwashing with water and soap, referred to as a basic hygiene service, as one of the top priorities for monitoring of progress towards the Sustainable Development Goals (SDG) Target 1.4 and Target 6.2 (WHO & UNICEF, 2018). Despite its importance, significant efforts are still required to increase the prevalence of hand hygiene with water and soap in many low- and middle-income countries (LMICs), particularly in poor and marginalised settings where people are disadvantaged by a lack of basic infrastructure and education (Loughnan et al., 2015; Renzaho, 2020). In a recent study, Brauer et al. (2020) estimated approximately two billion people globally still lacked access to basic handwashing facilities with water and soap at home in 2019, with barriers to universal access mostly in LMICs. At subnational level, large disparities in access were found across geographic locations and between urban and rural settings (Brauer et al., 2020; Jiwani & Antiporta, 2020). Access to a hygiene facility with water and soap can be extremely low even in urban areas in countries such as Malawi where local cleansing agents (e.g. ash, mud, etc.) are often chosen over soap as cheaper and more acceptable alternatives (Nguyen, 2015), although their effects on preventing disease transmission remain uncertain (Paludan-Müller, Boesen, Klerings, Jørgensen, & Munkholm, 2020). In this context, implementing hand hygiene interventions in response to an emergency situation such as the ongoing coronavirus diseases 2019 (COVID-19) pandemic can be challenging (Jiwani & Antiporta, 2020). Such circumstances call for rapid resource deployments by governments and development partners to scale up access to hand hygiene facilities with water and soap and shift social and cultural norms on handwashing (UNICEF & WHO, 2020), which in turn requires knowledge about hand hygiene facilities and behaviours at the sub-provincial level.

Nationally representative household surveys such as the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS) are often the key sources of data on hand hygiene for LMICs. Being designed for multiple purposes, DHS and MICS surveys rely on rapid observations of hygiene facilities to balance cost-effectiveness and representativeness (Ram, 2013). In comparison with other small-scale data collected via rigorous methods such as structured observations on behaviours, these surveys cover larger geographic extents and a wider range of demographic and health-related characteristics. However, due to survey sampling design and confidentiality protection, these household surveys often cannot be disaggregated spatially to a greater extent (ICF International, 2012). Other recent efforts to estimate coverage of basic hygiene services have also been limited to the national or the first subnational administrative levels (Brauer et al., 2020). To gain a more detailed view of access to basic hygiene services from household surveys and in order to inform development policy, resource deployment and intervention implementation, local coverage estimates should be produced for all locations within a country.

A common approach to producing geospatially explicit estimates for all locations is through Bayesian geostatistical modelling (Diggle, Tawn, & Moyeed, 1998) which quantifies the relationship between the prevalence at observed locations with potential predictive covariates whilst accounting for spatial dependence via a covariance matrix of a Gaussian process with location-specific random effects (Karagiannis-Voules, Scholte, Guimarães, Utzinger, & Vounatsou, 2013; Lai, Zhou, Utzinger, & Vounatsou, 2013). Implementation of such Bayesian geostatistical models is often through the Markov Chain Monte Carlo (MCMC) algorithm. Because of the large covariate matrices involved, MCMC suffers drawbacks such as lack of convergence, high storage requirements, and high computational cost (Lai et al., 2013; Mayala, Dontamsetti, Fish, & Croft, 2019). There have been efforts to address these issues, among which the Integrated Nested Laplace Approximation (INLA) methodology (Rue, Martino, & Chopin, 2009) has become increasingly popular (Dwyer-Lindgren et al., 2019; Local Burden of Disease WaSH Collaborators, 2020; Mayala et al., 2019; Mosser et al., 2019; Osgood-Zimmerman et al., 2018; Reiner et al., 2018). INLA constructs a triangulation (namely 'mesh') over the study area, and computes the spatial autocorrelation structure of the dataset at the mesh vertices using a stochastic partial differential equation (SPDE) approach (Lindgren, Rue, & Lindström, 2011). In comparison with MCMC methods, INLA provides a significant speed boost with high accuracy (Lindgren & Rue, 2015). However, this computationally efficient approach generates an approximation with potential drawbacks such as the boundary effect in the covariance approximation due to boundary conditions of the SPDE (Lindgren et al., 2011).

In this study, we adopted an ensemble model of machine learning algorithms as an alternative to geostatistical models to produce geospatially explicit estimates for all locations within the case study countries. In comparison with geostatistical models, machine learning predictive models often require fewer statistical assumptions and can be flexibly automated (Hengl, Nussbaum, Wright, Heuvelink, & Gräler, 2018). To the best of our knowledge, such machine learning models have not yet been used for mapping the prevalence of basic hygiene services. The main objectives of this study are (1) to produce geospatially explicit estimates of basic hygiene prevalence across the case study countries; (2) to examine the applicability of the ensemble machine learning model for such applications; (3) to examine the relative importance of covariates for predicting basic hygiene services; and (4) to quantify geographic disparities in access to basic hygiene services using the resultant geospatial estimates of basic hygiene services.

Materials and Methods

Study countries and sample data

In this study, we selected five case study countries, namely Malawi, Nigeria and Uganda in sub-Saharan Africa, and Nepal and Pakistan in South Asia, for two reasons. Firstly, these developing countries are classified into different groups (i.e. low- and middle-income) by the World Bank (World Bank, 2020b), located in separate geographic regions, and represent differing socio-economic and demographic contexts. Estimated national coverage of basic hygiene services in these five countries ranges from 9% (uncertainty interval: 6%, 12%) in Malawi to 63% (uncertainty interval: 53%, 72%) in Pakistan (Brauer et al., 2020). Secondly, all five countries had recent (post-2015) georeferenced DHS datasets and recent geospatial datasets characterising factors affecting access to basic hygiene services, particularly poverty. **Figure 1** shows the locations of the five study countries and clusters (i.e. the groupings of households participated in the DHS campaign) from the latest DHS. We obtained country-specific standard DHS data from the most recent round via the DHS Program portal (ICF, 2018), covering household survey and geographic data for Malawi DHS 2015-16 (National Statistical Office (NSO) [Malawi] & ICF, 2017), Nepal DHS 2016 (Ministry of Health [Nepal], New ERA, & ICF, 2017), Nigeria DHS 2018 (National Population Commission (NPC) [Nigeria] & ICF, 2019), Pakistan DHS 2017-18 (National Institute of Population Studies (NIPS) [Pakistan] & ICF, 2019) and Uganda DHS 2016 (Uganda Bureau of Statistics (UBOS) & ICF, 2018). These DHS household surveys included observation-based measurements of handwashing facilities. During the survey, interviewers asked respondents to show them the location most often used by household members for handwashing (Croft, Marshall, Allen, & et al., 2018; ICF, 2019). Where feasible, presence of water and cleansing agent were directly observed at the place for handwashing and recorded by the interviewer. The derived data therefore allow us to capture the proportion of population living in a household with an observed place for handwashing where both water and soap are available.

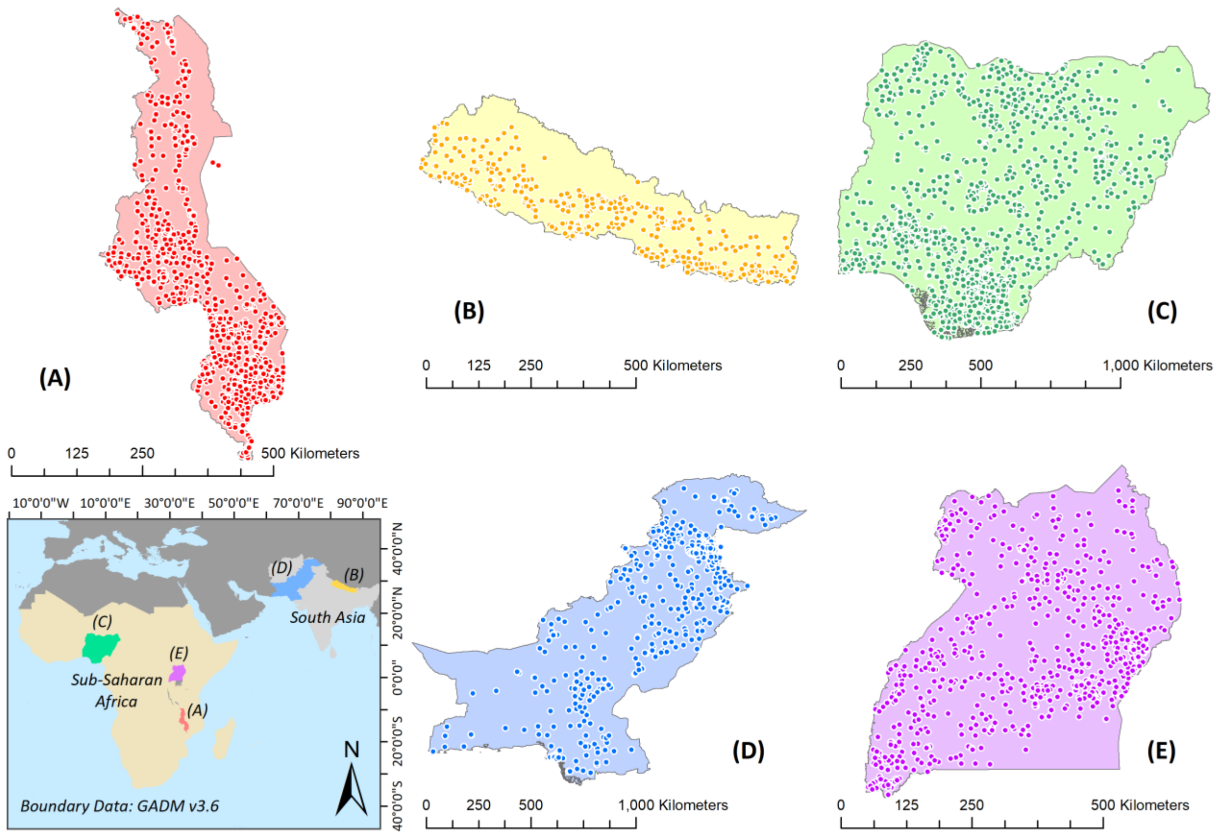


Figure 1. Map showing the geographic locations of (A) Malawi, (B) Nepal, (C) Nigeria, (D) Pakistan and (E) Uganda, together with cluster point locations (coloured dots) for the latest DHS.

The DHS household surveys were based on a stratified two-stage cluster design (Croft et al., 2018) and weights are used to adjust for non-response and disproportionate sampling in regions with smaller and larger populations. In this study, the DHS household weight was multiplied by the number of *de jure* household members (i.e. usual residents living in the surveyed household) in order to produce estimate for the proportion of the population. Derived information on individuals living in a household with access to basic hygiene was aggregated to cluster level and then combined with the DHS geographic data. Households with missing data on handwashing with water and soap (only in Pakistan, <0.1%) were assumed to lack basic hygiene. Generated tabular data therefore included the outcome variable – the proportion of *de jure* population (hereafter referred to as 'population' in this study) living in a household with an observed handwashing facility where water and soap were present - alongside Global Positioning System (GPS) coordinates (longitude and latitude) for the DHS clusters. These GPS coordinates were provided with displacement of up to 2km for urban locations and up to 5km for rural locations (and up to 10km for every 100th rural location) to protect respondent confidentiality (Burgert, Colston, Roy, & Zachary, 2013; Perez-Heydrich, Warren, Burgert, & Emch, 2013). By overlaying these georeferenced cluster points with pre-processed geospatial covariate layers, those with implausible coordinates (i.e. without GPS coordinates or outside the spatial extent of the pre-processed covariate layers) were excluded. Multiple clusters located within the same 5km grid cell were merged to form a new cluster, with the outcome variable calculated from the total number of household members and those with basic hygiene within that 5km grid cell. **Table 1** summarises the characteristics of the georeferenced DHS household survey data describing basic hygiene services and the number of cluster points included as the final sample for each study country.

Table 1. Characteristics of the obtained country-specific DHS data describing basic hygiene services

	Malawi	Nepal	Nigeria	Pakistan	Uganda
Sample size (number of households)	26,361	11,040	40,427	14,540	19,588
Number of households with hygiene information	26,361 (100.0%)	11,040 (100.0%)	40,427 (100.0%)	14,535 (>99.9%)	19,588 (100.0%)
Number of households without hygiene information or missing data	0 (0.0%)	0 (0.0%)	0 (0.0%)	5 (<0.1%)	0 (0.0%)
Number of households having water and soap at the observed place for handwashing	2,596 (9.8%)	4,991 (45.2%)	11,042 (27.3%)	8,892 (61.2%)	4,788 (24.4%)
Survey respondent:					
Household's head	14,618 (55.5%)	7,425 (67.3%)	31,533 (78.0%)	4,514 (31.1%)	12,481 (63.7%)
Wife/husband of household's head	9,898 (37.5%)	2,331 (21.1%)	6,342 (15.7%)	6,344 (43.6%)	5,690 (29.0%)
Son/daughter of household's head	1,021 (3.9%)	586 (5.3%)	1,713 (4.2%)	1,822 (12.5%)	803 (4.1%)
Total number of household members in the sampled households	119,326	47,026	186,450	98,895	89,202
Total number of clusters	850	383	1,389	561	696
Number of clusters excluded – without geographic data	0 (0.0%)	0 (0.0%)	30 (2.2%)	1 (0.2%)	11 (1.6%)
Number of clusters excluded – located outside the spatial extent of the covariate layers	66 (7.8%)	8 (2.1%)	14 (1.0%)	2 (0.4%)	3 (0.4%)
Number of clusters merged – had other neighbouring cluster(s) located within the same 5km grid cell	164 (19.3%)	33 (8.6%)	148 (10.7%)	106 (18.9%)	91 (13.1%)
Number of clusters used in this study (final sample size)	620 (72.9%)	342 (89.3%)	1,197 (86.2%)	454 (80.9%)	591 (84.9%)

Geospatial covariates

We identified candidate covariates for their potential to predict basic hygiene prevalence based on theory and existing literature on either factors influencing hygiene practices (Luby & Halder, 2008; Wolf et al., 2019) or mapping of water, sanitation and hygiene (WASH) access (Brauer et al., 2020; Ekumah et al., 2020; Gething, Tatem, Bird, & Burgert-Brucker, 2015; Mayala et al., 2019; Yu et al., 2019). For socio-economic factors, we included candidate geospatial covariates characterising population density, proximity to human settlement, and proximity to cropland areas which may reflect the socio-economic and demographic contexts and in turn potentially affect access to hygiene facilities. In addition, since piped water has previously been applied as a covariate in mapping global access to basic hygiene (Brauer et al., 2020), and since the presence of handwashing items may relate to household amenities given the need for handwashing after toilet use for example (Wolf et al., 2019), we included existing modelled map surfaces of improved water sources and lack of sanitation from the DHS Spatial Data Repository (Gething et al., 2015; Mayala et al., 2019) as proxies. We also included literacy as an educational attainment outcome, also previously included as a component of a socio-demographic index in predicting coverage of handwashing with water and soap (Brauer et al., 2020). Moreover, since wealth has previously been shown to correlate with handwashing (Luby & Halder, 2008), we calculated a global poverty index following the methodology described in Elvidge et al. (2009) using recent population and night-time lights datasets obtained from WorldPop (WorldPop & Center for International Earth Science Information Network (CIESIN) Columbia University, 2018) and the Earth Observation Group (EOG) at the National Oceanic and Atmospheric Administration (NOAA)/National Geophysical Data Center (NGDC) (C. D. Elvidge, Zhizhin, Ghosh, Hsu, & Taneja, 2021) . Floating point radiance values from the Visible Infrared Imaging Radiometer Suite (VIIR) Day/Night Band (DNB) sensor were rescaled to 1-100 before calculation so as to avoid any numeric difficulties. Since satellite-observed stable night-time lights have been found to correlate with economic activities (Pinkovskiy & Sala-i-Martin, 2016), we also included stable night-time lights directly as a proxy locational metric of economic status. Further geospatial data characterising health conditions were also selected (see details in **Table S.1** in **Supplementary Material 1)**, including women with anaemia, child mortality and objective undernutrition measures, and infectious disease (i.e. diarrhoea, lower respiratory infection and human immunodeficiency virus infection) prevalence as indicators reflecting socio-economic status and thereby poor hygiene (Curtis et al., 2011; Ejemot, Ehiri, Meremikwu, & Critchley, 2008; Fewtrell et al., 2005; Huang & Zhou, 2007; Jefferson et al., 2011; Luby et al., 2005; Luby & Halder, 2008; Rah et al., 2015). For environmental factors, we included candidate covariates describing elevation, slope, precipitation, aridity and potential evapotranspiration (PET) previously applied in predictive mapping of water sources and sanitation facilities (Gething et al., 2015; Mayala et al., 2019; Yu et al., 2019). Wherever possible, we obtained geospatial datasets from publicly available data sources for the same year as the most recent DHS surveys to create candidate covariate layers of these factors (as detailed in **Table S.1** in **Supplementary Material 1**). The final inclusion of candidate covariates was country-specific subject to data availability. Exceptions were some environmental covariates which were either long-term means (e.g. prediction, aridity, and potential evapotranspiration) or assumed to be temporally static (e.g. elevation and slope).

Subject to data availability, we kept the data source of any given candidate covariate identical across the five study countries as far as possible. All covariate layers were prepared at a spatial resolution of 0.05 degrees (approximately 5km) due to the random displacement of DHS GPS cluster point locations (Burgert, Zachary, & Colston, 2013; Perez-Heydrich et al., 2013). Covariate data with higher resolutions were aggregated (for population counts and night-time light radiance values) or resampled (by nearest neighbour for discrete data and by bilinear interpolation for continuous data) to 5km resolution. Large water bodies identified in any source data layers were excluded, retaining the same spatial extent for all covariate layers. To reduce collinearity, we excluded strongly correlated covariate pairs ($|r| > 0.7$) as recommend by previous studies (Merow, Smith, & Silander, 2013; Yu et al., 2019), retaining the covariate in each pair least correlated overall with other covariates. Data pre-processing was performed using ArcGIS 10.4.1 (ESRI, Redlands, CA, USA).

Mapping prevalence of basic hygiene with an ensemble model

We adopted an ensemble machine learning model based on a model stacking (Wolpert, 1992) approach to predict basic hygiene prevalence. Model stacking is a robust ensemble method that combines outputs of multiple modelling algorithms to improve prediction, in which the final combination rule is a generalised modelling algorithm instead of voting or averaging. The process of model stacking is often arbitrarily implemented via many levels, where the predictions generated by the modelling algorithms in a level become the inputs in the next level, until being generalised into the final prediction by the final level modelling algorithm (known as the 'meta-model' in machine learning). This ensemble approach has been applied in global mapping of basic hygiene at national level (Brauer et al., 2020), water and sanitation (Local Burden of Disease WaSH Collaborators, 2020), as well as other demographic and health-related indicators of interest such as child growth failure (Osgood-Zimmerman et al., 2018), vaccine coverage (Mosser et al., 2019), and disease prevalence (Dwyer-Lindgren et al., 2019; Reiner et al., 2018). It has been shown to out-perform other conventional ensemble techniques (e.g. averaging) (Clarke, 2003). In this study, we adopted a multiple-level model stacking system as depicted in **Figure 2**. In the first level, we fitted four sub-models with the pre-processed cluster point data and geospatial covariate layers using four assumption-free, nonparametric predictive modelling algorithms for regression problems: (a) Boosted Generalised Additive Model (BGAM) (Amodio & Meulman, 2011); (b) Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991); (c) Classification and Regression Trees (CART) (L. Breiman, Friedman, Stone, & Olshen, 1984); and (d) Gradient Boosting Machines (GBM) (Friedman, 2001). These modelling algorithms were selected based on predictive accuracy, computational cost, availability of efficient built-in feature selection and ease of automatic parameter tuning in the R computational environment. The sub-model predictions generated using the geospatial covariates were then used as exploratory predictors in the next level meta-model for a generalised final prediction. We employed Random Forest (RF) (Leo Breiman, 2001) as the meta-model algorithm given its good predictive performance and advantages such as having fewer hyper-parameters to tune (Stevens, Gaughan, Linard, & Tatem, 2015). Since such machine learning predictive model does not account for spatial autocorrelation, the meta-model's predictors additionally included a raster layer generated from sample cluster points using inverse distance weighting (IDW) interpolation (Philip & Watson, 1982). This was based on the rationale of random forest for spatial predictions (RFsp) (Hengl et al., 2018), but with buffer distances replaced by a simpler deterministic estimation method to reduce the computational cost. The outcome variable was therefore modelled as a function of the sub-model predictions and the inverse distance weighted interpolation. Model performance was evaluated by calculating the coefficient of determination (R-squared), root mean squared error (RMSE), and mean absolute error (MAE).

Model fitting, performance evaluation, prediction and analysis were carried out using R 3.5.2 (The R Core Team, 2020) in RStudio v1.1.383 (RStudio Team, 2016); whilst the IDW interpolation was performed using ArcGIS 10.4.1 (ESRI, Redlands, CA, USA). Each model was fitted using a random subsample of 70% of the selected clusters, whilst the remainder was used to test model performance based on Monte Carlo cross validation. For each country, we repeated the subsampling of training/test data, model fitting, model evaluation and prediction 100 times, and then computed summary grids (mean and 95% confidence level) of final predictions and model performance metrics from all 100 model runs. Since our outcome variable was a proportion, an empirical logit transformation was performed before the meta-model. The logit transformation ensures that the final predictions can be converted back to a proportion value bounded by 0 and 1, so as to be easily interpreted and combined with population data for further analysis and validation. Data output and R code for key steps are provided in an online repository at:
https://figshare.com/articles/online_resource/BasicHygieneMapping_R/14529135.

As an additional means of model performance evaluation, we aggregated the generated 5km-grid level prediction of each country to national and subnational boundaries depicted via publicly available data sources. Data on population, urban-rural divide and subnational level administrative boundaries were respectively derived from the WorldPop gridded population (Tatem, 2017), the European Commission's Global Human Settlement Layer (GHSL) Settlement Model grid (GHS-SMOD) datasets (Pesaresi, Florczyk, Schiavina, Melchiorri, & Maffenini, 2019), the Global Administrative Areas (GADM) database v3.6 (Hijmans, Garcia, & Wieczorek, 2018), and the new Nepal administrative boundary database digitised by the Hermes GIS team (Hermes GIS team, 2020) . For countries where the corresponding DHS report included estimates of basic hygiene coverage for subnational areas, Spearman's Rho was employed to examine the correlation across subnational areal units between the DHS-reported coverage and our resultant estimates.

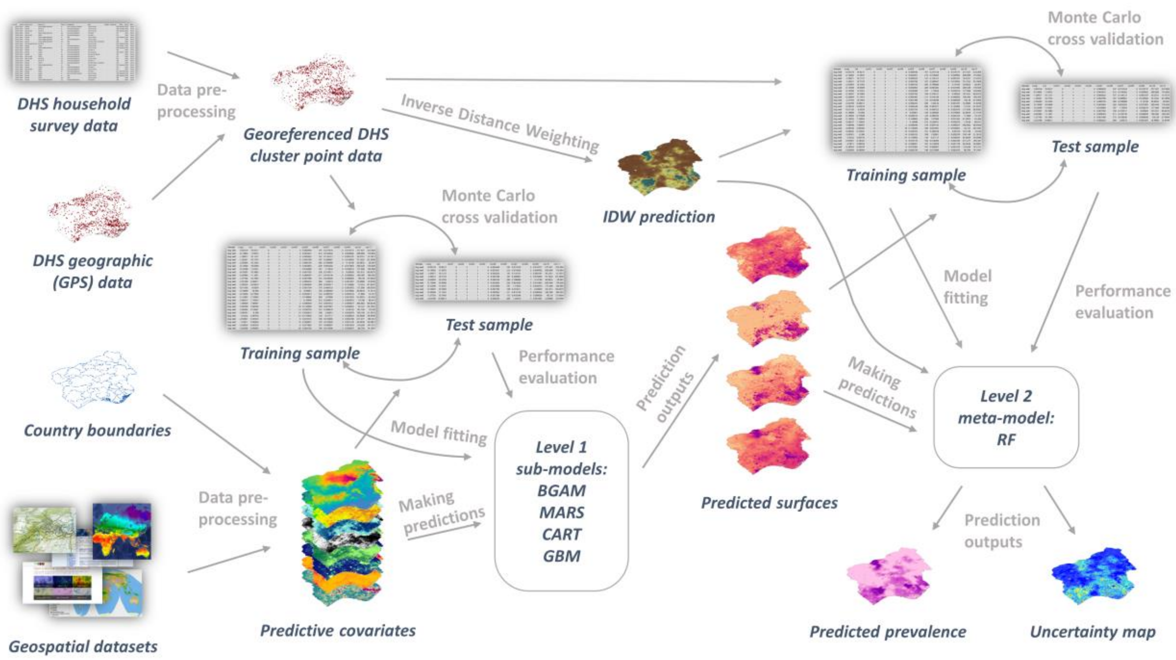


Figure 2. Flowchart of the model stacking methodology used in this study.

Examining spatial disparities in access to basic hygiene

To examine subnational disparities in basic hygiene services by area, the predicted prevalence of basic hygiene was aggregated to administrative level 2 for illustration by integrating with gridded population data (Tatem, 2017). Using the resultant basic hygiene estimates, we calculated a dissimilarity index (Duncan & Duncan, 1955; Yu, Bain, Mansour, & Wright, 2014) as a measure of geographic inequality in access to basic hygiene services. This index measures the proportion of people in the total population who would have to shift location for basic hygiene services to be completely evenly distributed throughout all areas. We chose this dissimilarity index as it has a simple functional form which can be more straightforwardly decomposed to different geographic levels. The calculations were conducted at all available administrative levels, given the scale-dependency of this dissimilarity measure (Yu et al., 2014). The administrative boundaries were derived from the GADM v3.6 database for most study countries, except Nepal where newly updated administrative boundaries digitised by the Hermes GIS team were used instead. Local contributions to national level inequality were also mapped, so as to reveal administrative level 2 areas contributing strongly to the overall disparity. Moreover, the predicted prevalence of basic hygiene was further broken down by type of human settlement following a classification system based on cluster population size, population density and built-up area density (Florczyk et al., 2019).

Ethics statement

Ethical clearance for this study was obtained from the Faculty of Social and Human Sciences, University of Southampton through the Ethics and Research Governance Online (ERGO) system (reference: 57472; approved on 17th June 2020).

Results

For all study countries, our ensemble models display good performance (all R^2 above 0.8; see **Table S.3** in **Supplementary Material 1**), suggesting the majority of variance in the data was explained. In most cases, the predictions at national and urban-rural levels are broadly in line with the figures in the DHS country reports

(**Table S.4** in **Supplementary Material 1**), with the exceptions of urban Nigeria, urban and rural areas in Nepal and Pakistan where the differences between DHS reports and predicted coverages were greater than 5%. At subnational level, our estimates show patterns consistent with the DHS reported areal coverages (Nepal: r_s=1.00, n=7; Nigeria: r_s=0.94, n=37; Pakistan: r_s=0.98, n=8; Uganda: r_s=0.99, n=15; insufficient disaggregated data for Malawi). For the performance of sub-model algorithms, GBM is shown to out-perform the others in all cases according to the sub-model performance evaluation metrics (**Table S.3** in **Supplementary Material 1**). However, for most of the sub-model algorithms except GBM, this ranking is inconsistent with the contribution of sub-model predictions to the meta-model (**Table S.5** in **Supplementary Material 1**).

The relative contribution of the covariates varied by sub-model algorithm and by country (**Table S.6** in **Supplementary Material 1**). However, in most cases, the covariate with the highest importance in a sub-model also had high importance in other parallel sub-models. The only exceptions included anaemia in women (ranked first in the CART sub-model but 9th out of 12 covariates in the BGAM sub-model) for Nigeria and night-time lights (ranked first in the CART sub-model but 16th out of 20 covariates in the BGAM sub-model) for Uganda. For Malawi, child underweight had the greatest contribution to all sub-models. For Nepal, children stunting was found to be the most influential covariate. Population density had the greatest contribution for Nigeria. In Pakistan, women's literacy and open defecation were found most important. In contrast, for Uganda, potential evapo-transpiration had the greatest contributions. Across the five study countries and different sub-models, literacy (for men and/or women) provided useful information for modelling basic hygiene services in most cases (**Table S.6** in **Supplementary Material 1**). Surprisingly, in most cases except Malawi, access to improved water did not provide useful information for modelling basic hygiene services. This is similar for lack of sanitation (open defecation) except in Pakistan.

Figures 3–4 respectively show the spatially continuous estimates of the proportion of people in households with access to basic hygiene for each 5km×5km grid cell for the five case study countries and corresponding uncertainty maps based on width between 95% confidence intervals. **Figure 5** shows the aggregated estimates for second administrative level areas, which highlights areas with high and low basic hygiene coverage. Household access to basic hygiene varies considerably by area – across geographic locations, by subnational areal units, and between urban-rural settings. Among the five case study countries, Malawi has comparatively little geographic variation in basic hygiene coverage, with all areas less than 50% at the 5km-grid level (**Figure 3A**), or less than 40% at the second administrative level (**Figure 5A**). Uganda, with a mostly rural population, also has poor basic hygiene services across most of the country, with relatively higher basic hygiene coverage in the south, particularly in areas bordering Lake Victoria. For the other three countries, basic hygiene coverage varies more geographically – for example, in Nepal, from 6.8% in Dolpa District, Karnali Pradesh to 84.5% in Kathmandu District, Bagmati Pradesh (**Figure 5B**). Pakistan has the highest coverage among the five study countries, with basic hygiene prevalent in northern Punjab, around Lahore (**Figure 5D**), and coastal areas, such as Karachi.

Patterns in basic hygiene coverage (**Figure 5**) differed from the mapped local contributions to the national level inequalities (**Figure 6**). While the former measured the proportion of people with access to basic hygiene services, the latter measured the magnitude of segregation of population sub-groups with access versus those without access. For example, in Nigeria, the geographic disparities in basic hygiene services are pronounced with higher (> 60%) basic hygiene coverage concentrated in the south and Katsina State in the north, with very low (< 10%) coverage in most parts of northern and central Nigeria (**Figures 3C & 5C**). However, the map of local contributions to the dissimilarity index illustrated more spatially homogenous patterns, with stronger contributors also noticeable among areas where access to basic hygiene services was low, such as in the North East and North West. For the national level dissimilarity index, greater inequality in access is apparent in Nigeria (**Figure 7**), whilst Malawi, with the lowest levels of basic hygiene coverage, has consistently lower dissimilarity index values reflecting lower spatial inequalities in access relative to the other four countries. All five countries show gaps between urban and rural areas, with lower coverage apparent in rural areas in all cases. **Table 2** shows inequalities in access to basic hygiene services by type of human settlement. Pakistan has the highest coverage over all types of human settlement (mostly greater than 60%), with the lowest coverage in rural areas of very low density (56.3%, uncertainty interval: 56.0%, 56.5%). The highest coverage of basic hygiene can be found in dense urban clusters in Malawi, in suburb/peri-urban areas in the other two Sub-Saharan African study countries, and in urban centres for the two South Asian study countries, respectively.

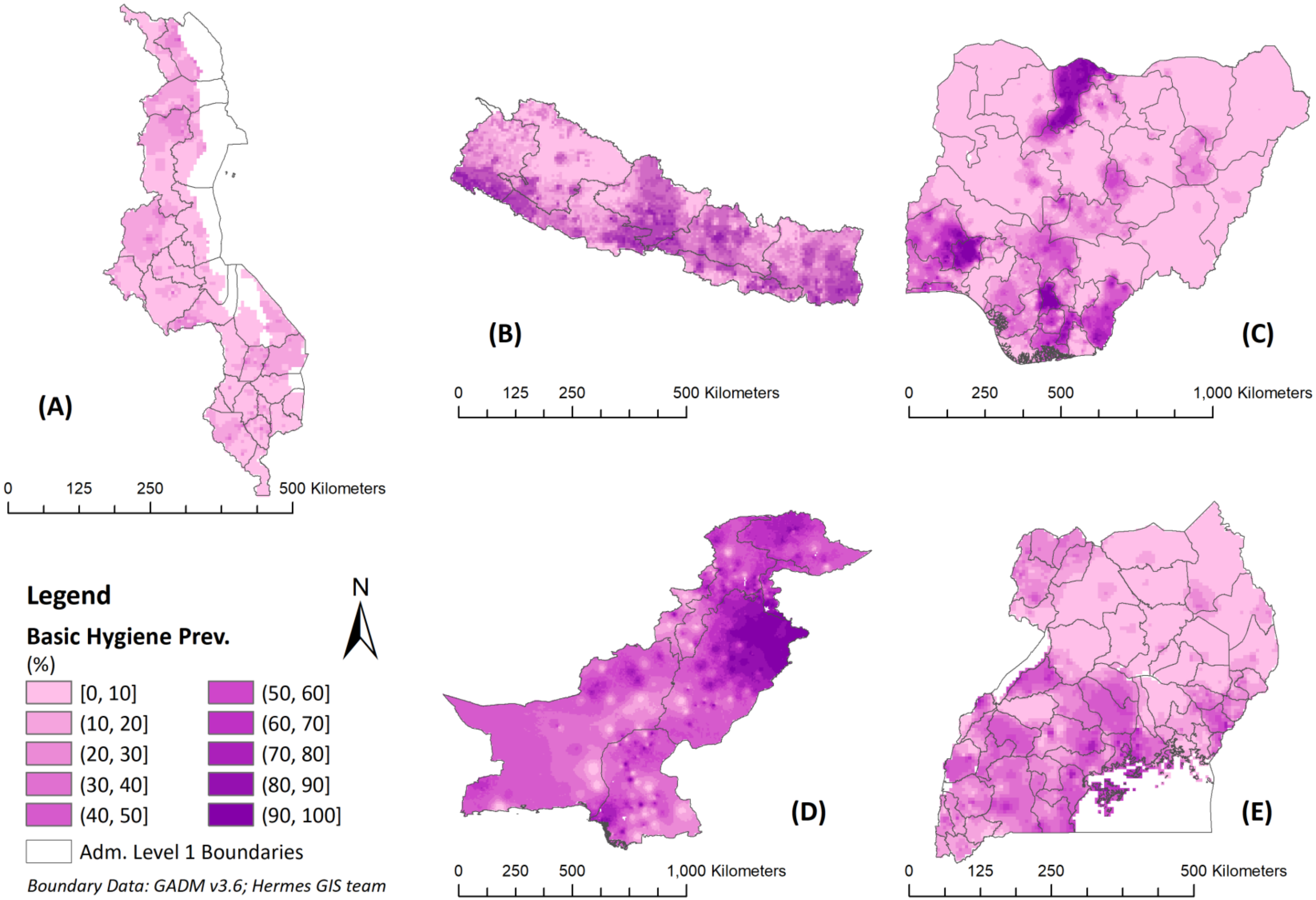


Figure 3. Modelled surfaces showing the estimated proportion of population living in a household with access to basic hygiene for (A) Malawi, (B) Nepal, (C) Nigeria, (D) Pakistan, and (E) Uganda

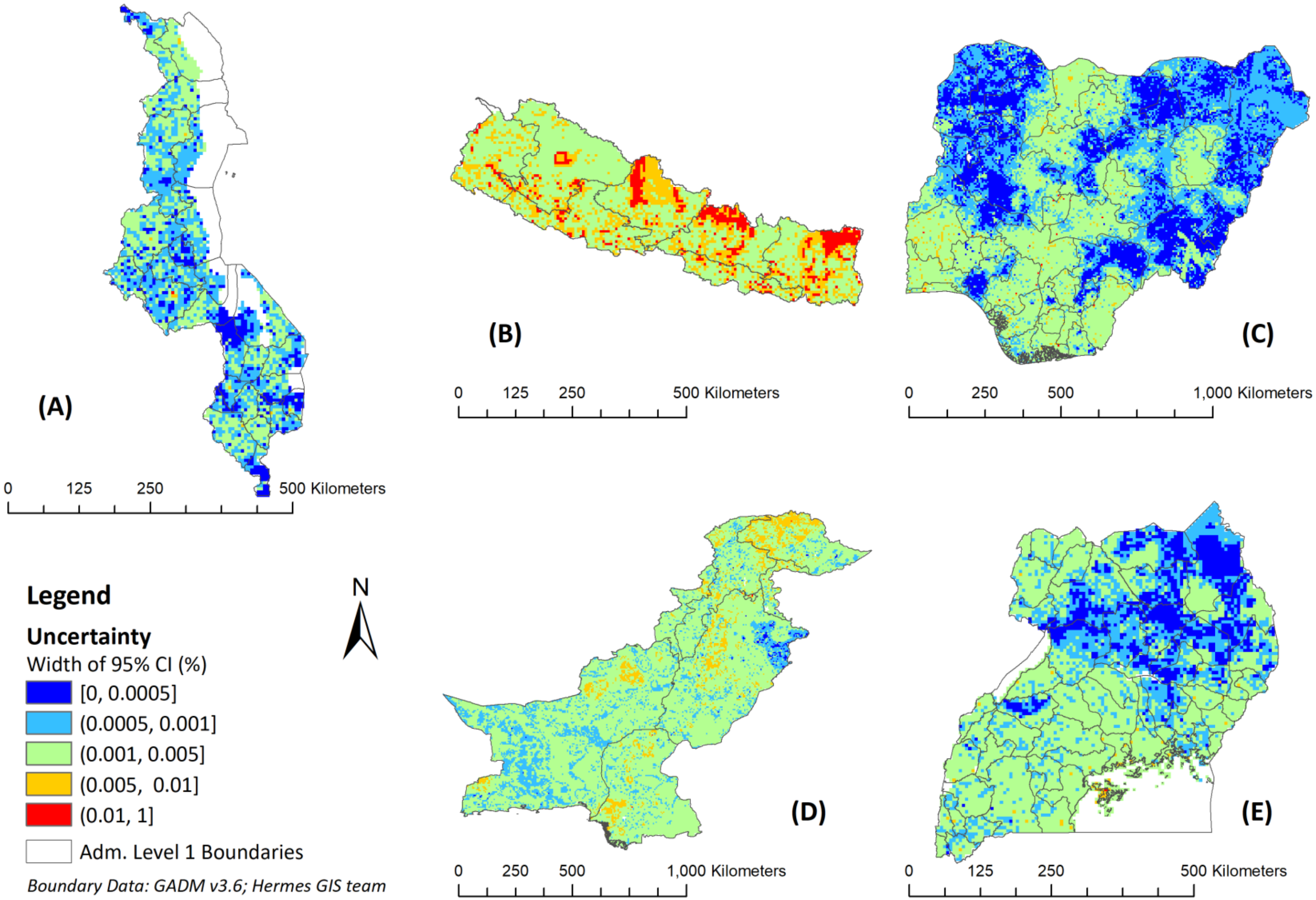
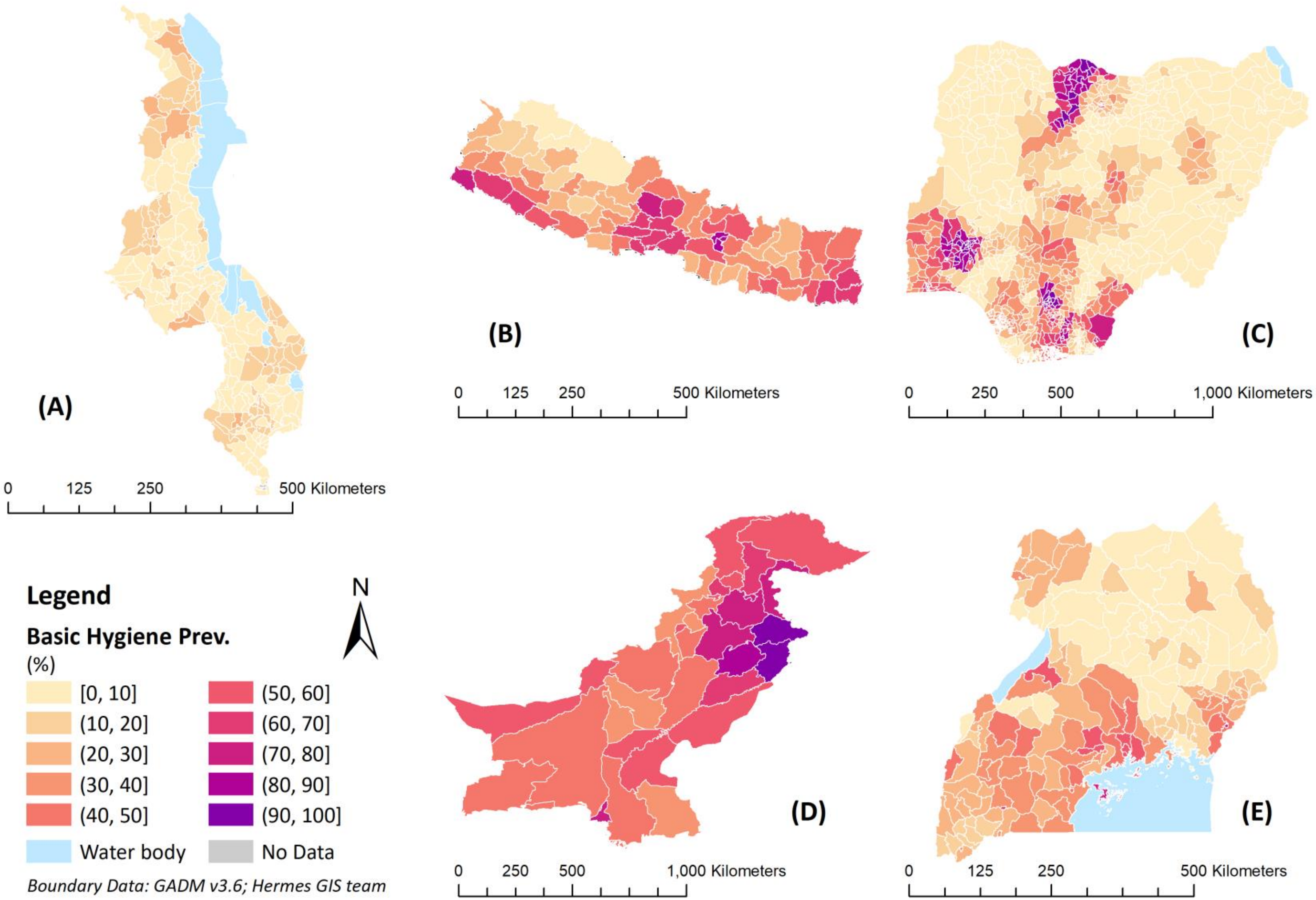


Figure 4. Uncertainty surfaces measured using the width of the 95% confidence intervals for (A) Malawi, (B) Nepal, (C) Nigeria, (D) Pakistan, and (E) Uganda



346

347 **Figure 5.** Estimated coverage of basic hygiene services at administrative level 2 for (A) Malawi, (B) Nepal, (C) Nigeria, (D) Pakistan, and (E) Uganda

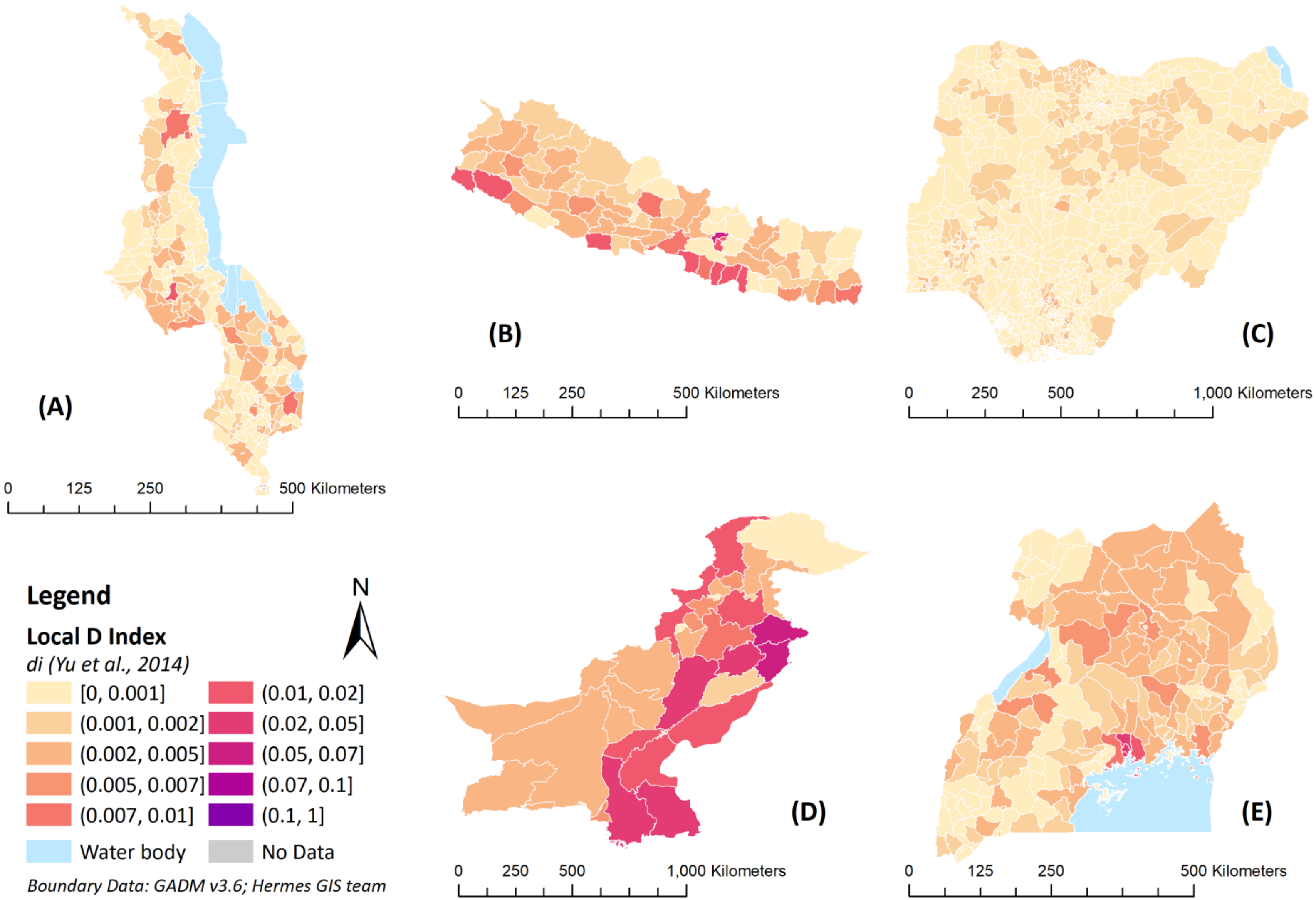


Figure 6. Map showing administrative level 2 local contributions to the national level dissimilarity index for basic hygiene services for (A) Malawi, (B) Nepal, (C) Nigeria, (D) Pakistan, and (E) Uganda

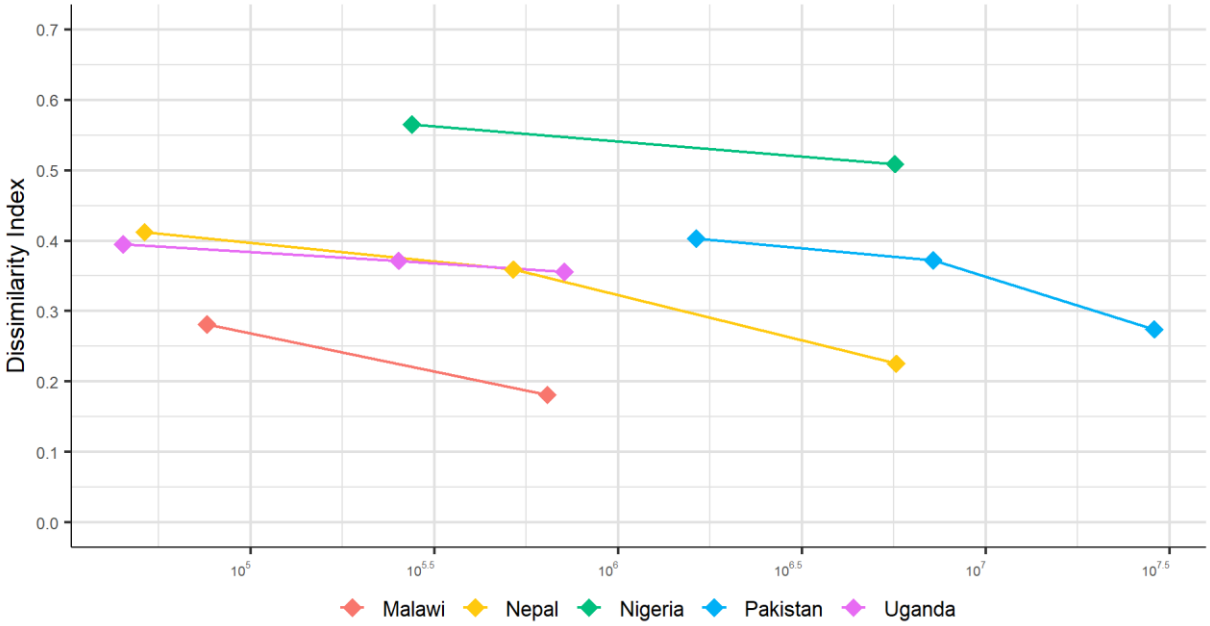


Figure 7. Dissimilarity indices for basic hygiene services

Y-axis represents the value of the national level dissimilarity index for each administrative tier; X-axis is the mean population size of the administrative units (from administrative level 1 on the left to level 2 or level 3 on the right for each country) in base-10 log scale.

Table 2. Percentage of population living in a household with basic hygiene by type of human settlement

Settlement Type	Malawi	Nepal	Nigeria	Pakistan	Uganda
Urban Centres	14.7 (14.3, 15.1)	78.6 (78.2, 79.0)	51.2 (50.9, 51.5)	77.2 (76.9, 77.4)	48.4 (48.1, 48.7)
Dense Urban Clusters	19.0 (18.8, 19.2)	47.0 (46.4, 47.6)	30.2 (30.0, 30.4)	67.4 (67.2, 67.6)	22.1 (21.9, 22.3)
Semi-Dense Urban Clusters	-	39.3 (38.3, 39.7)	27.1 (26.9, 27.4)	63.6 (63.4, 63.8)	-
Suburb/Peri-Urban Areas	6.3 (6.2, 6.3)	45.2 (44.6, 45.7)	51.3 (51.0, 51.6)	70.5 (70.3, 70.7)	48.6 (48.3, 49.0)
Rural Clusters	9.0 (8.8, 9.1)	43.1 (42.6, 43.6)	27.3 (27.2, 27.5)	68.5 (68.3, 68.7)	29.6 (29.4, 29.8)
Low Density Rural Areas	8.2 (8.1, 8.3)	38.5 (38.0, 39.0)	31.0 (30.8, 31.2)	68.1 (67.9, 68.3)	24.5 (24.4, 24.7)
Very Low Density Rural Areas	8.4 (8.3, 8.5)	48.3 (47.8, 48.9)	22.1 (21.9, 22.2)	56.3 (56.0, 56.5)	21.7 (21.6, 21.9)
National Total	9.2 (9.1, 9.4)	50.9 (50.4, 51.4)	30.8 (30.6, 31.0)	66.3 (66.1, 66.5)	25.7 (25.5, 25.9)

Discussion

To the best of our knowledge, this analysis represents the first nationwide estimation of basic hygiene prevalence in LMICs using machine learning predictive models, and represents a pioneering work examining

 DOI: https://doi.org/10.1016/j.apgeog.2021.102549

geospatial disparities in access to basic hygiene services. In the context of the SDG targets for universal access to basic services for all by 2030 (WHO & UNICEF, 2018), our results reveal substantial disparities in access to basic hygiene services across geographic locations. Notably, high levels of access to basic hygiene services often benefit people living in affluent areas, whereas the vast majority who lack access to basic hygiene live in economically disadvantaged communities. For example, in Nigeria, basic hygiene is more prevalent in coastal areas in the south, and less so in most of the north, home to 87% of the poor (World Bank, 2020a). The substantial local disparities in coverage observed in this study suggest that estimates at national and provincial level are insufficient for monitoring progress towards universal access. While aggregated estimates often mask small pockets of low coverage, our modelled surfaces at the 5km-grid level and the adopted inequality metrics offer tools for a better understanding of the underlying local disparities in hygiene access hidden by national and regional averages.

Machine learning models such as RF have been adopted in various applications concerning spatial distributions. Such applications, either concerning environmental suitability or susceptibility across an area or looking for geographic weights for dasymetric population redistribution (Stevens et al., 2015), often ignore spatial autocorrelation (Hengl et al., 2018). Hence, studies looking at the prevalence of demographic and health-related indicators often draw on Bayesian geostatistical methodologies (Mayala et al., 2019). In this study, we employed a model stacking technique with several machine learning algorithms to predict basic hygiene prevalence, where an IDW interpolation was employed as an additional predictive covariate to account for spatial autocorrelation. The adoption of this IDW estimator incorporated geographical proximity effects into the model to ensure optimal prediction, and offered simplicity and a significant speed boost in computation. Both the performance evaluation metrics and the consistency between our predictions and DHS-reported figures suggest our resultant output is plausible. For Malawi, the comparatively little geographic variation in basic hygiene coverage with low prevalence in all areas may be the main reason for the relatively low model performance. The adoption of the model stacking technique was shown to improve model performance significantly, as reflected in the gain in performance evaluation metrics between levels (**Table S.3** in **Supplementary Material 1**).

As an additional benefit of the predictive machine learning technique, our ensemble model gained insights into landscape-level characteristics associated with basic hygiene prevalence. In this five country study, literacy had a high covariate contribution to basic hygiene prevalence in all countries except Uganda. This finding may suggest a link between educational attainment and demand for handwashing facilities and materials. Education has previously been shown to associate positively with water and soap presence at handwashing facilities (Loughnan et al., 2015) as well as handwashing behaviour (White, Thorseth, Dreibelbis, & Curtis, 2020). In this regard, reducing the educational attainment gap may help widen access to basic hygiene services in LMICs. In addition, household wealth status is known to relate to handwashing behaviour and facility status in some domestic settings (Loughnan et al., 2015; White et al., 2020). This is reflected in the high importance of covariates in modelling basic hygiene prevalence, including child growth failure for Malawi, Nepal, Nigeria and Uganda; stable night-time lights for Pakistan and Uganda. In contrast, despite the known association between water source availability and handwashing (White et al., 2020), access to an improved water surprisingly provided little useful information for modelling basic hygiene services in all countries except Malawi. This may be caused by the inclusion of improved water sources that are non-piped, off premises or perceived as poor quality within this metric, since the proximity and perception of water sources also affect handwashing behaviour (Luby et al., 2009; White et al., 2020) and in turn may affect presence of handwashing facilities and materials. Similarly, while handwashing behaviour can often be linked to toileting and potential faecal contact (White et al., 2020; Wolf et al., 2019), lack of sanitation was only found among the most important covariates in Pakistan. This may also be because the sanitation indicator did not distinguish those improved sanitation types that have stronger associations with handwashing using water and soap (White et al., 2020).

Beyond local coverage gaps, a variety of inequality metrics have been employed in measuring geographic inequalities in water and sanitation access (Cetrulo, Marques, Malheiros, & Cetrulo, 2020), among which many can be decomposed at a regional level (Chaudhuri & Roy, 2017; He et al., 2018; Pullan, Freeman, Gething, & Brooker, 2014; Yu et al., 2014). Drawing on the decomposition of a dissimilarity index (Yu et al., 2014), this study examined the degree of segregation of the two groups of population with and without access to basic hygiene across geographic-sub-divisions. This dissimilarity index, through its calculation, offers simplicity and takes into consideration population size. It therefore provides additional information on the effect of population size hidden by local coverage. For example, in Malawi, a very high level of inequality in access to basic hygiene services in Lilongwe City can be found in the local dissimilarity index map (**Figure 6**), given its large population size as the capital city. In contrast, this contribution to overall inequality is masked in the local coverage map (**Figure 5**). This feature of the local dissimilarity index could provide a foundation for identifying population at risk, quantifying burden and guiding resource deployment by locating areas with high population density and low levels of access.

This study is subject to several limitations as follows: firstly, the precise coordinates for the DHS clusters were not available in order to protect respondents' confidentiality (Burgert, Zachary, et al., 2013; Perez-Heydrich et al., 2013). The displacement of cluster locations restricts DHS spatial precision, and thus undermines the utility of the output in estimation at very fine spatial scales. Secondly, since handwashing has only recently been measured through household survey campaigns, and since currently existing geospatial data sources on certain hygiene-related factors are limited, our study was based on a cross-sectional design which cannot examine development over time. Thirdly, our model may inherit the limitations and uncertainties of the input data and methods adopted in this study. This includes, for example, the inaccuracy caused by the 1% of further-displaced (up to 10km) rural cluster points in the model input; potential bias caused by the exclusion of samples in politically unstable areas from the DHS campaign (National Population Commission (NPC) [Nigeria] & ICF, 2019); underestimation of basic hygiene presence caused by handling of missing data; misleading information captured in stable night-time lights due to petroleum industry in certain areas such as the Niger Delta; potential inconsistency in urban-rural classification between our referenced data and the DHS-adopted definitions; potential effect of "no permission to see" in the DHS-reported figures; uncertainty in the geospatial data products used as predictive covariates; distortion caused by data pre-processing; and any drawbacks in the adopted modelling algorithms. In addition, the scale-dependency of the dissimilarity index undermines its utility in national comparisons, as the population size of subnational divisions varies between countries. Furthermore, this study was cross-sectional in design, which thus precludes causal inferences (Kesmodel, 2018). Lastly, subject to context, the predicted prevalence of basic hygiene should be interpreted with caution in studying handwashing behaviour, given the low rate of actual post-faecal-contact handwashing practice in households with access to basic hygiene facilities in low- and middle-income countries found in previous studies (Prüss-Ustün et al., 2019; Wolf et al., 2019).

In this study, the machine learning algorithm RF was employed to generalise the final prediction in the stacked modelling framework. However, many existing efforts to map demographic and health-related indicators adopted a Bayesian geostatistical model using an ensemble approach, whereby the estimates reflect both geospatial and temporal dimensions. Subject to sufficient data being available, a future study could implement such an approach to produce estimates of basic hygiene coverage with extended temporal coverage. Such a study could systematically investigate the strengths and limitations of machine learning models in comparison with Bayesian geostatistical models. Additionally, future studies could explore the geospatial distribution of more detailed aspects of handwashing services (e.g. households with a facility lacking water/soap, fixed versus mobile handwashing stations, etc.). Furthermore, there would be scope to conduct a similar study examining predictors of basic hygiene services in other settings, including schools and healthcare facilities, which are priorities for SDG monitoring and infectious diseases prevention. In the context of the currently ongoing COVID-19 pandemic, there would also be scope to expand the analysis to other countries and to examine spatial patterns for hygiene in conjunction with other priority indicators.

Conclusion

In view of geographical disparities in basic hygiene services, meeting the SDG target for universal access for improved public health requires monitoring at geospatially explicit scales. This study produced estimates of access to basic hygiene services at the 5km-grid scale for five low- and middle-income study countries using an ensemble model, reflecting the capability of machine learning and the value of existing geospatial datasets in predicting the prevalence of basic hygiene services. The methodology provided insights into geospatial patterns of basic hygiene services and their association with landscape-level characteristics. Both educational attainment and wealth status were found to be important in explaining the prevalence of basic hygiene services. By triangulating with subnational administrative data, local coverage and inequality metrics were calculated to reveal apparent disparities in access to basic hygiene services, particularly highlighting areas with large populations. Such outputs can be used as alternative or supplementary information alongside the aggregated estimates. With extended geographic and temporal coverages in the future, they could become important tools to support planning of efficient and precise deployment to scale up access to hand hygiene facilities with water and soap and shift social and cultural norms on handwashing, and ultimately achieve universal access to basic hygiene and improved public health for all.

Role of the funding source

The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

CRediT Author Statement

Conceptualization: WY, RESB, JY, VA, WDG, YL and JAW; Methodology: WY, RESB, JY, VA, WDG, YL and JAW; Software: WY; Validation: WY; Formal analysis: WY; Investigation: WY; Resources: WY, RESB, JY, YL and JAW; Data Curation: WY, YL and JAW; Writing – Original Draft: WY, RESB, VA, and JAW; Writing – Review & Editing: RESB, JY, VA, WDG, YL and JAW; Visualization: WY, RESB and JAW; Supervision: RESB, VA, YL and JAW; Project Administration: WY, VA, YL and JAW; Funding Acquisition: WY, VA, YL and JAW

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant 41771449. Victor Alegana is funded as a Wellcome Trust Training Fellow (number 211208).

References

Amodio, S., & Meulman, J. (2011). Generalized boosted additive models. University of Naples Federico II.

Brauer, M., Zhao, J. T., Bennitt, F. B., & Stanaway, J. D. (2020). Global Access to Handwashing: Implications for COVID-19 Control in Low-Income Countries. *Environmental Health Perspectives*, *128*(5), 1–6. https://doi.org/https://doi.org/10.1289/EHP7200

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. CRC.

Burgert, C. R., Colston, J., Roy, T., & Zachary, B. (2013). *Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys. DHS Spatial Analysis Reports No. 7*.

Calverton, Maryland, USA. Retrieved from http://dhsprogram.com/pubs/pdf/SAR7/SAR7.pdf

Burgert, C. R., Zachary, B., & Colston, J. (2013). *Incorporating geographic information into demographic and health surveys: A field guide to GPS data collection*. Calverton, Maryland USA.

Cetrulo, T. B., Marques, R. C., Malheiros, T. F., & Cetrulo, N. M. (2020). Monitoring inequality in water access: Challenges for the 2030 Agenda for Sustainable Development. *Science of the Total Environment*.

Chaudhuri, S., & Roy, M. (2017). Rural-urban spatial inequality in water and sanitation facilities in India: A cross-sectional study from household to national level. *Applied Geography*, *85*, 27–38. https://doi.org/10.1016/j.apgeog.2017.05.003

Clarke, B. S. (2003). Comparing Bayes model averaging and stacking when model approximation error cannot be ignored. *The Journal of Machine Learning Research*, *4*, 683–712. https://doi.org/https://doi.org/10.1162/153244304773936090

Croft, T. N., Marshall, A. M. J., Allen, C. K., & et al. (2018). *Guide to DHS Statistics*. Rockville, Maryland, USA.

Curtis, V., Schmidt, W., Luby, S., Florez, R., Touré, O., & Biran, A. (2011). Hygiene: new hopes, new horizons. *The Lancet Infectious Diseases*, *11*(4), 312–321. https://doi.org/https://doi.org/10.1016/S1473-3099(10)70224-3

Diggle, P. J., Tawn, J. A., & Moyeed, R. A. (1998). Model−based geostatistics. *Journal of the Royal Statistical Society Series C – Applied Statistics*, *47*, 299–326.

Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, *20*(2), 210–217.

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

Ejemot, R. I., Ehiri, J. E., Meremikwu, M. M., & Critchley, J. A. (2008). Hand washing for preventing diarrhoea. *Cochrane Database of Systematic Reviews*, *23*(1). https://doi.org/10.1002/14651858.CD004265.pub2

Ekumah, B., Armah, F. A., Yawson, D. O., Quansah, R., Nyieku, F. E., Owusu, S. A., … Afitiri, A.-R. (2020). Disparate on-site access to water, sanitation, and food storage heighten the risk of COVID-19 spread in Sub-Saharan Africa. *Environmental Research*, *189*. https://doi.org/https://doi.org/10.1016/j.envres.2020.109936

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, *35*(8), 1652–1660. https://doi.org/https://doi.org/10.1016/j.cageo.2009.01.009

Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F. C., & Taneja, J. (2021). Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing*, *13*(5), 922. https://doi.org/10.3390/rs13050922

Fewtrell, L., Kaufmann, R. B., Kay, D., Enanoria, W., Haller, L., & Colford Jr, J. M. (2005). Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. *Lancet Infectious Diseases*, *5*(1), 42–52.

Florczyk, A. J., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., … Zanchetta, L. (2019). *GHSL Data Package 2019*. Luxembourg. https://doi.org/10.2760/290498

Friedman, J. H. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics*, *19*(1), 1–67. https://doi.org/10.1214/09-AOAS284

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

Gething, P., Tatem, A., Bird, T., & Burgert-Brucker, C. R. (2015). *Creating spatial interpolation surfaces with DHS Data, DHS Spatial Analysis Reports No. 11*. Rockville, Maryland, USA. Retrieved from

http://dhsprogram.com/publications/publication-SAR11-Spatial-Analysis-Reports.cfm
He, W.-J., Lai, Y.-S., Karmacharya, B. M., Dai, B.-F., Hao, Y.-T., & Xu, D. R. (2018). Geographical heterogeneity and inequality of access to improved drinking water supply and sanitation in Nepal. *International Journal for Equity in Health*, *17*(40), 1–14.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518. https://doi.org/10.7717/peerj.5518

Hermes GIS team. (2020). Nepal Administrative Boundary (WGS 1984). Retrieved June 18, 2020, from https://download.hermes.com.np/nepal-administrative-boundary-wgs/

Hijmans, R., Garcia, N., & Wieczorek, J. (2018). GADM: database of global administrative areas. Version 3.6.

Huang, D. B., & Zhou, J. (2007). Effect of intensive handwashing in the prevention of diarrhoeal illness among patients with AIDS: a randomized controlled study. *Journal of Medical Microbiology*, *56*(5), 659–663. https://doi.org/https://doi.org/10.1099/jmm.0.46867-0

ICF. (2018). Demographic and Health Surveys (various) [Datasets]. Funded by USAID. Rockville, Maryland, USA: ICF.

ICF. (2019). *Demographic and Health Survey Interviewer's Manual*. Rockville, Maryland, USA.

ICF International. (2012). *Demographic and Health Survey Sampling and Household Listing Manual*. Calverton, Maryland, U.S.A.

Jefferson, T., Del Mar, C. B., Dooley, L., Ferroni, E., Al-Ansary, L. A., Bawazeer, G. A., … Conly, J. M. (2011). Physical interventions to interrupt or reduce the spread of respiratory viruses (Review). *Cochrane Database of Systematic Reviews*, (7). https://doi.org/10.1002/14651858.CD006207.pub4

Jiwani, S. S., & Antiporta, D. A. (2020). Inequalities in access to water and soap matter for the COVID-19 response in sub-Saharan Africa. *International Journal for Equity in Health*, *19*(82). https://doi.org/https://doi.org/10.1186/s12939-020-01199-z

Karagiannis-Voules, D.-A., Scholte, R. G. C., Guimarães, L. H., Utzinger, J., & Vounatsou, P. (2013). Bayesian Geostatistical Modeling of Leishmaniasis Incidence in Brazil. *PLoS Neglected Tropical Diseases*, *7*(5), e2213.

Kesmodel, U. S. (2018). Cross-sectional studies – what are they good for ? *Acta Obstetricia et Gynecologica Scandinavica*, *97*, 388–393. https://doi.org/10.1111/aogs.13331

Lai, Y.-S., Zhou, X.-N., Utzinger, J., & Vounatsou, P. (2013). Bayesian geostatistical modelling of soil-transmitted helminth survey data in the People's Republic of China. *Parasites & Vectors*, *6*(359). https://doi.org/10.1186/1756-3305-6-359

Lindgren, F., & Rue, H. (2015). Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, *63*(19). https://doi.org/10.18637/jss.v063.i19

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B*, *73*(4), 423–498. https://doi.org/https://doi.org/10.1111/j.1467-9868.2011.00777.x

Local Burden of Disease WaSH Collaborators. (2020). Mapping geographical inequalities in access to drinking water and sanitation facilities in low-income and middle-income countries, 2000–17. *Lancet Global Health*, *8*, e1162-85. https://doi.org/10.1016/S2214-109X(20)30278-3

Loughnan, L. C., Ram, P. K., & Luyendijk, R. (2015). Measurement of handwashing behaviour in Multiple Indicator Cluster Surveys and Demographic and Health Surveys, 1985–2008. *Waterlines*, *34*(4), 296–313.

Luby, S. P., Agboatwalla, M., Feikin, D. R., Painter, J., Billhimer, W., Altaf, A., & Hoekstra, R. M. (2005). Effect of Handwashing on Child Health: A Randomised Controlled Trial. *Lancet*, *366*(9481), 225–233.

https://doi.org/10.1016/S0140-6736(05)66912-7
Luby, S. P., & Halder, A. K. (2008). Associations Among Handwashing Indicators, Wealth, and Symptoms of Childhood Respiratory Illness in Urban Bangladesh. *Tropical Medicine & International Health*, *13*(6), 835–844. https://doi.org/10.1111/j.1365-3156.2008.02074.x
Luby, S. P., Halder, A. K., Tronchet, C., Akhter, S., Bhuiya, A., & Johnston, R. B. (2009). Household Characteristics Associated with Handwashing with Soap in Rural Bangladesh. *The American Journal of Tropical Medicine and Hygiene*, *81*(5), 882–887. https://doi.org/10.4269/ajtmh.2009.09-0031
Mayala, B. K., Dontamsetti, T., Fish, T. D., & Croft, T. N. (2019). *Interpolation of DHS Survey Data at Subnational Administrative Level 2. DHS Spatial Analysis Reports No. 17*. Rockville, Maryland, USA.
Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x
Ministry of Health [Nepal], New ERA, & ICF. (2017). *Nepal Demographic and Health Survey 2016*. Kathmandu, Nepal.
Mosser, J. F., Gagne-Maynard, W., Rao, P. C., Osgood-Zimmerman, A., Fullman, N., Graetz, N., … Hay, S. I. (2019). Mapping Diphtheria-pertussis-tetanus Vaccine Coverage in Africa, 2000–2016. *The Lancet*, *393*(10183), 1843–1855. https://doi.org/https://doi.org/10.1016/S0140-6736(19)30226-0
National Institute of Population Studies (NIPS) [Pakistan], & ICF. (2019). *Pakistan Demographic and Health Survey 2017-18*. Islamabad, Pakistan, and Rockville, Maryland, USA.
National Population Commission (NPC) [Nigeria], & ICF. (2019). *Nigeria Demographic and Health Survey 2018*. Abuja, Nigeria, and Rockville, Maryland, USA.
National Statistical Office (NSO) [Malawi], & ICF. (2017). *Malawi Demographic and Health Survey 2015-16*. Zomba, Malawi, and Rockville, Maryland, USA.
Nguyen, T. (2015). Hand washing with ash and mud, an accepted practice in Malawi: findings from a knowledge, attitudes and practice study. *Tropical Medicine & International Health*, *20*(360).
Osgood-Zimmerman, A., Millear, A. I., Stubbs, R. W., Shields, C., Pickering, B. V., Earl, L., … Hay, S. I. (2018). Mapping Child Growth Failure in Africa between 2000 and 2015. *Nature*, *555*(7684), 41–47. https://doi.org/https://doi.org/10.1038/nature25760
Paludan-Müller, A. S., Boesen, K., Klerings, I., Jørgensen, K. J., & Munkholm, K. (2020). Hand cleaning with ash for reducing the spread of viral and bacterial infections: a rapid review (Review). *Cochrane Database of Systematic Reviews*, (4). https://doi.org/10.1002/14651858.CD013597
Perez-Heydrich, C., Warren, J. L., Burgert, C. R., & Emch, M. E. (2013). *Guidelines on the Use of DHS GPS Data. Spatial Analysis Reports No. 8*. Calverton, Maryland, USA.
Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., & Maffenini, L. (2019). GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015), R2019A. European Commission, Joint Research Centre (JRC). https://doi.org/10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218
Philip, G. M., & Watson, D. F. (1982). A Precise Method for Determining Contoured Surfaces. *Australian Petroleum Exploration Association Journal*, *22*, 205–212.
Pinkovskiy, M., & Sala-i-Martin, A. (2016). Lights, Camera … Income! Illuminating the National Accounts-Household Surveys Debate. *The Quarterly Journal of Economics*, *131*(2), 579–631. https://doi.org/https://doi.org/10.1093/qje/qjw003
Prüss-Ustün, A., Wolf, J., Bartram, J., Clasen, T., Cumming, O., Freeman, M. C., … Johnston, R. (2019). Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. *International Journal of Hygiene*

and Environmental Health, *222*(5), 765–777. https://doi.org/10.1016/j.ijheh.2019.05.004

Pullan, R. L., Freeman, M. C., Gething, P. W., & Brooker, S. J. (2014). Geographical inequalities in use of improved drinking water supply and sanitation across sub-Saharan Africa: Mapping and spatial analysis of cross-sectional survey data. *PLoS Medicine*, *11*(4). https://doi.org/10.1371/journal.pmed.1001626

Rah, J. H., Cronin, A. A., Badgaiyan, B., Aguayo, V. M., Coates, S., & Ahmed, S. (2015). Household sanitation and personal hygiene practices are associated with child stunting in rural India: a cross-sectional analysis of surveys. *BMJ Open*, *5*(2), e005180. https://doi.org/10.1136/bmjopen-2014-005180

Ram, P. (2013). *Practical Guidance for Measuring Handwashing Behavior: 2013 Update* (Water and Sanitation Program Working Paper). New York, NY, USA.

Reiner, R. C., Graetz, N., Casey, D. C., Troeger, C., Garcia, G. M., Mosser, J. F., … Hay, S. I. (2018). Variation in Childhood Diarrheal Morbidity and Mortality in Africa, 2000–2015. *New England Journal of Medicine*, *379*(12), 1128–1138. https://doi.org/10.1056/NEJMoa1716766

Renzaho, A. M. N. (2020). The Need for the Right Socio-Economic and Cultural Fit in the COVID-19 Response in Sub-Saharan Africa : Examining Demographic , Economic Political , Health , and Socio-Cultural Di ff erentials in COVID-19 Morbidity and Mortality. *International Journal of Environmental Research and Public Health*, *17*(10), 3445. https://doi.org/https://doi.org/10.3390/ijerph17103445

RStudio Team. (2016). RStudio: Integrated development for R. Boston: RStudio Inc.

Rue, H., Martino, S., & Chopin, N. (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.

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. *PLoS ONE*, *10*(2), 1–22. https://doi.org/10.1371/journal.pone.0107042

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, *4*(170004). https://doi.org/https://doi.org/10.1038/sdata.2017.4

The R Core Team. (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Uganda Bureau of Statistics (UBOS), & ICF. (2018). *Uganda Demographic and Health Survey 2016*. Kampala, Uganda and Rockville, Maryland, USA.

UNICEF, & WHO. (2020). *Hand hygiene for all*. Retrieved from https://www.unicef.org/reports/hand-hygiene-for-all-2020

Warren-gash, C., Fragaszy, E., & Hayward, A. C. (2012). Hand hygiene to reduce community transmission of influenza and acute respiratory tract infection: a systematic review. *Influenza and Other Respiratory Viruses*, *7*(5), 738–749. https://doi.org/10.1111/irv.12015

White, S., Thorseth, A. H., Dreibelbis, R., & Curtis, V. (2020). The determinants of handwashing behaviour in domestic settings: An integrative systematic review. *International Journal of Hygiene and Environmental Health*, *227*, 113512. https://doi.org/10.1016/j.ijheh.2020.113512

WHO, & UNICEF. (2018). *Core questions on water, sanitation and hygiene for household surveys - 2018 update*. New York.

WHO, & UNICEF. (2020). *Water, sanitation, hygiene and waste management for the COVID-19 virus*.

Wolf, J., Johnston, R., Freeman, M. C., Ram, P. K., Slaymaker, T., Laurenz, E., & Prüss-Ustün, A. (2019). Handwashing with soap after potential faecal contact: global, regional and country estimates. *Infectious Disease*, *48*(4), 1204–1218. https://doi.org/10.1093/ije/dyy253

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. https://doi.org/https://doi.org/10.1016/S0893-6080(05)80023-1

World Bank. (2020a). *Advancing Social Protection in a Dynamic Nigeria (English)*. Washington, D.C., U.S.A. Retrieved from http://documents.worldbank.org/curated/en/612461580272758131/Advancing-Social-Protection-in-a-Dynamic-Nigeria

World Bank. (2020b). World Bank Country and Lending Groups. Retrieved June 8, 2020, from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

World Health Organization. (2009). *Pandemic Influenza Preparedness and Response: A WHO Guidance Document*. Geneva, Switzerland.

World Health Organization. (2020). *Recommendations to Member States to improve hand hygiene practices to help prevent the transmission of the COVID-19 virus*.

WorldPop, & Center for International Earth Science Information Network (CIESIN) Columbia University. (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://doi.org/10.5258/SOTON/WP00671

Yu, W., Bain, R., Mansour, S., & Wright, J. A. (2014). A cross-sectional ecological study of spatial scale and geographic inequality in access to drinking-water and sanitation. *International Journal for Equity in Health*, *13*(1), 113. https://doi.org/10.1186/s12939-014-0113-3

Yu, W., Wardrop, N. A., Bain, R. E. S., Alegana, V., Graham, L. J., & Wright, J. A. (2019). Mapping access to domestic water supplies from incomplete data in developing countries : An illustrative assessment for Kenya. *PLoS ONE*, *14*(5), e0216923. https://doi.org/https://doi. org/10.1371/journal.pone.0216923
Editor: