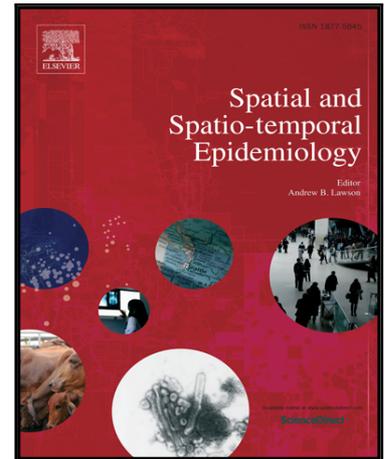# Accepted Manuscript

Modelling and Forecasting Spatio-temporal Variation in the Risk of Chronic Malnutrition Among Under-Five Children in Ghana

Justice Moses K. Aheto, Benjamin M. Taylor, Peter J. Diggle, Thomas J. Keegan

Highlights

- This is the first study to explore spatio-temporal patterns in < 5 stunting

- We develop a spatio-temporal model to capture variation in childhood stunting

- Our method facilitates and enriches modelling and forecasting of future stunting

- Substantial spatio-temporal variation exists in the risk of under 5 child stunting

- Our method allows identification of higher risk communities for policy decisions

1

Modelling and Forecasting Spatio-temporal Variation in the Risk of Chronic Malnutrition Among Under-Five Children in Ghana

Justice Moses K. Aheto*, Benjamin M. Taylor, Peter J. Diggle, Thomas J. Keegan

Lancaster Medical School,Faculty of Health and Medicine, Lancaster University, Bailrigg, Lancaster, LA1 4YB, United Kingdom.

Abstract

Background: Spatio-temporal variation in under-five-year-old children malnutrition remains unstudied in most developing countries like Ghana. This study explores and forecasts the spatio-temporal patterns in childhood chronic malnutrition among these children. We also investigate the effect of maternal education on childhood malnutrition.

Methods: We analysed data on 10,036 children residing in 1,516 geographic locations. A spatio-temporal model was fitted to the data and was used to produce predictive maps of spatio-temporal variation in the probability of stunting.

Results: The study found substantial spatio-temporal variation in the prevalence of stunting. Also, higher levels of mother's education were associated with decreased risk of being stunted.

Conclusion: Our spatio-temporal model captured variations in childhood stunting over place and time. Our method facilitates and enriches modelling and forecasting of future stunting prevalence to identify areas at high risk. Improving maternal education could be given greater consideration within an overall strategy for addressing childhood malnutrition.

Keywords: Spatio-Temporal modelling, Public health, Childhood

Email addresses: `j.aheto@lancaster.ac.uk` (Justice Moses K. Aheto*), `b.taylor1@lancaster.ac.uk` (Benjamin M. Taylor), `p.diggle@lancaster.ac.uk` (Peter J. Diggle), `t.keegan@lancaster.ac.uk` (Thomas J. Keegan)

malnutrition.

1. Introduction

Childhood malnutrition is associated with poor mental and physical development, and with increased risk of morbidity and mortality.[1, 2, 3, 4] It has been established that morbidity and mortality increase exponentially with deteriorating nutritional status.[5] Childhood malnutrition also prevents children from reaching their full growth potential.[2, 4, 6] Childhood malnutrition is also associated with economic and health complications in adult life and constitutes a significant public health challenge globally, especially in developing countries.[2, 4, 6, 7, 8]

Spatial and temporal variations in the effects on growth of different levels of childhood nutrition can be attributable to socioeconomic, environmental and policy-related factors. Spatio-temporal analysis of under-fives malnutrition could therefore help to improve our understanding of the problem and to identify better and more targeted nutritional interventions, which is crucial to public health policy given the limited resources available in developing countries. Spatio-temporal variation in malnutrition amongst under-five-year-old children remains largely unstudied in most developing countries such as Ghana.

Studies into health outcomes such as malnutrition, diarrhoea and diarrheal mortality using space-time analysis [9], spatial analysis [10, 11] and exploratory analysis [12] has contributed to immense improvement in our understanding of such health outcomes.

The need to examine spatial and temporal patterns in the risk of childhood chronic malnutrition in this study has been motivated by previous studies that observed that certain health outcomes exhibit spatial, temporal, spatiotemporal and seasonal patterns and that such findings are crucial to public health policy and to designing effective public health intervention, treatment and prevention strategies.[13, 14, 10, 12, 15, 16, 9, 11] Specifically, a study conducted in northwest Ethiopia to examine spatiotemporal clustering and seasonal differences in

3

childhood diarrhoea reported substantial spatial, temporal, and spatiotemporal clustering.[14] A study conducted in Mexico to investigate spatiotemporal patterns of diarrhoeal mortality also observed spatiotemporal trends [13]. A study conducted in Somalia to examine wasting among children aged below 5 years reported substantial spatial and temporal variability in wasting.[9] The present study is the first to provide forecast maps for the risk of chronic malnutrition.

In a previous study using the 2008 Ghana Demographic and Health Survey (GDHS), we examined spatial variation in the risk of childhood malnutrition in Ghana, and identified higher risk communities in the country. Results from that study showed substantial spatial variation in the risk of childhood malnutrition in the country (Aheto, et al. 2015, Unpublished). The present study explores the spatio-temporal patterns in the risk of childhood malnutrition and provides a forecast for the risk of chronic malnutrition in Ghana.

Globally, malnutrition accounts for about 30% of under-five-year-old mortality [1] and this percentage is higher in developing countries, at over 50%.[17] Childhood malnutrition is the cause of about 40% of deaths among children under the age of 5 years in Ghana [8, 18, 19, 20] and remains a major public health challenge in the country; the percentages of children aged less than 5 years who were stunted were 33% in 1993, 31% in 1998, 35% in 2003 and 28% in 2008.[21]

The prevalence of malnutrition measured through nutrition surveys such as the Ghana Demographic and Health Survey is characterized by both short- and long-term variability. The prevalence need to be interpreted in relation to adverse climatic conditions, emergency factors and seasonality, all of which can vary geographically. Thus, this information will be useful for assessing the severity of malnutrition at one time and place and early prediction of the rates at which malnutrition prevalences are likely to rise or decrease.[9]

In this paper we investigate the spatio-temporal patterns in the risk of childhood chronic malnutrition measured on height-for-age Z-scores (HAZ, a measure of stunting) among children aged less than 5 years while taking into account the cross-sectional nature of each survey and adjusting for potential risk factors.

4

60 Thus, we want to predict HAZ both in space and time. We produce forecast maps for spatio-temporal variations in stunting over Ghana and identify areas at highest risk of stunting. We also evaluate the impact on the prevalence of malnutrition that would take place if we could change each child's covariates in certain ways. Specifically, we consider a policy change with regards to ma-

65 ternal education. In 2008, 55% of mothers had less than 6 years of education. To test the effect of increasing the level of maternal education we forecast what would have happened had each mother with less than 6 years education received exactly 6 years of education.

We emphasise the importance of benefits to maternal and child health of

70 increased in women's education. In our previous study using the 2008 GDHS dataset, we observed that an increase in the level of maternal education was associated with decreased risk of child malnutrition.[20] Also, previous studies have established an association between maternal education and child health outcomes. Specifically, maternal education is linked to knowledge of good

75 practices for the mother and her child (e.g. feeding and health seeking behaviour) but is also associated with household socioeconomic status and access to food.[20, 8, 22, 23].

2. Methods

2.1. Study population and design

80 This study uses four data-sets from the quinquennial Ghana Demographic and Health Surveys (GDHS), conducted in 1993, 1998, 2003 and 2008.[21, 24, 25, 26] These data-sets included the same variables in each survey iteration, used the same survey methodology and were conducted by the same organization. The data were provided by the Ghana Statistical Service and DHS Program. They

85 contain information on the health of women and their children under 5 years of age, their geographic location and anthropometric measurements (weight and height) at the time of the survey, breastfeeding and infant feeding practices, childhood illness and mortality, domestic violence, fertility, awareness and use

5

of family planning. Anthropometric measurements of female respondents of

[90] reproductive age (15-49 years) and children under 5 years of age were collected from only those households selected for the individual interviews.

The GDHS used a two-stage cluster sampling design for each of the four surveys. The clusters were enumeration areas or sample points. The 1984 Ghana Population Census provided the master sampling frame from which the clusters

[95] were selected for the 1993 and 1998 GDHSs, whilst the 2000 Ghana Population and Housing Census (GPHC) provided the master sampling frame for the selection of the clusters for the 2003 and 2008 GDHSs. The clusters were selected with probability proportional to the number of households listed in each cluster. The selection of the clusters was followed by a complete listing of all households

[100] in the selected clusters, which provided the sampling frame for the selection of households in the second stage. Different samples and sample sizes were used in each of the four surveys, i.e. a serial cross-sectional rather than a longitudinal design. The data therefore have a hierarchical structure, with households nested within clusters and children nested within sampled households. There were a

[105] total of 2,204, 3,298, 3,844 and 2,992 children aged below 5 years in 1993, 1998, 2003 and 2008 GDHS datasets respectively. In our final analysis, we analysed data on 1,922 (87.2%) in the 1993 GDHS, 2,736 (83.0%) in the 1998 GDHS, 3,074 (80.0%) in the 2003 GDHS, and 2,304 (77.0%) in the 2008 GDHS.

2.2. Geographical data

[110] The 1993, 1998, 2003 and 2008 GDHSs identify the geographical locations of households with the centroid of the cluster from which they were selected. This resulted in a total of 319, 395, 402 and 400 distinct geographical coordinates for 1993, 1998, 2003 and 2008 GDHS data-sets, respectively. Data were not collected from one of the clusters sampled in the 2008 survey. Geographic

[115] coordinates were not recorded for 81, 10, 5 and 11 clusters in the 1993, 1998, 2003 and 2008 surveys, respectively, and these clusters were therefore excluded from our analysis. Out of the combined number of 1,623 clusters used across the surveys, 399 featured in all the 4 surveys, 1 featured in three out of four, 12

6

on two out of four. We used the 1,516 unique centroid cluster locations in our
spatio-temporal analysis.

2.3. Outcome variable

The three most commonly used nutritional status measures for under-five
children are weight-for-age (WAZ), weight-for-height (WHZ) and height-for-age
(HAZ) Z-scores We use the HAZ score, which is considered to be the most stable
of the three as it is not easily influenced by temporal variations in food supply
or impact of severe and recurrent diseases, and captures multiple dimensions
of children's development, health and the environment in which they reside.[27,
28, 29]. Amongst these three measures, the HAZ score also gives the highest
national prevalence of malnutrition in each of the survey years.[21, 20]. A child
is classified as stunted or chronically malnourished if their HAZ score is below
-2 standard deviations from the median of the reference population.[30]

2.4. Explanatory variables

Months of breastfeeding, child's age, mother's educational attainment and
household size have previously been identified in the literature to be predictive of
under-five nutritional outcomes in developing countries[20, 31, 22, 32, 33, 34, 35]
and were available in the data from all four surveys. For our data, the backward
elimination method selected these four covariates for inclusion in our final spatio-
temporal model.

2.5. Numbers of eligible participants

In accordance with WHO guidelines, eligible HAZ scores are those in the
range −6 to 6; values lying outside this range were removed as outliers[36].
The numbers of children with incomplete data on height and age, ineligible
HAZ scores or unidentifiable geographical coordinates in the 1993, 1998, 2003
and 2008 data-sets were 282, 562, 770 and 688, respectively. This resulted
in 10,036 children who could be included in our analysis, consisting of 1,922
(87.2%), 2,736 (83.0%), 3,074 (80.0%) and 2,304 (77.0%) in 1993, 1998, 2003 and

7

2008, respectively. Response rates from occupied households were 98.4%, 99.1%, 98.7% and 98.9% for 1993, 1998, 2003 and 2008, respectively [21, 24, 25, 26].

2.6. Statistical analysis

[150] The analysis needs to accommodate: (i) the fact that some of the sampled locations change between successive surveys; (ii) our wish to be able to obtain spatially continuous maps of the probability of malnutrition. In this section we first introduce our statistical model for HAZ, accounting for the changes in sampled locations over the four surveys. We then present the details of our [155] method for producing predictive maps.

2.6.1. Model and Methods

Our dynamic linear model which is a linear state-space model for the HAZ outcome, denoted by Y, is given by the state and observation equations, respectively 1 and 2 below.

$$\theta_t = A\theta_{t-1} + BW_t, \qquad W_t \sim MVN(0, \Omega) \qquad (1)$$

$$Y_t = M_t\theta_t + V_t, \qquad V_t \sim MVN(0, \sigma_v^2 I_{n_t}), \qquad (2)$$

[160] where n_t is the number of children included at time t and I_{n_t} is an $n_t \times n_t$ identity matrix. Also,

$$
\begin{aligned}
\theta_t &= \begin{bmatrix} S_t \\ \beta_t \end{bmatrix}, \\
S_t &= [S_{1,t}, S_{2,t}, \ldots, S_{1516,t}]^T \\
\beta_t &= [\beta_{1,t}, \beta_{2,t}, \ldots, \beta_{5,t}]^T \\
A &= \operatorname{diag}\left[\operatorname{rep}(\alpha, 1516), \operatorname{rep}(1, 5)\right] \\
B &= \operatorname{diag}\left[\operatorname{rep}\left(\sqrt{1-\alpha^2}, 1516\right), \operatorname{rep}(0, 5)\right] \\
\Omega &= \left[\begin{array}{c|c} \Sigma & 0_{1516\times 5} \\ \hline 0_{5\times 1516} & 0_{5\times 5} \end{array}\right] \\
M_t &= [D_t \mid X_t]
\end{aligned}
$$

8

In the above expressions: $\mathrm{rep}\,(x, y)$ is a vector of length y whose elements are all equal to x; $\mathrm{diag}\,[v]$ is a diagonal matrix with the elements of v on the diagonal; $0_{x \times y}$ is a matrix of zeros of dimension x by y; Σ is a matrix of dimension $1516 \times$

165 1516 whose (i, j) element is the covariance between the centroid of the ith cluster and the centroid of the jth cluster, which we model as $\sigma^2 \exp(-d_{ij}/\phi)$, where d_{ij} is the distance between cluster i and cluster j (see below for a justification); the (i, j) element of the $n_t \times 1516$ matrix D_t is 1 if child i at time t was from cluster j and zero otherwise; the $n_t \times 5$ matrix, X_t, is the design matrix for all

170 observations at time t; $\alpha \in [0, 1]$; and Y_t is a vector of observation at time t of length n_t.

This formulation is a linear state-space model [37], whose state vector, θ_t, includes the spatio-temporal random effects S_t and survey-specific covariate effects β_t. Note that Ω is not a valid covariance matrix since it has determinant

175 zero. However, because our initial conditions for the state vector *are* valid, the Kalman filtering recursions can nevertheless be implemented with Ω defined as above. With α, A and B defined as above, the implication for the spatio-temporal process S_t is that (unconditionally) if $S_{t-1} \sim \mathrm{N}(0, \Sigma)$, then $S_t \sim \mathrm{N}(0, \Sigma)$ also, i.e. S_t is stationary in time and has a separable covariance function.

180 Note that the stationarity assumption only applies to the residual process, and that it is a common assumption in regression models.

The four model parameters are σ^2, σ_v^2, ϕ and α. Their interpretations are as follows: σ^2 is the unconditional variance of each $S_{i,t}$; σ_v^2 is the conditional variance of each element of Y_t given the state vector θ_t; ϕ controls the rate

185 at which the correlation between the $S_{i,t}$ at different locations decays with increasing distance between them; α is the autocorrelation between $S_{i,t}$ and $S_{i,t+1}$. We could have included the term $X_t \beta$ into the observation equation (2), but this would have considerably increased the computational cost.

To complete the model, we set the initial conditions of the filter to be Gaus-

190 sian with mean μ_0 and covariance Σ_0. This choice guarantees that the *time t filtering distribution* $[\theta_t | Y_{1:t}]$, where $Y_{1:t}$ denotes the set of all Y_s for $s = 1, 2, ..., t$, is also Gaussian for all t. We chose the initial conditions as follows. For μ_0, we

9

fitted a linear mixed effect model to the 1993 data-set using geographic location as a single, spatially and temporally uncorrelated grouping variable, and set 195 μ_0 as its maximum likelihood estimate under this working model. For Σ_0, we conducted a sensitivity analysis using the choices $\Sigma_0 = \text{diag}\left[\text{rep}\left(u, 1521\right)\right]$ with $u \in \{1, 10, 20, 50, 60\}$. All choices led to substantially the same results; we used $u = 10$ in our main analysis.

We estimated the four parameters in our model, α, ϕ, σ^2, σ_v^2, using maxi- 200 mum likelihood. For the parameters ϕ, σ^2 and σ_v^2, we initialised the maximum likelihood procedure by conducting a variogram analysis of the 1993 data-set. We set the initial value of the autocorrelation parameter, α, to be 0.5, the mid-point of its support. The variogram analysis was also used to choose the functional form for ρ, a choice between the Exponential, Gaussian and Spher- 205 ical correlation functions. We obtained 95% confidence intervals for our fixed effects from the filtering distribution; Table 1 gives β_t for the last time point, conditional on having observed data from all surveys up to and including 2008.

We then used the fitted model to compute the probability that a child at an unsampled location will be chronically malnourished (stunted), i.e. for which 210 HAZ < -2, and constructed predictive maps for these probabilities across Ghana. We also constructed forecast maps of the effect of varying levels of maternal education on the probabilities of chronic malnutrition across Ghana.

All computations were performed using the R package 'miscFuncs', which implements the Kalman Filter and parameter estimation for Gaussian dynamic 215 linear models [38, 39].

2.6.2. Predicting the Risk of Stunting Across the Whole of Ghana

One of our aims is to forecast the probability that in the year of a future survey, here 2013, a child living at an arbitrary location will be stunted. In order to produce maps, we forecast the probability of stunting on a grid of 220 points covering Ghana. This requires us to know or assume the values of all relevant covariates. For forecasting, we propose to draw from the empirical distribution of covariates at the most recent survey time point (i.e. 2008) and use

10

these samples to construct forecast probabilities of stunting across Ghana. This involves two assumptions. The first is that the distribution of the characteristics
₂₂₅ of the population in 2008, namely duration of breastfeeding, mothers' education level, household size and the children's age as measured in the survey will not change substantially in the time we are intending to forecast. The second is that the distribution of each of these population characteristics does not vary spatially over the whole of Ghana; see below for further comment.

₂₃₀ From our model, we can compute the one-step ahead forecast density of $[\theta_{t+1}|Y_{1:t}]$ exactly. Using the mean and variance of this density, we can then predict the distribution of the spatial random effect on the finely spaced grid covering Ghana. This is a multivariate Gaussian distribution with mean $S_G^{(t+1)}$ and variance $V_G^{(t+1)}$. To predict the probability of malnutrition on this grid,
₂₃₅ we again need to know or assume the values of individual-level covariates at each point on the grid. As in the purely spatial case, when individual-level covariates are not available at locations for which a prediction is required it is necessary to use one or more scenarios to represent the values of the covariates. One possible solution would be to take a spatial average of risk factor values
₂₄₀ over a suitable neighbourhood of each grid point, and use these to compute the required probabilities. However, this approach has the following drawbacks:

1. it would not take into account the amount of individual variability in the data;

2. the effect of individual-level covariates is not the same as the effects of
₂₄₅ group-level covariates; ignoring this would lead to ecological bias in the resulting estimates.

We therefore propose to use empirical individual-level covariate data to capture individual-level variability, but with a homogeneous distribution over space.

To obtain the forecast mean and variance of HAZ for 2013 on the grid, we
₂₅₀ used the covariates of the ith child in 2008, $X_{i,t+1}$, to compute the forecast

11

mean and variance for a child with the same covariates over Ghana,

$$
\begin{aligned}
M^{(i)} &= X_{i,t+1}\beta_{t+1} + S_G^{(t+1)} \\
V^{(i)} &= (X_{i,t+1})^T V_\beta^{(t+1)} X_{i,t+1} + V_S^{(t+1)} + \hat{\sigma}_v^2
\end{aligned}
$$

where $V_\beta^{(t+1)}$ and $V_S^{(t+1)}$ are the appropriate sub-matrices of $V_G^{(t+1)}$ and $\hat{\sigma}_v^2$ is estimated nugget effect. Finally, we use equations (3) and (4) to combine these estimates and obtain the mean and variance of the forecast distribution:

$$
\begin{aligned}
\mathbb{E}_Y(Y_{t+1}|\text{data}) &= \mathbb{E}_{X_{t+1}}[\mathbb{E}_{Y_{t+1}|X_{t+1}}(Y_{t+1}|X_{t+1},\text{data})], &(3)\\
\mathbb{V}_Y(Y_{t+1}|\text{data}) &= \mathbb{V}_{X_{t+1}}[\mathbb{E}_{Y_{t+1}|X_{t+1}}(Y_{t+1}|X_{t+1},\text{data})] \\
&\quad + \mathbb{E}_{X_{t+1}}[\mathbb{V}_{Y_{t+1}|X_{t+1}}(Y_{t+1}|X_{t+1},\text{data})]. &(4)
\end{aligned}
$$

255   We then evaluate exceedance probabilities, i.e, $\mathbb{P}(\text{HAZ} < -2)$, using the standard normal cumulative distribution function. In (3) and (4) we use Monte Carlo integration to evaluate the outer integral, selecting covariate values, $X_{i,t+1}$, by sampling from their empirical distribution in 2008.

2.6.3. Exploring the Effect of Policy Changes

260   A second aim is to produce predictive maps of the effect of varying levels of maternal education on the risk of childhood chronic malnutrition. We achieved this by modifying $M^{(i)}$ and $V^{(i)}$ above, potentially changing each child's covariates depending on whether that child's mother received (i) less than 6 years or (ii) less than 12 years education. In case (i), we changed all maternal education

265   covariates to equal 6 years, if the child's mother received less than 6 years of education. In case (ii), we changed all maternal education covariates to equal 12 years, if the child's mother received less than 12 years of education. We again used (3) and (4) to obtain the mean and variance of the policy-modified predictive distributions.

12

3. Results

Out of the 10,036 children with eligible HAZ scores and geographical coordinates 3,306 (33%) were identified as being stunted. Table 1 presents the estimates of the fixed effects in our spatio-temporal model for HAZ score. Months of breastfeeding, child's age, and mother's years of education were statistically significant risk factors for HAZ score, whereas number of household members had no statistically significant association with HAZ score. Longer breastfeeding duration and older ages of children were negatively associated with HAZ score for children, while higher levels of mother's education were positively associated with HAZ scores for children.

Table 1: The effect estimate (β) for the associations between risk factors and child nutritional status (height-for-age Z-score or HAZ score) for spatio-temporal model (n=10,036)

	Parameter Estimates for HAZ (95%CI)
Intercept	-1.288 (-1.421, -1.156)
Months of breastfeeding	-0.445 (-0.485, -0.404)
Child's age	-0.195 (-0.236, -0.154)
Mother's education	0.105 (0.070, 0.140)
Household size	0.008 (-0.024, 0.040)

Table 2 presents the parameter of the fitted spatio-temporal model for HAZ. The range, which measures the rate at which spatial correlation decays to zero with increasing distance, is about 95km. The estimated value of the autocorrelation between successive values of S_{it} at the same location is $\hat{\alpha} = 0.07$. To formally test the effect of the dependency between states, we used a likelihood ratio (deviance) test of the hypothesis that $\alpha = 0$. The deviance statistic was 3.6 on 1 degree of freedom, which is not conventionally significant ($p = 0.0578$). Nevertheless, we included the estimate $\hat{\alpha}$ in our spatio-temporal model to enable prediction of future stunting patterns.

To show how well our model works, we regressed the observed Height-for-Age

13

Table 2: Parameters of the fitted spatio-temporal model for height-for-age Z-score (or HAZ)

.

	Partial Sill $(\sigma^2)^A$	Nugget $(\sigma_v^2)^B$	Range $(\phi)^C$	α^D
Estimates	0.37	1.51	94.85 km	0.07

A = marginal variance associated with the Gaussian spatial processes.

B = variance associated with individual Gaussian residuals.

C = the scale parameter controlling the rate at which correlation decays to zero with increasing distance.

D = dependence (correlation) between states.

z-scores (HAZ) on the predicted HAZ for each year and obtained the intercept, the slope and the R-squared values. The computed R-squared values for the years 1993, 1998, 2003 and 2008 are 0.959, 0.966, 0.930, and 0.909, respectively. We presented detailed results in Table 3 under supplementary material.

Figure 1 presents the maps of the geographical variation in the probability that a child will be stunted in 1993, 1998, 2003 and 2008, using the same grey-scale for all four years. The ranges of the predicted probabilities of stunting in 1993, 1998, 2003 and 2008 are about 0.23 to 0.52, 0.19 to 0.44, 0.37 to 0.65 and 0.17 to 0.37, respectively. The highest stunting probability range of about 0.56 to 0.66 was observed in parts of Northern and Western regions in the year 2003.

At the time of writing, the 2008 Ghana Demographic and Health Survey is the most recent survey data-set available to the public. To illustrate the performance of our fitted spatio-temporal model, we therefore used it to forecast the probabilities of stunting for the year 2013. Figure 2 shows our forecast map, using the same grey-scale as in Figure 1. The predicted stunting probability ranges between 0.36 and 0.65. Stunting probabilities in the range 0.54 to 0.65 were predicted in parts of Northern and Western regions of the country, while probabilities in the range 0.42 to 0.53 were observed in parts of Upper West, Upper East, Ashanti, Volta, Eastern, Brong Ahafo, Central and Greater Accra regions of the country. The geographical distribution in the risk of childhood chronic malnutrition observed in our forecast maps over Ghana is similar to
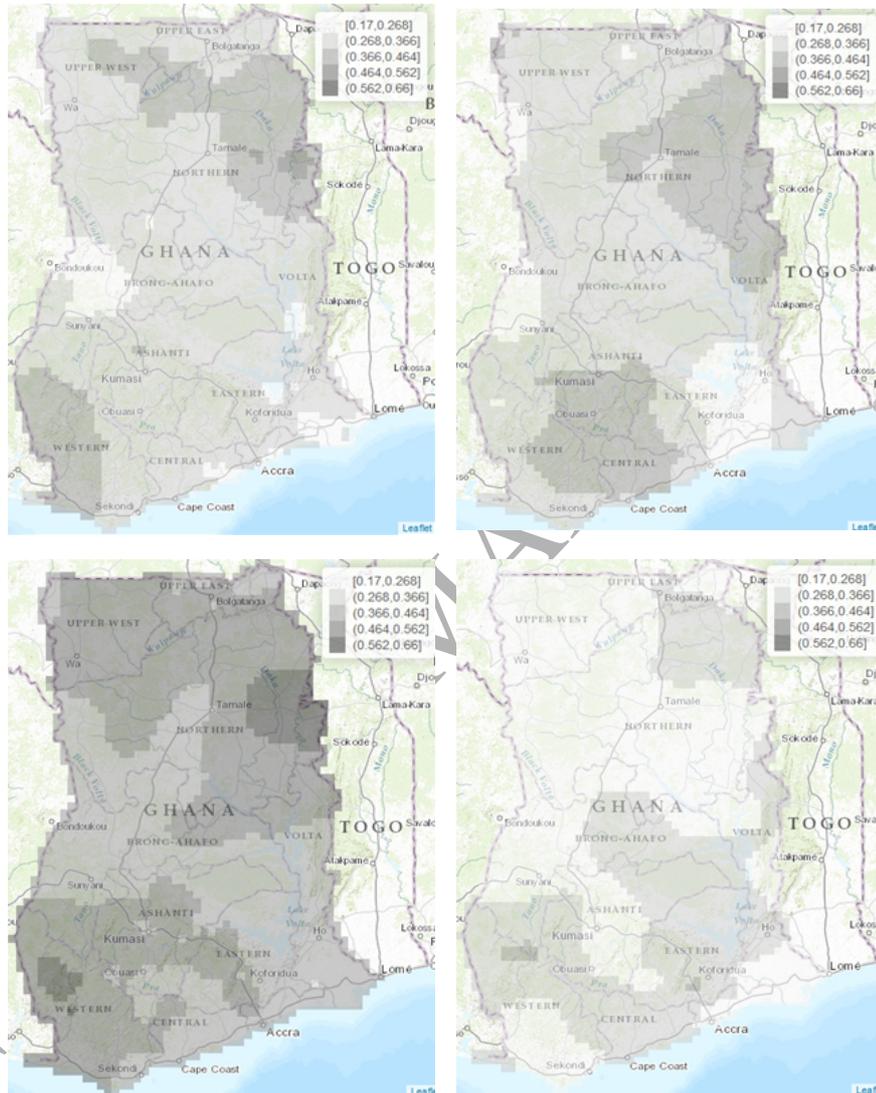
14

Figure 1: Probabilities of stunting in Ghana in 1993 (top left), 1998 (top right), 2003 (bottom left) and 2008 (bottom right) in children < 5 years
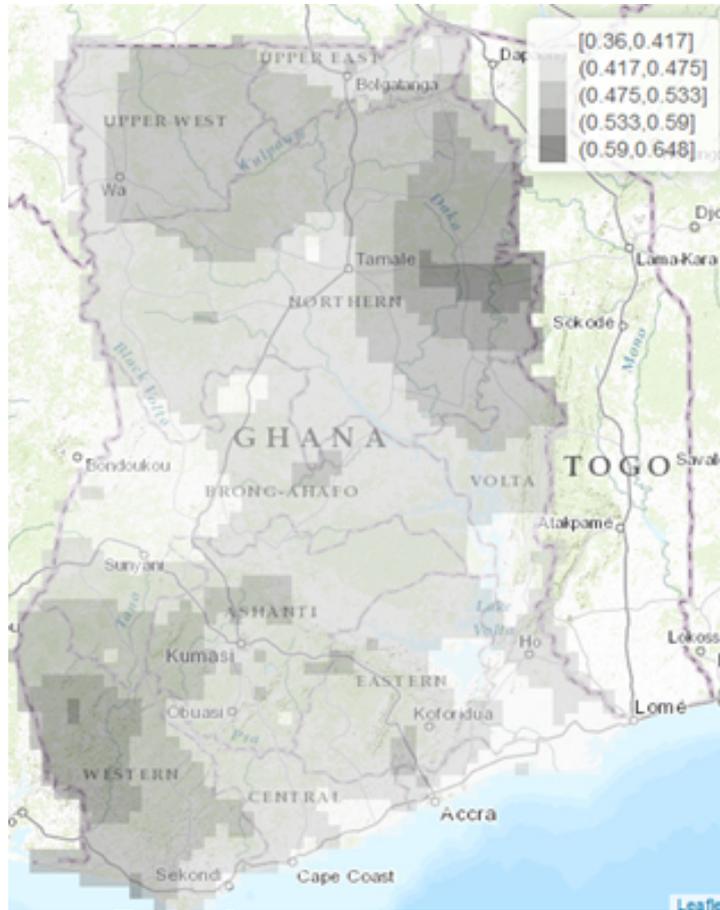
Figure 2: Overall probabilities of stunting forecast in Ghana in 2013 in children < 5 years

those observed in each of the actual surveys.

We also investigated the effect of mother's education on the risk of childhood chronic malnutrition across Ghana by making predictions of childhood malnutrition under different levels of maternal education in the country. Figure 315 3 presents our predicted maps of probabilities of childhood stunting based on mothers who had a minimum of 6 years and 12 years of education; note that both panels use the same grey-scale, but that this is different from the grey-scale used in Figures 1 and 2. The predicted stunting probability ranges from about
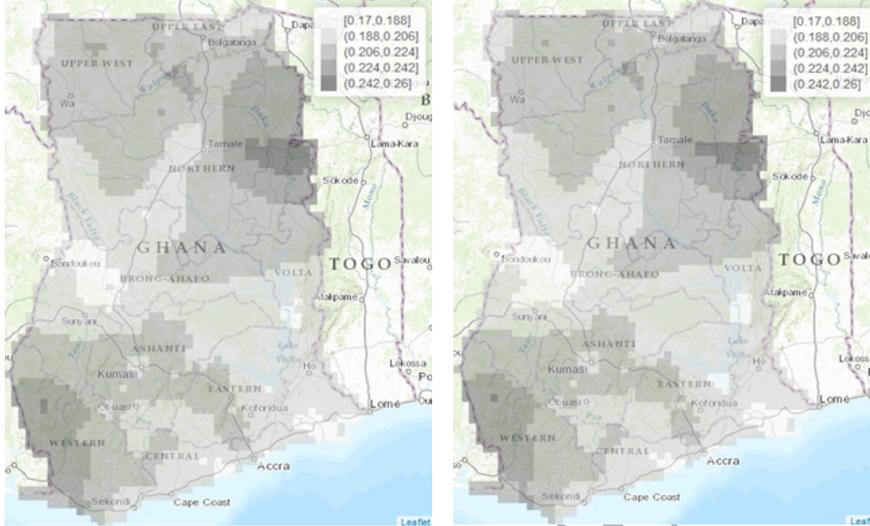
Figure 3: Effect of a minimum of 6 (left) and 12 (right) years of mother's education on probabilities of stunting forecast in Ghana in 2013 in children < 5 years

0.17 to 0.26.

4. Discussion

In this study, our aim has been to contribute to a better understanding of childhood malnutrition problems in developing countries like Ghana and to inform pre-emptive nutrition interventions. We constructed maps showing predicted probabilities of chronic malnutrition in children for the years 1993, 1998, 2003, 2008 and 2013. We have taken into account the uncertainties in our model parameters and in childhood characteristics implicit in these maps (see the methods section).

We have assumed that the residual process in our model has a stationary covariance function, allowing our measured covariates to describe the first-order spatial variation in malnutrition. One alternative would be to use a non-stationary covariance function for the residual process as in [9] but we did not pursue this and further research is warranted in this area to explore this issue.

17

Despite this, our spatiotemporal model works well: we observed high R-squared values, and closeness of our intercept values to zero (0) and their respective slope

335 values to one (1) as presented in Table 3 under supplementary material. This suggests that the regression of observed HAZ on predicted HAZ has a regression line reasonably close to the line of equality.

One advantage of our spatio-temporal model is that it is able to cope with the changing locations and quantity of sampled clusters in each of the surveys.

340 One of the assumptions we have made in our analysis is that the data gathered from the survey samples are representative of the underlying population. While the GDHS survey sample design attempts to achieve this, it is possible that there may be regions in which responses to the survey were not representative of the population; in such a case, more sophisticated analysis techniques would

345 be required in order to make inferences.

Our model produces predictions at each of the cluster locations that have historically occurred. If a new cluster location occurred in the next GDHS survey, we would simply expand the parameter space to include random effects for that cluster. These could then be updated along with other parameters and

350 we could make historical or future predictions of prevalence at that location.

We found substantial spatial and temporal variations in the probabilities of chronic malnutrition in Ghana, with median predicted prevalence of 33%, 33%, 47% and 26% for the years 1993, 1998, 2003 and 2008, respectively. We observed that living in parts of Northern or Western regions was associated

355 with relatively high probability of chronic malnutrition. Our forecast median stunting prevalence for the year 2013 in the country was 46%; living in parts of Northern and Western regions was again associated with significantly and materially higher probabilities of chronic malnutrition. Among all the ten regions of Ghana, Northern region was among the three regions in the northern

360 part of the country that are consistently the poorest; the other two are Upper East and Upper West.[40, 41, 23] This could partly explain the high chronic malnutrition prevalence patterns observed in Northern region. Our finding indirectly supports previous studies that reported that certain health outcomes

18

such as diarrhoea, wasting, and diarrhoea mortality exhibit spatial and tempo-

365 ral patterns.[13, 14, 10, 12, 15, 16, 9, 11] Our study is the first of it kind to use

spatio-temporal models to provide forecast maps for the risk of chronic malnu-

trition using repeated cross-sectional survey datasets, especially in developing

countries.

Using the same confounders, the risk of stunting is greatest in 2003 than

370 other years. This may be due to random variations which could not be controlled

for in this study. However, we did not rule out the possibility that the observed

pattern in the risk of stunting in 2003 could be real and further research in this

area could explore this issue.

According to the 'Feed the Future' program, a United States Government

375 global hunger and food security initiative, chronic malnutrition in the northern

regions of Ghana is related to household poverty levels, inadequate sanitation fa-

cilities and poor infant-feeding practices, and leads to higher disease burden.[42]

Our finding also supports a previous study conducted in Ghana among children

aged less than 5 years, which reported higher prevalence of malnutrition in the

380 northern part compared to the southern part of the country.[8, 23]

Children from mothers with higher levels of education were at less risk of

malnutrition. This might be because a mother's education is associated with

knowledge of good practices for herself and her child (e.g. feeding and health

seeking behaviour) but is also associated with household socioeconomic status

385 and access to food.[20, 8, 22, 23] In our predictive maps shown in Figure3,

increasing the level of maternal education was shown to reduce the prevalence

of malnutrition throughout Ghana. However, policies and interventions aimed

at improving the level of maternal education to combat childhood malnutrition

must be supported with efforts to improve the general standard of living among

390 households in the communities.

Our unexpected finding that months of breastfeeding have negative associ-

ation with HAZ could be the result of poverty among households in Ghana;

consequently, mothers may continue to breast feed beyond the recommended 6

months without supplementation.[20, 43] Also, our finding that child's age is

19

negatively associated with HAZ could be attributable to presence of progressive childhood diseases and deficit in proper complementary foods.[34, 35]

The parameter estimates from our spatio-temporal model characterise the nature of the spatio-temporal variations in nutritional outcomes of children. We observed a weak, and statistically non-significant, time-dependence between successive surveys, suggesting that a spatial model for each of the survey years separately could have fitted the data as well as our spatio-temporal model. However, one of our main aims in this paper is to forecast the risk of childhood chronic malnutrition, which is only possible through a spatio-temporal model. A longitudinal survey design, with repeated measurements in the same household, would undoubtedly have led to stronger correlation between successive surveys

The study's strengths include representativeness and national coverage, which makes the findings relevant to the wider population of Ghana. In addition, our spatio-temporal model permitted us to borrow strength from similar previous surveys conducted in 1993, 1998, 2003 and 2008 to provide a forecast for stunting prevalence in the country in 2013.

The study's limitations include lack of access to other possible risk-factors such as data on interventions (nutrition, health care and socioeconomic impacts) and distance to health care facilities in the communities. Also, despite the fact that cluster sampling is a cost-saving approach without the need to enumerate all the households in the country, statistically it poses analytical challenges in that independence in observational units might not be guaranteed. As a result, statistical analyses that rely upon the assumption of independence of households drawn will no longer be a valid assumption. Finally, the Demographic and Health Survey typically displaces the cluster locations in order to protect respondent confidentiality and this inherent nature could possibly have an impact on the final predictions since the calculated distances between clusters are likely to be measured with increased error.

Previous studies in developing countries like Ghana have shown that socioeconomic, cultural and demographic factors, feeding and care practices are associated with nutritional outcomes of children under five years of age.[18, 20,

20

31, 22, 34, 44, 45, 46] However, these studies did not explore or forecast spatio-temporal variations in nutritional outcomes.

Generally, lack of food availability, poverty, lack of dietary diversity, inadequate care and feeding practices coupled with inadequate or complete lack of access to health care services are among the major factors responsible for malnutrition in children. [27, 40, 23]

Our results can be used to target public health nutrition interventions for high-risk communities so as to improve overall health and nutrition of Ghanaian children in two different ways. Firstly, as with our example of maternal education, our model enables prediction of the effectiveness of a public health intervention that changes the distribution of an included explanatory variable. Secondly, our maps can identify priority areas for implementation of an intervention that is known to be effective but is too expensive to implement universally.

5. Conclusion

Our forecast maps of probabilities of stunting for the year 2013 illustrate how our spatio-temporal modelling approach can be used as a tool for identifying high risk communities. This can help to prioritise and target nutrition and health policies that can promote effective and sustainable public health interventions in Ghana and other developing countries, with the overall aim of improving childhood nutrition and health. We also intend to apply our current spatio-temporal modelling techniques to update our model and provide forecasts for the year 2019, as soon as the 2014 GDHS data-set is made available.

References

[1] R. E. Black, L. H. Allen, Z. A. Bhutta, L. E. Caulfield, M. de Onis, M. Ezzati, C. Mathers, J. Rivera, M. Group, C. U. Study, Maternal and child undernutrition: global and regional exposures and health consequences, Lancet 371 (9608) (2008) 243–60.

21

[2] M. de Onis, M. Blssner, The World Health Organization Global Database on Child Growth and Malnutrition: methodology and applications, Int J Epidemiol 32 (4) (2003) 518–26.

[3] D. L. Pelletier, E. A. Frongillo, Changes in child survival are strongly associated with changes in malnutrition in developing countries, J Nutr 133 (1) (2003) 107–19.

[4] UNICEF, WHO, The World Bank, Joint child malnutrition estimates: Levels and trends. [last accessed 21/07/2015], `http://www.who.int/nutgrowthdb/estimates/en/index.html` (29/03/2013 2012).

[5] D. Pelletier, E. Frongillo, J. Habicht, Epidemiological evidence for a potentiating effect of malnutrition on child mortality, American journal of Public Health 83 (1993) 1133–1139.

[6] The World Bank, Repositioning nutrition as central to development: A strategy for large-scale action (2006).

[7] M. Ezzati, A. D. Lopez, A. Rodgers, S. Vander Hoorn, C. J. Murray, C. R. A. C. Group, Selected major risk factors and global and regional burden of disease, Lancet 360 (9343) (2002) 1347–60.

[8] ORC Macro, Nutrition of young children and mothers in Ghana: Findings from the 2003 Ghana Demographic and Health Survey, `http://www.measuredhs.com/publications/publication-ANC16-Africa-Nutrition-Chartbooks.cfm` (2005).

[9] D. K. Kinyoki, J. A. Berkley, G. M. Moloney, E. O. Odundo, N. B. Kandala, A. M. Noor, Space-time mapping of wasting among children under the age of five years in somalia from 2007 to 2010, Spat Spatiotemporal Epidemiol 16 (2016) 77–87.

[10] N. Chaikaew, N. K. Tripathi, M. Souris, Exploring spatial patterns and hotspots of diarrhea in Chiang Mai, Thailand, Int J Health Geogr 8 (2009) 36.

22

[11] S. Pande, M. A. Keyzer, A. Arouna, B. G. Sonneveld, Addressing diarrhea prevalence in the West African Middle Belt: social and geographic dimensions in a case study for Benin, Int J Health Geogr 7 (2008) 17.

[12] D. Hamer, J. Simon, D. Thea, G. Keusch, Childhood diarrhea in sub-saharan africa, Child Health Res Project Special Rep 2 (1998) 32.

[13] W. J. Alonso, R. Acua-Soto, R. Giglio, J. Nuckols, S. Leyk, C. Schuck-Paim, C. Viboud, M. Miller, B. McCormick, Spatio-temporal patterns of diarrhoeal mortality in Mexico, Epidemiol Infect 140 (1) (2012) 91–9.

[14] M. Azage, A. Kumie, A. Worku, A. C. Bagtzoglou, Childhood diarrhea exhibits spatiotemporal variation in northwest ethiopia: A satscan spatial statistical analysis, PLoS One 10 (2015) e0144690.

[15] S. Leyk, B. J. McCormick, J. R. Nuckols, Effects of varying temporal scale on spatial models of mortality patterns attributed to pediatric diarrhea, Spat Spatiotemporal Epidemiol 2 (2011) 91–101.

[16] S. Leyk, T. Phillips, J. Smith, J. R. Nuckols, Spatiotemporal trends of diarrheal mortality of children in association with hydrographic regions of brazil, Cart Geogr Inform Sc 38 (2) (2011) 223–232.

[17] WHO, Handbook: Integrated Management of Childhood Illness (IMCI)., `http://apps.who.int/iris/bitstream/10665/42939/1/9241546441.pdf` (2005).

[18] S. Antwi, Malnutrition: missed opportunities for diagnosis, Ghana Medical Journal 42 (2008) 101–104., `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2643431/`.

[19] UNICEF–Ghana, Issues facing children in Ghana. [last accessed 21/07/2015], `http://www.unicef.org/wcaro/Countries_1743.html` (2008).

[20] J. M. K. Aheto, T. J. Keegan, B. M. Taylor, P. J. Diggle, Childhood malnutrition and its determinants among under-five children in Ghana, `http://onlinelibrary.wiley.com/doi/10.1111/ppe.12222/abstract` (2015).

[21] Ghana Statistical Service, Ghana Health Service, ICF Macro, Ghana Demographic and Health Survey 2008 (2009).

[22] N. B. Kandala, T. P. Madungu, J. B. Emina, K. P. Nzita, F. P. Cappuccio, Malnutrition among children under the age of five in the Democratic Republic of Congo (DRC): does geographic location matter?, BMC Public Health 11 (2011) 261.

[23] E. Van de Poel, A. R. Hosseinpoor, C. Jehu-Appiah, J. Vega, N. Speybroeck, Malnutrition and the disproportional burden on the poor: the case of Ghana, Int J Equity Health 6 (2007) 21.

[24] Ghana Statistical Service, Macro International Inc., Ghana Demographic and Health Survey 1998 (1999).

[25] Ghana Statistical Service, Macro International Inc., Ghana Demographic and Health Survey 1993 (1994).

[26] Ghana Statistical Service, Noguchi Memorial Institute for Medical Research, ORC Macro, Ghana Demographic and Health Survey 2003 (2004).

[27] M. de Onis, M. Blssner, WHO global database on child growth and malnutrition, `http://whqlibdoc.who.int/hq/1997/WHO_NUT_97.4.pdf` (1997).

[28] WHO, Physical status: the use and interpretation of anthropometry. report of a WHO expert committee, World Health Organ Tech Rep Ser 854 (1995) 1–452.

[29] M. Pradhan, D. E. Sahn, S. D. Younger, Decomposing world health inequality, J Health Econ 22 (2) (2003) 271–93.

24

[30] WHO Multicentre Growth Reference Study Group, WHO Child Growth Standards based on length/height, weight and age, Acta Paediatr Suppl 450 (2006) 76–85.

535 [31] N. J. Bomela, Social, economic, health and environmental determinants of child nutritional status in three Central Asian Republics, Public Health Nutr 12 (10) (2009) 1871–7.

[32] R. Babatunde, F. Olagunju, S. Fakayode, F. Sola-Ojo, Prevalence and determinants of malnutrition among under-five children of farming households 540 in Kwara state, Nigeria, Journal of Agricultural Science 3 (3) (2011) 173– 181.

[33] J. Kabubo-Mariara, G. K. Ndenge, D. K. Mwabu, Determinants of childrens nutritional status in Kenya: Evidence from demographic and health surveys, Journal of African Economies 18 (3) (2009) 363387, `http://jae.` 545 `oxfordjournals.org/content/18/3/363.abstract`.

[34] V. T. Adekanmbi, G. A. Kayode, O. A. Uthman, Individual and contextual factors associated with childhood stunting in Nigeria: a multilevel analysis, Maternal and Child Nutrition 9 (2) (2013) 244–59.

[35] J. Alom, M. A. Quddus, M. A. Islam, Nutritional status of under-five 550 children in Bangladesh: a multilevel analysis, J Biosoc Sci 44 (5) (2012) 525–35.

[36] WHO, WHO Anthro for personal computers, version 3.2.2, 2011: Software for assessing growth and development of the world's children., `http://www.` `who.int/childgrowth/software/en/` (2010).

555 [37] R. E. Kalman, A new approach to linear filtering and prediction problems, `https://www.cs.unc.edu/~welch/kalman/media/pdf/` `Kalman1960.pdf` (1960).

[38] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing (2015).

25

560 [39] B. Taylor, miscfuncs: Miscellaneous useful functions. R package version 1.2-7, `http://CRAN.R-project.org/package=miscFuncs` (2015).

[40] Food and Agriculture Organization, Nutrition country profiles-ghana. [last accessed may 2015], `http://www.fao.org/ag/AGN/nutrition/gha_en.stm` (2009).

565 [41] Ghana Statistical Service, Ghana Living Standards Survey: Report of the fifth round (GLSS 5), `http://www.statsghana.gov.gh/docfiles/glss5_report.pdf` (2008).

[42] Feed the Future, Ghana country profile. [last accessed october, 2015], `http://www.feedthefuture.gov/country/ghana` (2015).

570 [43] E. Van de Poel, A. R. Hosseinpoor, C. Jehu-Appiah, J. Vega, N. Speybroeck, Socioeconomic inequality in malnutrition in developing countries., Bulletin of the World Health Organization 86 (2008) 282–291.

[44] R. Brugha, J. Kevany, Determinants of nutrition status among children in the eastern region of Ghana, J Trop Pediatr 40 (5) (1994) 307–11.

575 [45] D. A. Amugsi, M. B. Mittelmark, A. Lartey, D. J. Matanda, H. B. Urke, Influence of childcare practices on nutritional status of Ghanaian children: a regression analysis of the Ghana Demographic and Health Surveys, BMJ Open 4 (11) (2014) e005340, `http://www.ncbi.nlm.nih.gov/pubmed/25366675`. `doi:10.1136/bmjopen-2014-005340`.

580 [46] V. T. Adekanmbi, O. A. Uthman, O. M. Mudasiru, Exploring variations in childhood stunting in Nigeria using league table, control chart and spatial analysis, BMC Public Health 13 (2013) 361.