Received: 14 September 2020 | Accepted: 2 August 2021

DOI: 10.1111/rssa.12746

ORIGINAL ARTICLE

A hidden Markov space–time model for mapping the dynamics of global access to food

Francesco Bartolucci[1] | **Alessio Farcomeni**[2]

[1]University of Perugia, Perugia, Italy

[2]University of Rome "Tor Vergata", Rome, Italy

Correspondence
Alessio Farcomeni, University of Rome "Tor Vergata", Via Columbia, 2, 00133 Rome, Italy.
Email: alessio.farcomeni@uniroma2.it

Abstract

In order to analyse worldwide data about access to food, coming from a series of Gallup's world polls, we propose a hidden Markov model with both a spatial and a temporal component. This model is estimated by an augmented data MCMC algorithm in a Bayesian framework. Data are referred to a sample of more than 750 thousand individuals in 166 countries, widespread in more than two thousand areas, and cover the period 2007–2014. The model is based on a discrete latent space, with the latent state corresponding to a certain area and time occasion that depends on the states of neighbouring areas at the same time occasion, and on the previous state for the same area. The latent model also accounts for area-time-specific covariates. Moreover, the binary response variable (access to food, in our case) observed at individual level is modelled on the basis of individual-specific covariates through a logistic model with a vector of parameters depending on the latent state. Model selection, in particular for the number of latent states, is based on the Watanabe–Akaike information criterion. The application shows the potential of the approach in terms of clustering the areas, data smoothing and prediction of prevalence for areas without sample units and over time.

KEYWORDS

data augmentation, data smoothing, MCMC, prediction, Watanabe–Akaike information criterion

1 | INTRODUCTION

Data with both spatial and temporal dimensions are nowadays produced in many fields, and require sophisticated statistical models and inferential methods to be analysed. In this paper we focus on data deriving from Gallup's world poll (GWP) surveys that, each year, are based on samples of individuals from several countries and territories. Gallup is an American firm that, since 1935, conducts public polls worldwide. GWP data are commonly used to investigate socio-economic issues, especially concerning aspects related to well-being (e.g. Deaton, 2008; Frongillo et al., 2017; Powdthavee et al., 2017).

Among the aspects examined by the GWP questionnaire, we focus in particular on lack of access to food, a theme that nowadays is gathering much interest in the study of poverty (e.g. Bhattacharya et al., 2004; Mahadevan & Hoang, 2016; Nord et al., 2008; Rose, 1999; Smith et al., 2017; Suryanarayana & Silva, 2007). This aspect is observed through a binary response variable that is individual specific and indicates whether the household was not able to afford food within the past 12 months. The available data are collected in 166 world countries and territories for the period 2007–2014 (eight waves). In our survey, each country is suitably divided in areas, for a total of more than two thousand areas. A limitation of the world poll is that, with simple models, accurate estimates can be aggregated at most at the country/territory level. It is well known that dramatic heterogeneity can be present within many countries, where certain regions can be at low risk of poverty, while other regions of the same country can be at high risk. It would be therefore useful for politicians, stakeholders, charities and non-governmental organizations to have access to estimates with a greater spatial disaggregation. Formally, our main interest is in estimating prevalence of people that, at area level, are at least occasionally not able to afford food for themselves or their family. We do so by pooling spatial and temporal information, in order to reduce variability due to the possible small number of individuals sampled in the specific area of interest. A series of covariates are also available for every individual which will be used to further decrease the mean squared error of the estimates. The geographic subregion of each country is our target for producing estimates.

For the analysis of such data we propose a spatio-temporal approach based on the assumption that the response variable follows a logistic model including the individual-specific covariates and depending on vectors of regression coefficients that are area and wave specific. This is formalized by associating a discrete latent variable (with a finite number of categories) to every area and wave. Each category corresponds to a different vector of regression coefficients. These latent variables are assumed to follow a Markov model with spatial and time dependence. In particular, every area- and wave-specific latent variable is assumed to depend on the latent variables associated with the neighbours at the same time occasion and to the latent variable for the same area at the previous time occasion. The latent state can be interpreted in terms of area- and wave-specific propensity to lack of access to food. Similar models have been used so far, to the best of our knowledge, only to tackle spatial or temporal aspects separately (e.g. Dotto et al., 2019; Li Donni & Marino, 2018).

The assumptions mentioned above give rise to a hierarchical model that may be seen as a temporal extension of a hidden Markov field model (Green & Richardson, 2002; Qian & Titterington, 1991; Spezia et al., 2018), with covariates; or as a spatial extension of a hidden Markov model (Bartolucci & Farcomeni, 2015; Bartolucci et al., 2013, 2014; Zucchini et al., 2017), again with covariates. As in many other approaches, in formulating the assumed latent structure, we take inspiration from the seminal paper of Besag (1986). For spatio-temporal approaches related to the proposed one, but applied in different fields, see Wei and Li (2008), Ailliot et al. (2009), Lawson (2013 Ch. 12) and Lin et al. (2015).

We adopt a Bayesian approach for model estimation. We refer the reader to Marshall (1991) for an introduction to spatial clustering, and to Best et al. (2005) and Lawson (2013) specifically for Bayesian models for spatial (disease) mapping. The proposed model is estimated by a Markov chain Monte Carlo (MCMC) algorithm based on data augmentation (Tanner & Wong, 1987), as we treat the latent variables on the same footing as the model parameters that are updated at each iteration of the algorithm. We also deal with label switching (Stephens, 2000), by post-processing the MCMC output, and model selection, on the basis of the Watanabe–Akaike information criterion (WAIC; Watanabe, 2010), so that the model marginal likelihood is not necessary. `R` code with an implementation of our approach is available from https://github.com/afarcome/LMsae. This implementation includes Fortran routines to speed up the computation given the very large sample size, which is over 750 thousand individuals.

The main feature of the proposed approach is that it allows us to cluster country areas into a finite number of groups in a dynamic fashion. Exploiting the dependence in space and time of the latent variables, this clustering is possible also when, for a specific wave, data are not available in a certain area because no individuals have been sampled. Moreover, it is possible to predict the prevalence of at least occasional lack of access to food for every area and wave. The prediction is a smoothed (less noisy) version of the observed one, when data are observed, or a pure prediction, when no data are observed. As we illustrate in the application, this allows us to make nice graphical representations in the form of dynamic maps of the characteristic of interest. The approach here proposed may be easily extended to deal with different situations when we observe more response variables, for each individual, even if these variables are mixed discrete and continuous.

The paper is organized as follows. In the next section we provide a description of the available data. The proposed model is illustrated in Section 3, whereas Bayesian inference based on the MCMC algorithm is described in Section 4. The results of the analysis are described in Section 5 and last section reports some conclusions.

2 | DATA DESCRIPTION

The GWP is a survey conducted by interviewing nationally representative samples of the adult population (aged 15 and older) in almost every world country or territory. The sampling is repeated each year, so that different individuals are interviewed within each country at different waves. The survey covers a range of topics including family economics, employment, human development and well-being. About 1,000 individuals for every country are included each year by Gallup, with some variability and the exception of about 3,000 for India and 5,000 for China. Some oversampling has occurred, for example in major cities, in certain areas. See Gallup organization (2020) for specific details. Sampling involves a random-digit-dial telephone survey in countries where more than 80% of the population has landline phones, and area frame designs for face-to-face interviewing otherwise.

In our final data set we include a total of 760,282 subjects, essentially representative of the world population for the years from 2007 to 2014. Our aim is to investigate the area-time-specific prevalence of subjects who replied ‘Yes’ to the following question:

> 'Have there been times in the past 12 months when you did not have enough money to buy food that you or your family needed?'

Over the period, 198,788 subjects replied 'Yes' to this question (coded as 1 in our analysis and tagged as 'food insecure' in the rest of the paper) and the remaining replied 'No' (coded as 0 in our analysis and tagged as 'food secure' in the rest of the paper). The 'food insecure' terminology is used only for simplicity: it shall be here noted that food insecurity is a broader concept than the mere lack of money to buy food, and that it shall therefore be kept in mind that our target is simply the prevalence of subjects who at least occasionally lack the money to buy food for them or their family. Before proceeding further, we must warn that some measurement error might be influencing the results. First of all, some subjects might not reply sincerely to this sensitive question. Additionally, given the way it is formulated, there might be some recall bias. Other issues involve measurement invariance and intended meaning. Some measurement errors might arise from the fact that the question at hand is asked to people of different cultures, speaking different languages. In order to minimize translation issues Gallup prepares and validates questionnaires in English, French and Spanish. A professional translator then uses one or more of the main questionnaires to prepare that in the language spoken in the sampled household. A second professional translator compares original and translated questionnaire, and suggests refinements.

The overall sample prevalence is 26.1%. This prevalence shows a strong spatial and temporal heterogeneity. In the upper panel of Table 1 we show, for illustration, raw prevalence estimates stratified by continent and year. In the Supplementary Material we give a complete graphical account reporting maps of raw and predicted prevalence estimates.

It is obvious that a very large spatial heterogeneity exists for the proportion of food insecure in the world, and also (and possibly more interesting) temporal variability. In addition to identifying a country for each respondent, Gallup also releases information on the main subregion of residence within the country. Each country is divided in a variable number of subregions, with several countries having more than 25 subregions (e.g. Russia, China, India, Brazil, etc.). Clearly, some areas might not have been sampled in certain years. In the bottom panel of Table 1 we show the number of missing areas, by continent and wave.

Our main objective in this study is to obtain a reliable ('smoothed') estimate of prevalence of food insecure within each country-specific area, by pooling information over neighbouring areas and time points. We end up working with 2,141 areas that cover all continents, with the exception of Antarctica.

In order to pool information we also use two different sets of covariates. Covariates used at site level are *latitude*, *longitude*, *the absolute value of latitude* and *indicator variables for the continent* (using the classification in six continents that distinguishes between North and South America). In this way we model fixed effects with the logic that geographically closer areas tend to be similar and at the same time we account for more or less globally disadvantaged world regions and the general effect of distance from the equator. We recall that these covariates affect the distribution of the area-specific latent variables that may be used to cluster directly these areas in distinct groups. At subject level we will use *age* (median: 40 years, inter-quartile range: 28), *squared age*, *gender* (46% females) and *quintiles of equivalized disposable income within the country* included by suitable indicator variables. These can be expected to be related to access to food, at least in some countries/territories. The average age stratified by continent and wave is reported in the second upper panel of Table 1, while in the third panel we report the stratified proportion of females. We observe that age of the respondents is slightly variable

TABLE 1 Empirical statistics stratified by continent and year. First panel: prevalence of food insecure. Second panel: average age. Third panel: proportion of females. Fourth panel: number of areas not sampled

		2007	2008	2009	2010	2011	2012	2013	2014
Prevalence of food insecure	Africa	0.49	0.53	0.42	0.43	0.43	0.43	0.52	0.51
	Americas	0.07	0.08	0.10	0.36	0.35	0.33	0.37	0.33
	Asia	0.25	0.23	0.22	0.23	0.21	0.22	0.26	0.26
	Europe	0.20	0.13	0.18	0.15	0.12	0.11	0.17	0.14
	Oceania	0.11	0.11	0.11	0.10	0.12	0.14	0.09	0.10
Average age	Africa	34.11	34.26	34.93	35.05	35.65	35.61	35.43	35.32
	Americas	52.84	49.00	50.40	41.75	42.40	42.12	42.39	44.02
	Asia	39.16	40.25	38.00	38.76	38.87	39.62	39.74	41.27
	Europe	44.57	48.27	48.20	49.01	50.54	50.98	47.95	48.44
	Oceania	44.10	44.32	46.95	50.81	50.29	49.21	53.20	54.90
Percentage of females	Africa	0.51	0.52	0.51	0.52	0.52	0.50	0.52	0.50
	Americas	0.47	0.43	0.45	0.44	0.43	0.43	0.43	0.44
	Asia	0.45	0.46	0.48	0.48	0.47	0.47	0.47	0.49
	Europe	0.40	0.43	0.41	0.40	0.43	0.44	0.43	0.46
	Oceania	0.49	0.49	0.44	0.38	0.38	0.43	0.41	0.42
Number of areas not sampled	Africa	618	503	480	399	353	349	351	0
	Americas	309	301	300	87	88	90	52	0
	Asia	424	239	210	205	156	112	61	6
	Europe	576	452	403	336	275	264	273	0
	Oceania	31	0	31	0	0	0	1	0

over time, with more substantial differences between continents. Gender proportions do not show a strong variability, as could be expected.

3 | SET UP AND MODEL ASSUMPTIONS

Let n denote the number of sites (areas) and let m_{jt} be the number of sampled units in site j at occasion t, with $j = 1, \dots, n$ and $t = 1, \dots, T$. Note that, by design, unit i in site j at occasion t is in general different from unit i in the same site j at another occasion t'. For each of these units we observe a column vector of covariates $\boldsymbol{x}_{ijt}$ and a binary outcome y_{ijt}; all response variables are collected in the set $\mathcal{Y}$ having $\sum_{j=1}^{n} \sum_{t=1}^{t} m_{jt}$ elements. In order to maintain a concise exposition, we use the same symbol for a random variable or one of its realizations. The same convention will be used for random vectors and matrices.

We also observe, at site level, column vectors of covariates denoted by $\boldsymbol{z}_{jt}$. The n-dimensional binary symmetric matrix $\boldsymbol{C}$ is used to identify neighbouring areas, with $c_{jj'} = 1$ if site j' is a neighbour of site j on a lattice that is not necessarily regular. The proposed multilevel spatial–temporal model is based on latent variable u_{jt} for every site j and time occasion t. These latent variables, collected in the $n \times T$ matrix $\boldsymbol{U}$, are discrete with k support points, labelled from 1 to k. A crucial assumption is that, given the covariates and the latent variables, the response variables y_{ijt} are conditionally independent with conditional Bernoulli distribution, that is,

$$y_{ijt}|u_{jt} \sim \text{Bin}(1, q_{ijt}(u_{jt})), \quad u_{jt} = 1, \ldots, k,$$

where the success probability is formulated according to the logistic model

$$q_{ijt}(u) = p(y_{ijt} = 1|u_{jt} = u) = \frac{\exp(\boldsymbol{x}'_{ijt}\boldsymbol{\beta}_u)}{1 + \exp(\boldsymbol{x}'_{ijt}\boldsymbol{\beta}_u)}, \tag{1}$$

based on regression parameters $\boldsymbol{\beta}_u$ depending on the latent state for site j at occasion t. These regression parameters are collected in the matrix $\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_k)'$. Vectors $\boldsymbol{x}_{ijt}$ are assumed to be known, and include an initial element for intercept. A similar assumption is adopted for the second-level covariates $\boldsymbol{z}_{jt}$. The success probabilities will also be denoted by $q_{ijt}(u, \boldsymbol{B})$ to stress their dependence on the parameters in $\boldsymbol{B}$.

Regarding the distribution of the latent variables, we assume a Markov chain that accounts for the spatial structure. In particular let $\tilde{\boldsymbol{u}}_{jt}$ be the vector of latent variables $u_{j't}$ if j' is a neighbour of j, namely $c_{jj'} = 1$. Taking inspiration from the auto-logistic model (Besag, 1974), we assume that the initial probabilities of the hidden Markov chain, conditional on the states of the neighbours, are equal to

$$\begin{aligned}\lambda_j(u|\tilde{\boldsymbol{u}}_{j1}) &= p(u_{j1} = u|\tilde{\boldsymbol{u}}_{j1})\\ &= \frac{1}{1 + \sum_{v=2}^{k} \exp(\boldsymbol{z}_{j1}(\tilde{\boldsymbol{u}}_{j1})'\boldsymbol{\gamma}_v)} \begin{cases} 1, & u = 1, \\ \exp(\boldsymbol{z}_{j1}(\tilde{\boldsymbol{u}}_{j1})'\boldsymbol{\gamma}_u), & u = 2, \ldots, k, \end{cases}\end{aligned}$$

where the denominator is the normalizing constant that obviously does not vary with u and, in general, $\boldsymbol{z}_{jt}(\tilde{\boldsymbol{u}}_{jt}) = (\boldsymbol{z}'_{jt}, \boldsymbol{f}(\tilde{\boldsymbol{u}}_{jt})')'$, with $\boldsymbol{f}(\tilde{\boldsymbol{u}}_{jt})$ being a column vector summarizing the states of the neighbours of site j at occasion t. In our implementation we assume that $\boldsymbol{f}(\tilde{\boldsymbol{u}}_{jt})$ is a k-dimensional vector with the uth element equal to the proportion of neighbours having latent state equal to u and $\boldsymbol{z}_{jt}(\tilde{\boldsymbol{u}}_{jt})$ is specified so as to avoid identifiability problems.

For the column vector $\lambda_j(\tilde{\boldsymbol{u}}_{j1}) = (\lambda_j(1|\tilde{\boldsymbol{u}}_{j1}), \ldots, \lambda_j(k|\tilde{\boldsymbol{u}}_{j1}))'$ we use an overall parametrization based on the matrix notation that simplifies the implementation of the estimation algorithm. Including all $\boldsymbol{\gamma}_u$ vectors in the matrix $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_k)'$, we have that

$$\lambda_j\left(\tilde{\boldsymbol{u}}_{j1}\right) = \frac{1}{\boldsymbol{1}'\exp(\boldsymbol{M}_1\boldsymbol{\Gamma}\boldsymbol{z}_{j1}(\tilde{\boldsymbol{u}}_{j1}))}\exp(\boldsymbol{M}_1\boldsymbol{\Gamma}\boldsymbol{z}_{j1}(\tilde{\boldsymbol{u}}_{j1})),$$

where in general $\boldsymbol{M}_u$ is an identity matrix of dimension k, without the uth column. In the following we also use the notation $\lambda_j(u|\tilde{\boldsymbol{u}}_{j1}, \boldsymbol{\Gamma})$ for the elements of this vector to stress their dependence on $\boldsymbol{\Gamma}$.

Regarding the transition probabilities we assume a similar parametrization, but using the starting state as reference category, that is,

$$\begin{aligned}\pi_{jt}(u|u', \tilde{\boldsymbol{u}}_{jt}) &= p(u_{jt} = u|u_{j,t-1} = u', \tilde{\boldsymbol{u}}_{jt})\\ &= \frac{1}{1 + \sum_{\substack{v=1\\ v\neq u'}}^{k} \exp(\boldsymbol{z}_{jt}(\tilde{\boldsymbol{u}}_{jt})'\boldsymbol{\delta}_{u'v})} \begin{cases} 1, & u = u', \\ \exp(\boldsymbol{z}_{jt}(\tilde{\boldsymbol{u}}_{jt})'\boldsymbol{\delta}_{u'u}), & u = 1, \ldots, k, \ u \neq u', \end{cases}\end{aligned}$$

for $u' = 1, \ldots, k$. Note that each area-time-specific latent state depends on the latent states of the neighbours for the same time occasion (taking again inspiration from the auto-logistic model) and on the previous latent state for the same site (as in a standard hidden Markov model). Also in this case we can adopt a matrix formulation. Including the elements in the u'th row of the transition matrix, that is, $\pi_{jt}(u|u', \tilde{\boldsymbol{u}}_{jt}), u = 1, \ldots, k$, in the vector $\boldsymbol{\pi}_{jt}(u', \tilde{\boldsymbol{u}}_{jt})$ and with $\boldsymbol{\Delta}_{u'} = (\boldsymbol{\delta}_{u'1}, \ldots, \boldsymbol{\delta}_{u',u'-1}, \boldsymbol{\delta}_{u',u'+1}, \boldsymbol{\delta}_{u'k})'$, we have that

$$\boldsymbol{\pi}_{it}(u', \tilde{\boldsymbol{u}}_{jt}) = \frac{1}{\boldsymbol{1}'\exp(\boldsymbol{M}_{u'}\boldsymbol{\Delta}_{u'}\boldsymbol{z}_{jt}(\tilde{\boldsymbol{u}}_{jt}))}\exp(\boldsymbol{M}_{u'}\boldsymbol{\Delta}_{u'}\boldsymbol{z}_{jt}(\tilde{\boldsymbol{u}}_{jt})), \quad u' = 1, \ldots, k.$$

We also denote by $\boldsymbol{\Delta}$ the matrix collecting parameter vectors $\boldsymbol{\delta}_{u'u}$ for $u', u = 1, \ldots, k$, with $u \neq u'$, and in the following we use the notation $\pi_{jt}(u|u', \tilde{\boldsymbol{u}}, \boldsymbol{\Delta})$ for the elements of the above vector.

Regarding prior distributions, we assume that all parameters are independent with

$$\begin{aligned}\boldsymbol{\beta}_u &\sim N(\boldsymbol{0}, \sigma^2_\beta \boldsymbol{I}), \quad u = 1, \ldots, k,\\ \boldsymbol{\gamma}_u &\sim N(\boldsymbol{0}, \sigma^2_\gamma \boldsymbol{I}), \quad u = 2, \ldots, k,\\ \boldsymbol{\delta}_{u'u} &\sim N(\boldsymbol{0}, \sigma^2_\delta \boldsymbol{I}), \quad u' = 1, \ldots, k, \; u = 2, \ldots, k,\end{aligned}$$

where $\boldsymbol{0}$ and $\boldsymbol{I}$ are a column vector of zeros and an identity matrix of suitable dimension, respectively, and σ^2_β, σ^2_γ and σ^2_δ are variance parameters that in our application are fixed at large values, such as 100. It shall be noted that in our application we did not observe relevant sensitivity to these prior inputs.

4 | BAYESIAN INFERENCE

Bayesian inference is based on the posterior distribution of model parameters and latent parameters, which may be expressed as

$$p(\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\Delta}, \boldsymbol{U}|\mathcal{Y}) \propto p(\boldsymbol{B})p(\boldsymbol{\Gamma})p(\boldsymbol{\Delta})p(\boldsymbol{U}|\boldsymbol{\Gamma}, \boldsymbol{\Delta})p(\mathcal{Y}|\boldsymbol{U}, \boldsymbol{B}), \tag{2}$$

where we explicitly indicate the dependence of the distributions of the parameters and we use the usual proportionality symbol $\propto$. In the previous expression, $p(\boldsymbol{B})$, $p(\boldsymbol{\Gamma})$ and $p(\boldsymbol{\Delta})$ refer to the prior distribution of the corresponding parameters and have density functions

$$\begin{aligned}p(\boldsymbol{B}) &= \prod_{u=1}^{k} p(\boldsymbol{\beta}_u),\\ p(\boldsymbol{\Gamma}) &= \prod_{u=2}^{k} p(\boldsymbol{\gamma}_u),\\ p(\boldsymbol{\Delta}) &= \prod_{u'=1}^{k} \prod_{\substack{u=1\\ u\neq u'}}^{k} p(\boldsymbol{\delta}_{u'u}).\end{aligned}$$

The distribution of the set of latent variables U_{jt}, collected in matrix $\boldsymbol{U}$, cannot be computed exactly. For computational convenience this distribution is substituted with

$$\tilde{p}(\boldsymbol{U}|\boldsymbol{\Gamma}, \boldsymbol{\Delta}) = \prod_{j=1}^{n} \left[\lambda(u_{j1}|\tilde{\boldsymbol{u}}_{j1}, \boldsymbol{\Gamma}) \prod_{t=2}^{T} \pi(u_{jt}|u_{j,t-1}, \tilde{\boldsymbol{u}}_{jt}, \boldsymbol{\Delta})\right], \tag{3}$$

which is a pseudo-probability in the sense of Besag (1975). For a discussion about this approximation for the case of cross-sectional spatial data see, among others, Spezia et al. (2018). These authors also rely on a method to avoid the approximation of $p(\boldsymbol{U}|\boldsymbol{\Gamma}, \boldsymbol{\Delta})$ that is based on auxiliary Monte Carlo steps within the estimation algorithm. In the present context, which is more complex due to the presence of data with both a spatial and temporal dimension, we prefer to directly use pseudo-probability (3) also because the very large size of the sample we analyse requires to limit the computational burden of the estimation algorithm. Moreover, the use of pseudo-distributions in place of the corresponding true distributions is not uncommon in spatial statistics and we do not expect it to strongly affect the estimation results. See also Spezia et al. (2017) for other examples of the use of pseudo-probabilities in related spatial contexts and Friel and Pettitt (2004), Friel et al. (2009) and Everitt (2012) for a discussion about the implication of their use.

Finally, regarding the conditional distribution of the response variables y_{ijt} collected in the set $\mathcal{Y}$, we have

$$p(\mathcal{Y}|\boldsymbol{U}, \boldsymbol{B}) = \prod_{j=1}^{n} \prod_{t=1}^{T} p\,(\boldsymbol{y}_{jt}|u_{jt}, \boldsymbol{B}),$$
$$p(\boldsymbol{y}_{jt}|u_{jt}, \boldsymbol{B}) = \prod_{i=1}^{m_{jt}} q\,(u_{jt}, \boldsymbol{B})^{y_{ijt}}[1-q(u_{jt}, \boldsymbol{B})]^{1-y_{ijt}},$$

where $\boldsymbol{y}_{jt}$ is the vector of the m_{jt} individual outcomes y_{ijt} at time t in site j.

4.1 | Markov chain Monte Carlo algorithm

In order to approximate the posterior distribution of model parameters, we rely on an MCMC algorithm, with an augmented parameter space. The algorithm is based on repeating a series of single steps for a suitable number of iterations R. The single steps of the proposed algorithm are described in the following.

- **Update of the $\boldsymbol{\beta}_u$ parameters:** For $u = 1, \ldots, k$, we propose $\boldsymbol{\beta}_u^*$ from distribution $N(\boldsymbol{\beta}_u, \tau_\beta^2\boldsymbol{I})$, namely a multivariate normal distribution centred on the current parameter vector and with variance depending on τ_β^2. The proposed parameter vector is accepted with probability

$$\alpha(\boldsymbol{\beta}_u^*, \boldsymbol{\beta}_u) = \min\left(1, \frac{p(\boldsymbol{\beta}_u^*)p(\mathcal{Y}|\boldsymbol{U}, \boldsymbol{B}_u^*)}{p(\boldsymbol{\beta}_u)p(\mathcal{Y}|\boldsymbol{U}, \boldsymbol{B})}\right),$$

  where $\boldsymbol{B}_u^*$ is obtained by substituting the uth column of the current matrix $\boldsymbol{B}$ with $\boldsymbol{\beta}_u^*$.
- **Update of the latent variables u_{jt}:** Using a Gibbs sampler, for $j = 1, \ldots, n$ and $t = 1, \ldots, T$, we draw u_{jt} from a categorical distribution with vector of probabilities $\boldsymbol{r}_{jt}$ obtained in a different way depending on the specific time occasion t. For the first time occasions, the elements of $\boldsymbol{r}_{j1}$ are

$$\frac{\lambda_j(u|\tilde{\boldsymbol{u}}_{j1}, \boldsymbol{\Gamma})s_{j1}(u|\boldsymbol{\Gamma})\pi_{j2}(u_{j2}|u, \tilde{\boldsymbol{u}}_{j2}, \boldsymbol{\Delta})p(\boldsymbol{y}_{j1}|u_{j1} = u, \boldsymbol{B})}{\sum_{v=1}^{k} \lambda_j(v|\tilde{\boldsymbol{u}}_{j1}, \boldsymbol{\Gamma})s_{j1}(v|\boldsymbol{\Gamma})\pi_{j2}(u_{j2}|v, \tilde{\boldsymbol{u}}_{j2}, \boldsymbol{\Delta})p(\boldsymbol{y}_{j1}|u_{j1} = v, \boldsymbol{B})}, \quad u = 1, \ldots, k,$$

with $s_{j1}(u|\boldsymbol{\Gamma}) = \prod_{\substack{j'=1 \\ c_{jj'}=1}}^{k} \lambda_{j'}(u_{j't}|\tilde{\boldsymbol{u}}_{j't}(u), \boldsymbol{\Gamma})$, where the product is extended to all units j' in the neighbourhood of j (so that $c_{jj'} = 1$); for time occasion $t = 2, \ldots, T-1$, the probability vector has elements

$$\frac{\pi_{jt}(u|u_{j,t-1}, \tilde{\boldsymbol{u}}_{jt}|\boldsymbol{\Delta})s_{jt}(u|\boldsymbol{\Delta})\pi_{j,t+1}(u_{j,t+1}|u, \tilde{\boldsymbol{u}}_{j,t+1}, \boldsymbol{\Delta})p(\boldsymbol{y}_{jt}|u_{jt}=u, \boldsymbol{B})}{\sum_{v=1}^{k} \pi_{jt}(v|u_{j,t-1}, \tilde{\boldsymbol{u}}_{jt}|\boldsymbol{\Delta})s_{jt}(v|\boldsymbol{\Delta})\pi_{j,t+1}(u_{j,t+1}|v, \tilde{\boldsymbol{u}}_{j,t+1}, \boldsymbol{\Delta})p(\boldsymbol{y}_{jt}|u_{jt}=v, \boldsymbol{B})}, \quad u=1, \ldots, k,$$

with $s_{jt}(u|\boldsymbol{\Delta}) = \prod_{\substack{j'=1 \\ c_{jj'}=1}}^{k} \pi_{j't}(u_{j't}|u_{j',t-1}, \tilde{\boldsymbol{u}}_{j't}(u), \boldsymbol{\Delta})$, which is defined for t greater than 1; for the last time occasion the elements of $\boldsymbol{r}_{jT}$ are

$$\frac{\pi_{jT}(u|u_{j,T-1}, \tilde{\boldsymbol{u}}_{jT}, \boldsymbol{\Delta})s_{jT}(u|\boldsymbol{\Delta})p(\boldsymbol{y}_{jT}|u_{jT}=u, \boldsymbol{B})}{\sum_{v=1}^{k} \pi_{jT}(v|u_{j,T-1}, \tilde{\boldsymbol{u}}_{jT}, \boldsymbol{\Delta})s_{jT}(v|\boldsymbol{\Delta})p(\boldsymbol{y}_{jT}|u_{jT}=v, \boldsymbol{B})}, \quad u=1, \ldots, k.$$

We also consider a version of the algorithm in which components $s_{j1}(u|\boldsymbol{\Gamma})$ and $s_{jt}(u|\boldsymbol{\Delta})$ are omitted from the previous updating rules; in fact, we experimented that the omission of these quantities does not significanty affect the estimation results, while making the estimation algorithm much faster.

- **Update of the $\boldsymbol{\gamma}_u$ parameters:** for $u = 2, \ldots, k$ we propose a new parameter vector $\boldsymbol{\gamma}_u^*$ which is accepted with probability

$$\alpha(\boldsymbol{\gamma}_u, \boldsymbol{\gamma}_u^*) = \min\left(1, \frac{p(\boldsymbol{\gamma}_u^*)\prod_{j=1}^{n} \lambda_j(u_{j1}|\tilde{\boldsymbol{u}}_{j1}, \boldsymbol{\Gamma}_u^*)}{p(\boldsymbol{\gamma}_u)\prod_{j=1}^{n} \lambda_j(u_{j1}|\tilde{\boldsymbol{u}}_{j1}, \boldsymbol{\Gamma}_u)}\right),$$

where $\boldsymbol{\Gamma}_u^*$ is the current matrix $\boldsymbol{\Gamma}$ with the uth row substituted by $\boldsymbol{\gamma}_u^*$.

- **Update of the $\boldsymbol{\delta}_{u'u}$ parameters:** for $u', u = 1, \ldots, k$, with $u \neq u'$, we propose a new parameter vector $\boldsymbol{\delta}_{u'u}$, denoted by $\boldsymbol{\delta}_{u'u}^*$, which is accepted with probability

$$\alpha(\boldsymbol{\delta}_{u'u}, \boldsymbol{\delta}_{u'u}^*) = \min\left(1, \frac{p(\boldsymbol{\delta}_{u'u}^*)\prod_{j=1}^{n}\prod_{t=2}^{T} \pi_{jt}(u_{jt}|u_{j,t-1}, \tilde{\boldsymbol{u}}_{jt}, \boldsymbol{\Delta}_{u'u}^*)}{p(\boldsymbol{\delta}_{u'u})\prod_{j=1}^{n}\prod_{t=2}^{T} \pi_{jt}(u_{jt}|u_{j,t-1}, \tilde{\boldsymbol{u}}_{jt}, \boldsymbol{\Delta}_{u'u})}\right),$$

where $\boldsymbol{\Delta}_{u'u}^*$ is the matrix $\boldsymbol{\Delta}$ with parameter vector $\boldsymbol{\delta}_{u'u}$ substituted by $\boldsymbol{\delta}_{u'u}^*$. The products in the previous expression may be restricted to the only cases in which $u_{j,t-1} = u'$.

At the end of each of the R iterations of the algorithm based on the single steps reported above, we obtain values of the parameters and latent variables drawn from the posterior distribution in Equation (2), which are denoted by $\boldsymbol{B}^{(r)}$, $\boldsymbol{\Gamma}^{(r)}$, $\boldsymbol{\Delta}^{(r)}$ and $\boldsymbol{U}^{(r)}$, $r = 1, \ldots, R$. A similar notation is used for each element of these matrices, while we use the short-hand notation $\boldsymbol{\theta}^{(r)}$ for the overall set of parameters and latent variables obtained at the end of iteration r.

The MCMC output may be elaborated in the usual way to obtain point estimates, standard errors and credible sets for the model parameters. To perform local decoding, that is, to predict the latent state of a certain area in a certain year, we adopt a maximum a posterior (MAP) rule; the predicted latent state is

$$\hat{u}_{jt} = \underset{u=1,\ldots,k}{\text{argmax}} \sum_{r=1}^{R} I(u_{jt}^{(r)} = u),$$

where $I(\cdot)$ denotes the indicator function. Moreover, to obtain a prediction for a specific area and year, corresponding in our application to the prevalence of food insecurity, we compute posterior averages

$$\hat{q}_{jt}(u) = \frac{1}{m_{jt}R}\sum_{i=1}^{m_{jt}}\sum_{r=1}^{R}\frac{\exp(\boldsymbol{x}'_{ijt}\boldsymbol{\beta}^{(r)}_{u^{(r)}_{jt}})}{1+\exp(\boldsymbol{x}'_{ijt}\boldsymbol{\beta}^{(r)}_{u^{(r)}_{jt}})},$$

which directly derive from Equation (1). Note that computing this quantity is only possible for areas with at least one sampled individual. For areas without any sampled individual at a certain time occasion, we sample individuals from neighbouring areas. In our implementation we sample as many individuals as the (rounded) average number of individuals observed in all of the neighbouring areas. In the few cases in which this number is also zero, we sample from individuals observed in the same area at the closest measurement occasion, breaking ties uniformly at random. Note that imputation is not needed for decoding, namely for the prediction of the latent states for each site and year, because site-specific covariates are usually known and/or time fixed. This is true for our specific application.

4.2 | Label switching

Given the model assumptions, a label switching problem (Stephens, 2000) clearly arises. We deal with this problem through a post-processing algorithm, along the same lines as Marin et al. (2005). Post-processing is based on finding the value of the parameters corresponding to the highest posterior density across all the MCMC iterations. This amounts to spotting the set of parameters among $\boldsymbol{\theta}^{(r)}$, $r = 1, \ldots, R$, that has the largest value of

$$LP^{(r)} = p(\boldsymbol{B}^{(r)})p(\boldsymbol{\Gamma}^{(r)})p(\boldsymbol{\Delta}^{(r)})\tilde{p}(\boldsymbol{U}^{(r)}|\boldsymbol{\Gamma}^{(r)},\boldsymbol{\Delta}^{(r)})p(\mathcal{Y}|\boldsymbol{U}^{(r)},\boldsymbol{B}^{(r)}), \quad (4)$$

which is the product between prior and likelihood and is an approximation of the right-hand side of expression (2). The corresponding matrix containing the $\boldsymbol{\beta}_u$ parameter vectors is denoted by $\hat{\boldsymbol{B}}$. Then, the value of the parameters and latent variables of each iteration r are re-examined by considering each possible permutation of the rows of $\boldsymbol{B}^{(r)}$ and the latent states are ordered so as to minimize the Euclidean distance between $\boldsymbol{B}^{(r)}$ and $\hat{\boldsymbol{B}}$.

4.3 | Model choice

For model choice we rely on the WAIC (Watanabe, 2010), in the version proposed by Vehtari et al. (2017), which does not need the model marginal likelihood and requires a very limited computational burden in addition to that of the estimation algorithm, an aspect that is of great importance given the very large sample size. In particular, the method measures the predictive accuracy by estimating the expected log-pointwise predictive density for a new data set (elpd) by the difference

$$\widehat{\text{elpd}}_{\text{waic}} = \widehat{\text{lpd}} - \hat{p}_{\text{waic}}, \quad (5)$$

where, in our context,

$$\widehat{\text{lpd}} = \sum_{j=1}^{n} \sum_{t=1}^{T} \log \hat{E}(p(\boldsymbol{y}_{jt} | \boldsymbol{\theta})) \tag{6}$$

is an estimate of the log-pointwise predictive density and

$$\hat{p}_{\text{waic}} = \sum_{j=1}^{n} \sum_{t=1}^{T} \hat{V}(\log p(\boldsymbol{y}_{jt} | \boldsymbol{\theta})). \tag{7}$$

In the previous expressions, $\hat{E}(p(\boldsymbol{y}_{jt} | \boldsymbol{\theta}))$ is simply the average of $p(\boldsymbol{y}_{jt} | \boldsymbol{\theta}^{(r)})$ and $\hat{V}(\log p(\boldsymbol{y}_{jt} | \boldsymbol{\theta}))$ denotes the variance of $\log p(\boldsymbol{y}_{jt} | \boldsymbol{\theta}^{(r)})$ across the R MCMC iterations. An advantage of this approach is that it also allows us to take into account the uncertainty about $\widehat{\text{elpd}}_{\text{waic}}$ in model selection. Uncertainty is measured on the basis of the sample variance (across sample units) of the single addends in expressions (6) and (7). For ease of reference, below we indicate this quantity as the standard error of $\widehat{\text{elpd}}_{\text{waic}}$.

5 | DATA ANALYSIS

In this section we proceed with the analysis of the data described in Section 2. We specify a neighbouring structure based on the four closest areas to every area in the same country, assuming then a first-order dependence. Any two contiguous areas that are in different countries are not flagged as neighbours, even within the European Union. Moreover, the 3% of observations are removed overall, corresponding to sample units for which at least one covariate is missing. For these data we estimate the proposed model for $k = 1, \ldots, 8$. We let our sampler run for $R = 50{,}000$ iterations, with a burn-in of 10,000 and thinning of 50 iterations. Convergence diagnostics are satisfactory in all cases.

In Table 2 we report summary information that may be used for model selection in terms of number of latent states (k). The table displays the average of (4), that is, the average product between prior and likelihood computed in each sampled vector of parameters, denoted by *LP*. It also displays the value of $\widehat{\text{elpd}}_{\text{waic}}$, as defined in Equation (5), its corresponding standard

TABLE 2 Data analysis results for $k = 1, \ldots, 8$; in bold the results referred to the selected model

k	*LP*	$\widehat{\text{elpd}}_{\text{waic}}$	$se(\widehat{\text{elpd}}_{\text{waic}})$	**Difference**
1	−405441.12	−405410.31	406.41	−
2	−360211.34	−350875.67	454.61	54612.89
3	−364339.40	−339406.02	459.65	11540.84
4	−383557.70	−335242.58	461.72	4248.55
5	−398849.34	−333357.25	461.20	2045.17
6	−420178.60	−332006.83	462.17	1433.42
7	**−443475.53**	**−331686.58**	**461.63**	**408.68**
8	−471498.74	−331432.65	463.11	579.92

TABLE 3 Posterior means of the parameters in β_u and 95% credibility intervals when $k = 7$

State	Intercept	Age	Age^2	Gender
1	−3.268 (−3.362, −3.123)	0.025 (0.001, 0.053)	−0.017 (−0.032, −0.004)	0.065 (−0.004, 0.131)
2	−2.325 (−2.441, −2.137)	−0.108 (−0.13, −0.087)	−0.077 (−0.088, −0.068)	−0.164 (−0.221, −0.104)
3	−1.846 (−1.897, −1.782)	0.062 (0.047, 0.081)	−0.034 (−0.039, −0.029)	−0.06 (−0.106, −0.015)
4	−1.034 (−1.088, −0.92)	0.071 (0.062, 0.081)	−0.031 (−0.035, −0.027)	−0.08 (−0.107, −0.053)
5	−0.273 (−0.318, −0.137)	0.087 (0.078, 0.097)	−0.021 (−0.024, −0.017)	−0.09 (−0.116, −0.065)
6	0.473 (0.429, 0.56)	0.086 (0.076, 0.095)	−0.027 (−0.03, −0.021)	−0.07 (−0.096, −0.04)
7	1.301 (1.227, 1.371)	0.077 (0.061, 0.092)	−0.011 (−0.021, −0.003)	−0.052 (−0.109, 0.004)
State	**IncomeQ.1**	**IncomeQ.2**	**IncomeQ.4**	**IncomeQ.5**
1	0.899 (0.829, 0.985)	0.271 (0.192, 0.346)	−0.204 (−0.349, −0.07)	−0.435 (−0.562, −0.323)
2	1.260 (1.158, 1.35)	0.614 (0.52, 0.69)	−0.425 (−0.555, −0.276)	−0.968 (-1.095, −0.828)
3	0.927 (0.884, 0.996)	0.415 (0.371, 0.476)	−0.431 (−0.48, −0.38)	−0.85 (−0.898, −0.8)
4	0.842 (0.775, 0.896)	0.368 (0.325, 0.418)	−0.33 (−0.368, −0.286)	−0.703 (−0.75, −0.653)
5	0.565 (0.507, 0.604)	0.332 (0.295, 0.365)	−0.303 (−0.348, −0.26)	−0.756 (−0.834, −0.707)
6	0.467 (0.402, 0.51)	0.234 (0.193, 0.275)	−0.273 (−0.311, −0.232)	−0.784 (−0.821, −0.729)
7	0.222 (0.148, 0.364)	0.174 (0.086, 0.365)	−0.185 (−0.254, −0.015)	−0.603 (−0.669, −0.446)

deviation, denoted by $se(\widehat{\text{elpd}}_{\text{waic}})$, and for $k \geq 2$ the difference in terms of $\widehat{\text{elpd}}_{\text{waic}}$ between the model with k states and the previous one.

Based on the results in Table 2, and accounting for the standard error of WAIC, we select a model with $k = 7$ latent states. In order to measure the goodness-of-fit of the selected model, we consider the statistic equal to the squared difference between the observed values of the response variable and the predicted values in the spirit of Copas (1989). The maximum of this statistic equals the overall number of observations, in case of completely wrong predictions. A posterior predictive p-value for this statistic is computed on the basis of the parameter draws at every MCMC step as described, among others, in Gelman (2013). More precisely, at step r of the algorithm we compute the following quantity

$$D^{(r)} = \sum_{j=1}^{n} \sum_{t=1}^{T} \sum_{i=1}^{m_{jt}} [y_{ijt} - q(u_{jt}^{(r)}, \boldsymbol{B}^{(r)})]^2, \tag{8}$$

on the basis of the values of the latent variables and the regression parameters drawn at that step, and denoted by $u_{jt}^{(r)}$ and $\boldsymbol{B}^{(r)}$ respectively. At the same step the corresponding simulated quantity, denoted by $\hat{D}^{(r)}$, is computed according to (8) with y_{ijt} substituted by $\hat{y}_{ijt}$, which is drawn from a Bernoulli distribution with parameter $q(u_{jt}^{(r)}, \boldsymbol{B}^{(r)})$. Finally, the posterior predictive p-value is computed as the proportion of times that $\hat{D}^{(r)}$ is at least equal to $D^{(r)}$. For the data at hand, and for the selected model with $k = 7$ latent states, the average value of $D^{(r)}$ across the MCMC steps is equal to 107,013.00, to be compared with a maximum value of 729,441 and with a posterior p-value of 0.585. Being this value close to 0.5, we then conclude that the selected model has an adequate fit (Gelman, 2013).

For reason of space, in Table 3 we report only posterior means and 95% credibility intervals for the regression parameters. According to these results, male gender is protective in all areas, confirming a need for gender equality in access to food (e.g., Garcia & Wanner, 2017). Moreover, a clear negative effect of quantile of income is seen everywhere, and could obviously be expected. The protective effect of income is stronger in low prevalence areas, where access to food is probably difficult only for the very poor.

We also summarize the latent distribution through average initial and transition probabilities in Table 4. These are obtained by averaging the individual and time-specific parameters. We observe that while low prevalence areas form a majority, there are also several high prevalence

TABLE 4 Posterior means of marginal initial and transition probabilities for the latent process U_{jt} when $k = 7$

	1	2	3	4	5	6	7
	Initial probabilities						
	0.08	0.17	0.14	0.18	0.17	0.15	0.11
	Transition probabilities						
1	0.81	0.05	0.06	0.03	0.03	0.00	0.00
2	0.01	0.92	0.05	0.02	0.00	0.00	0.00
3	0.01	0.05	0.83	0.09	0.02	0.00	0.00
4	0.00	0.00	0.10	0.75	0.10	0.04	0.00
5	0.00	0.00	0.02	0.10	0.77	0.08	0.02
6	0.00	0.00	0.00	0.04	0.07	0.83	0.06
7	0.00	0.00	0.00	0.01	0.04	0.07	0.87

areas. Transitions among latent states are not uncommon but mostly to adjacent latent states. For instance, for an area in the high prevalence state 5 at a certain time point there is a 77% chance of persisting in the same latent state, 10% chance of moving to the better scenario represented by latent state 4, but it is very unlikely to jump to latent states 1 or 2. Moreover, the marginal transition matrix is not symmetric, with generally larger probabilities to move to higher propensity of food insecurity. This is an indication that the number of areas with higher propensity to food insecurity has increased over time.

The clustering capability of our approach can be validated by obtaining latent state profiles with respect to some endogenous and exogenous variables. We proceed to do so by grouping measurements according to the predicted latent state of an area at a given time point, and then taking descriptive statistics. A consequence is that areas in the same latent state at a given time point can be deemed to be similar with respect to food insecurity, tremendously reducing the complexity of the problem. For instance, a country like the United States of America, which is divided into 50 areas, is finally clustered into only three groups over the observation period. A similar phenomenon is observed for most of the countries considered. In Table 5 we report the average prevalence of food insecurity, its standard deviation, the average income and its standard deviation, the average household size, the proportion of households with at least one member employed full time, and the modal continent of areas in each latent state at a given year. It is clear that areas assigned to each latent state are different with respect to all variables considered.

We use Equation (1) to compute the state-subject-occasion specific probability of each subject to be food insecure. These values are then averaged over states using subject-occasion specific posterior probabilities of each latent state, and further averaged over area indicators to obtain an area-occasion specific prevalence of food insecurity. For imputation in cases in which $m_{jt} = 0$, we proceed as described at the end of Section 4.1. We note here that resulting prevalence estimates only minimally change if repeating the procedure, showing therefore little sensitivity to random sampling. Additionally, prevalence estimates agree also if random sampling is replaced with a mean imputation method in which a single covariate profile is fixed as the mean covariate profile from neighbouring areas at the same time occasion, and in the same area at the previous and following time occasions. A total of 2141×8 prevalence estimates are therefore finally obtained, and are shown in Figures 1 and 2. A similar strategy can be used for extrapolation, obtaining forecasts at time occasions that have not been sampled, before or after the observation period.

TABLE 5 Profiles of latent states obtained by averaging measurements over areas assigned to each latent state in a given period

Latent state	1	2	3	4	5	6	7
Prevalence	0.04	0.08	0.13	0.24	0.39	0.57	0.77
sd(Prevalence)	0.03	0.07	0.07	0.08	0.09	0.09	0.10
Income	20543.0	11046.9	9822.3	4998.2	3335.4	2088.9	1216.4
sd(Income)	12718.8	13743.1	8547.6	9344.6	8740.2	2578.4	1936.4
Av. household size	2.84	3.53	3.81	4.23	4.83	5.11	5.86
Employed Full time	0.60	0.52	0.50	0.46	0.37	0.29	0.17
Modal continent	Europe	Asia	Asia	Asia	Asia	Africa	Africa

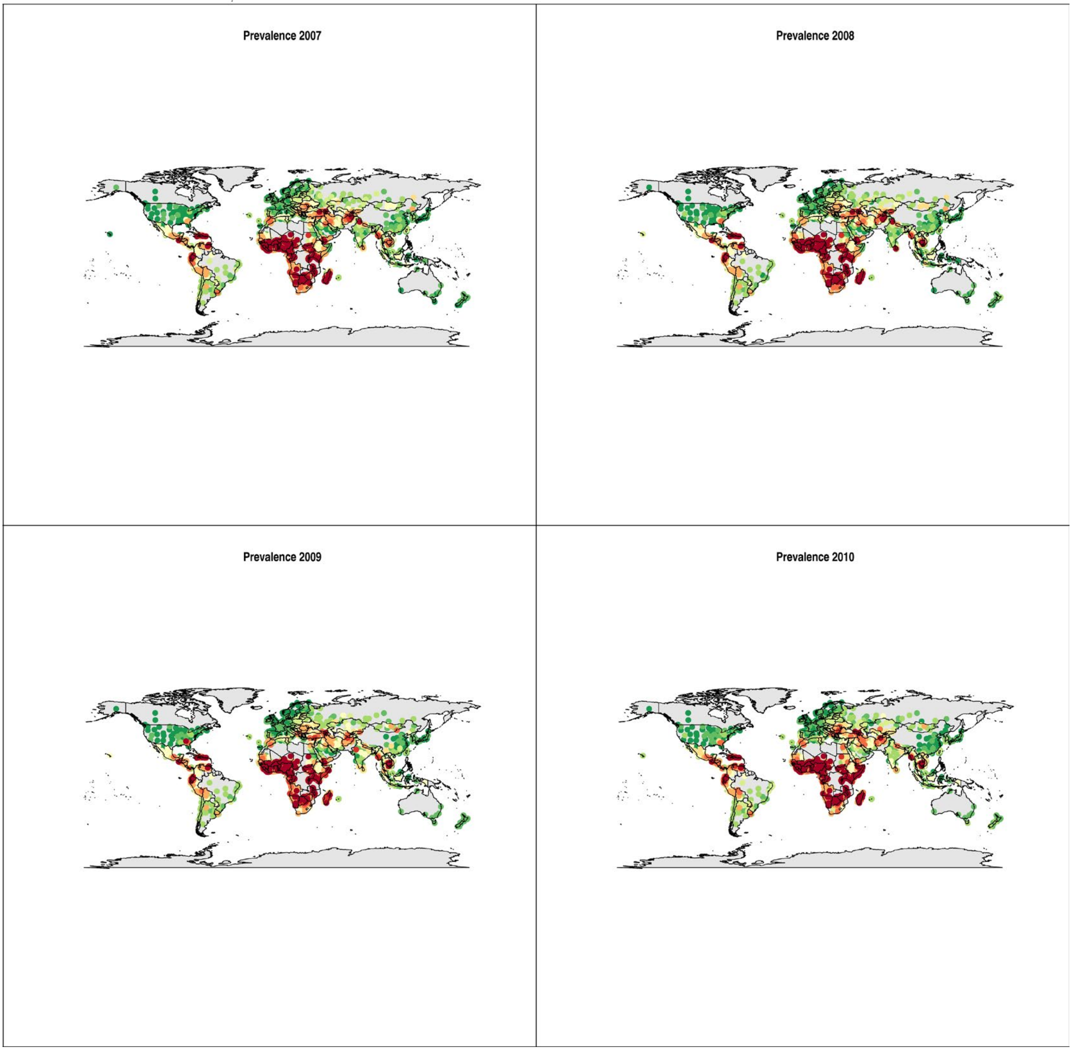

FIGURE 1 Predicted prevalence in 2007, 2008, 2009 and 2010. One dot per area is geolocalized through Google Maps. Each area is coloured from dark green (lowest predicted prevalence) to dark red (highest predicted prevalence)

Gallup uses in many cases non-standard area boundaries, and to the best of our knowledge raster data for these areas are not available; hence a map with area boundaries could not be plotted. We proceeded by mapping area names to a latitude and longitude through the Google Maps service. The coordinates given by Google do not necessarily point to a geographical centre of the area, but can be used to produce a spatial point pattern with dots on those coordinates as done in the figures. From Figures 1 and 2 a strong geographical pattern is observed, with some areas within countries that differ slightly from the other areas in the same country, especially in Asia and Latin America. Additionally, a clear effect of the 2008 economic crisis can be identified in several areas of the world and an interesting time pattern emerges for several areas (e.g. areas within Madagascar). Finally, to corroborate our results, we also plot clusters of areas as identified by the posterior expected latent states in Figures 3 and 4, where areas with the same colour (in the same and also in different years) belong to the same cluster of propensity to access to food.

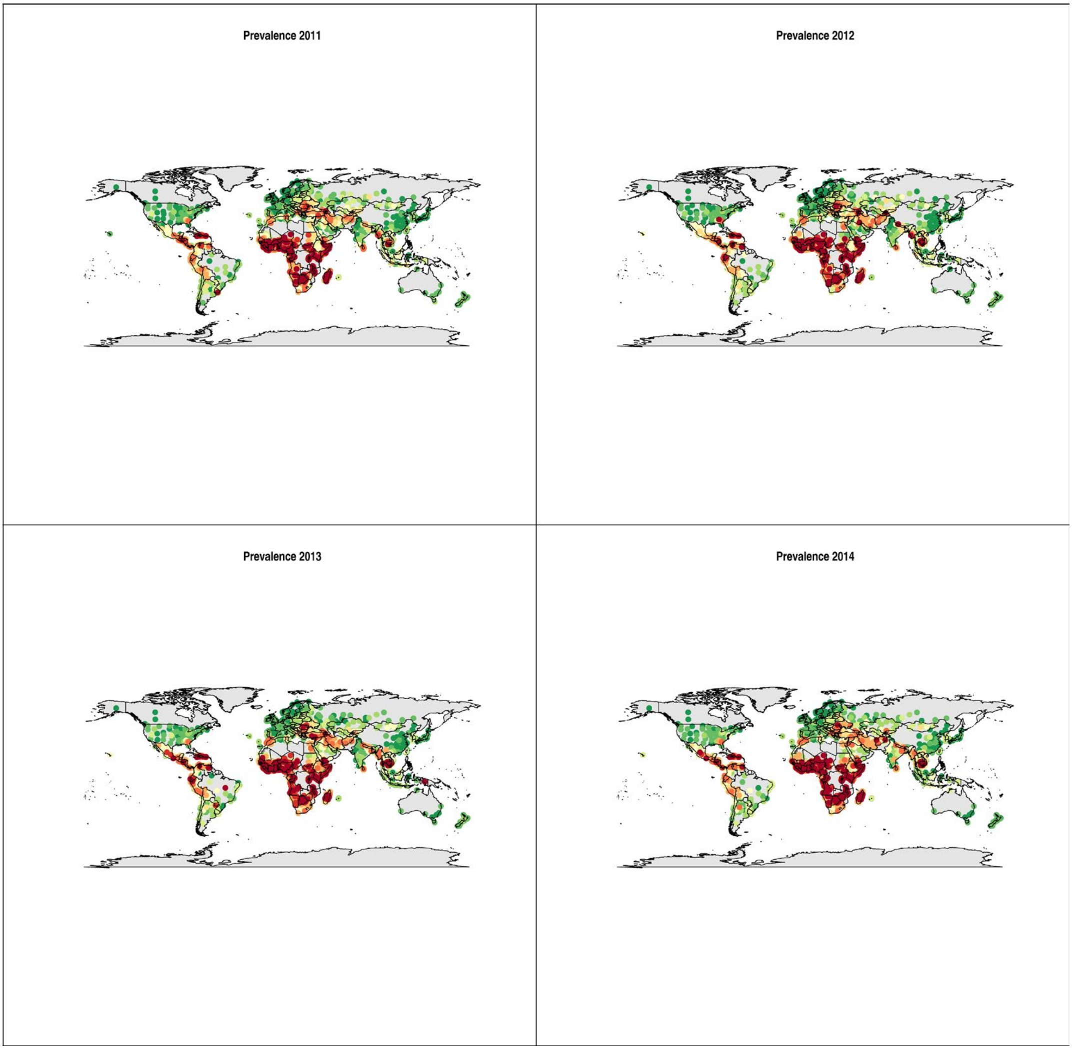

FIGURE 2 Predicted prevalence in 2011, 2012, 2013 and 2014. One dot per area is geolocalized through Google Maps. Each area is coloured from dark green (lowest predicted prevalence) to dark red (highest predicted prevalence)

In the Supplementary Material we give further insights into the properties of our model by showing the same maps obtained from raw data. A comparison clearly supports our claims about the smoothing effects of our methodology, and the advantages of pooling information over time and space. Indeed, maps of raw prevalence exhibit for some areas in some years quite extreme or counter-intuitive estimates, due to the high uncertainty associated with sampling a small number of subjects in an area.

To further illustrate, we restrict our attention to certain areas of interest, whose predicted prevalences are reported in Table 6. We consider Somalia's area of Gedo, where in mid-2008 the third phase of Somali's civil war started after withdrawal of Ethiopian troops. We observe a clear rise in the prevalence as a consequence of this event, with the area predicted to have switched from latent state 6 to 7 in 2009. In contrast, for a stable and rich area like Stockholm

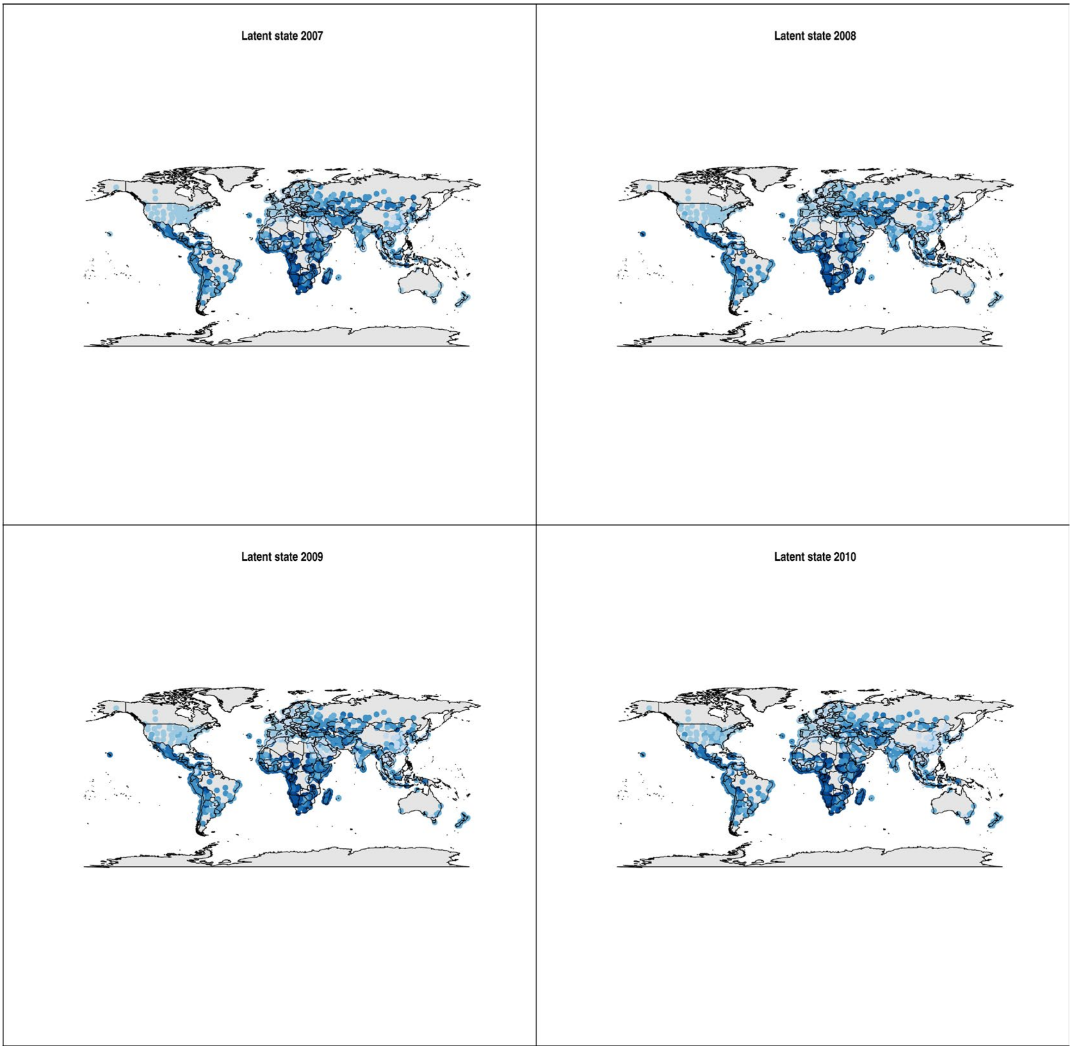

FIGURE 3 Predicted latent state in 2007, 2008, 2009 and 2010. One dot per area is geolocalized through Google Maps. Each area is coloured from light blue (first latent state) to dark blue (seventh latent state)

in Sweden, both prevalence and latent state are constant over the observation period. Clearly, even in Stockholm there might be suburbs of poor people who struggle to obtain food, but unfortunately we do not have information at suburb level. Furthermore, the prevalence in the area is clearly, comparatively, low and stable. A latent transition instead occurred in the area of Minsk, in Belarus, in 2013. This is probably linked to the fact that GDP in Belarus jumped from 65.69$ billion to 75.53$ billion. Finally, we consider Bhutan's regions of the capital Thimphu and Monggar. Bhutan quickly raised from one of the most poor countries in the world to a safe and healthy place, whose inhabitants are happy and well fed. This was an explicit political struggle, which is well known to have been successful. It can be seen from the results in Table 6 that, while this transition probably applies to the entire country, it might not have occurred simultaneously in all areas.

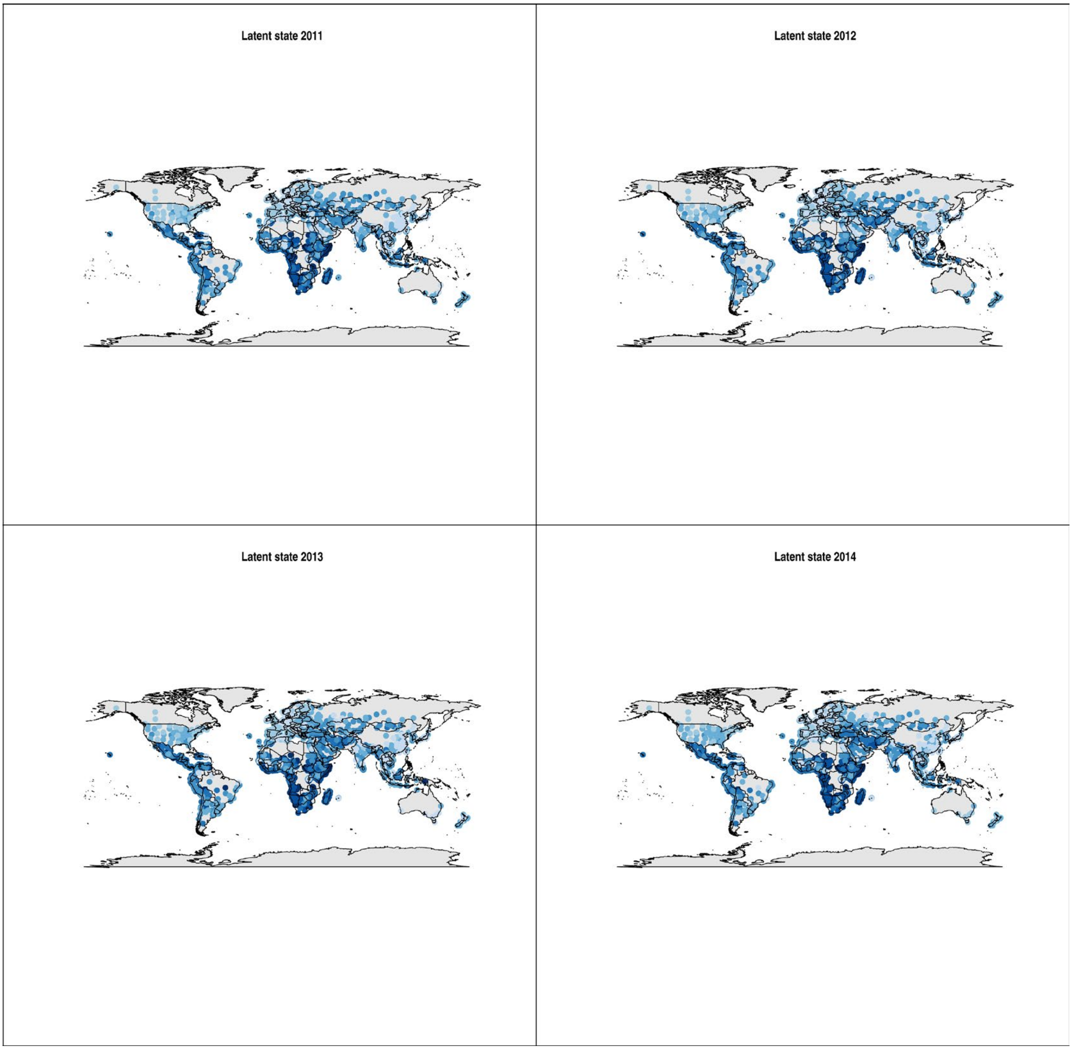

FIGURE 4 Predicted latent state in 2011, 2012, 2013 and 2014. One dot per area is geolocalized through Google Maps. Each area is coloured from light blue (first latent state) to dark blue (seventh latent state)

TABLE 6 Predicted prevalence and latent state (in parentheses) for selected areas

Area	2007	2008	2009	2010	2011	2012	2013	2014
Gedo (Somalia)	0.51 (6)	0.51 (6)	0.72 (7)	0.75 (7)	0.71 (7)	0.71 (7)	0.72 (7)	0.73 (7)
Stockholm (Sweden)	0.01 (1)	0.01 (1)	0.01 (1)	0.01 (1)	0.01 (1)	0.01 (1)	0.01 (1)	0.01 (1)
Minsk (Belarus)	0.20 (4)	0.19 (4)	0.20 (4)	0.20 (4)	0.20 (4)	0.20 (4)	0.09 (3)	0.10 (3)
Thimphu (Bhutan)	0.04 (1)	0.03 (1)	0.04 (1)	0.04 (1)	0.03 (1)	0.04(1)	0.03 (1)	0.03 (1)
Monggar (Bhutan)	0.11 (3)	0.10 (3)	0.10 (3)	0.12 (3)	0.10 (3)	0.10 (3)	0.15 (3)	0.05 (1)

6 | CONCLUSIONS

We have presented a flexible spatio-temporal model which is based on certain conditional independence assumptions. On one hand, the use of a discrete latent variable allows us to model unobserved heterogeneity with reduced parametric assumptions, as even continuous mixing distributions can be expected to be well approximated by a discrete distribution (see, for instance, Bartolucci & Farcomeni, 2009, on this point). On the other hand, it is possible to cluster several (here, 2141×8) measurements into a few (here, $k = 7$) groups. Our model provides also a comprehensive approach to assess relationships between access to food and certain predictors at global level. Time dynamics are also easily evaluated. The most useful feature of our approach is the fact that we can obtain prevalence estimates at the area level by pooling information over neighbours and time points. This is especially useful for areas without any sampled subjects at a certain wave, and in general it can be expected to reduce the mean squared error of predictions. Indeed, while observed prevalence at area level can exhibit patterns that might only be due to sampling variability, our predictions are often stable and interpretable.

To the best of our knowledge, prevalence estimates are routinely computed only at country and region (e.g. continent) level, and this is the first study to present reliable estimates at a smaller spatial scale. These estimates are clearly useful for the development of appropriate strategies to fight poverty. Given the fact that high variability can be expected for certain countries, we are confident that the lower spatial aggregation we provide can be very useful for policy makers and program evaluation. The complete results are available upon request. It shall be mentioned, as a further evidence, that a strong agreement can be found between our prevalence estimates (at national level) and the official prevalence estimate for food insecurity officially published in FAO et al. (2019).

Finally, for simplicity in this presentation we have ignored sampling weights computed within the GWP poll. Indeed, we have tried also fitting simplified versions of the proposed model in which sampling weights were taken into account. Results were very similar, with the model with uniform weights being more stable in terms of convergence of the MCMC. It shall be noted that this does not modify our results in any substantial respect. Moreover, given the very large sample size and the complexity of the data structure, which has both a spatial and a temporal dimension, we adopt an estimation algorithm that relies on an approximation of the distribution of the latent variables given the model parameters that we do not expect to strongly affect results. This approximation, which relies on the use of a pseudo-probability rather than the true probability of the latent variables, is not uncommon in spatial statistics. Nevertheless, further research could be devoted to a methodological study of the impact of this approximation, which is expected to also affect the performance of the Watanabe–Akaike information criterion (Watanabe, 2010) that we use for model selection, and it is a way to by-pass cumbersome computation of the marginal likelihood. A simulation study of the behaviour of this selection criterion specifically for the model proposed in this paper could be also of interest. Possible extensions also include the case of time-varying number of latent states as in Anderson et al. (2019), and count data as in Bartolucci and Farcomeni (2021).

ACKNOWLEDGEMENTS

We are grateful to two referees for constructive suggestions; and to Gallup, Inc. and the Voices of the Hungry project, FAO, for access to the GWP/FIES data.

ORCID

Francesco Bartolucci https://orcid.org/0000-0001-7057-1421
Alessio Farcomeni https://orcid.org/0000-0002-7104-5826

REFERENCES

Ailliot, P., Thompson, C. & Thomson, P. (2009) Space–time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions. *Journal of the Royal Statistical Society: Series C*, 58, 405–426.

Anderson, G., Farcomeni, A., Pittau, M.G. & Zelli, R. (2019) Rectangular latent Markov models for time-specific clustering, with an analysis of the well being of nations. *Journal of the Royal Statistical Society: Series C*, 68, 603–621.

Bartolucci F., & Farcomeni A. (2021) A spatio-temporal model based on discrete latent variables for the analysis of COVID-19 incidence. *Spatial Statistics*, 100504. https://doi.org/10.1016/j.spasta.2021.100504

Bartolucci, F. & Farcomeni, A. (2009) A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, 104, 816–831.

Bartolucci, F. & Farcomeni, A. (2015) A discrete time event-history approach to informative drop-out in mixed latent Markov models with covariates. *Biometrics* 71, 80–89.

Bartolucci, F., Farcomeni, A. & Pennoni, F. (2013) *Latent Markov models for longitudinal data*. Boca Raton, FL: Chapman & Hall/CRC Press.

Bartolucci, F., Farcomeni, A. & Pennoni, F. (2014) Latent Markov models: A review of a general framework for the analysis of longitudinal data with covariates. *TEST*, 23, 433–465.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36, 192–225.

Besag, J. (1975) Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D*, 24, 179–195.

Besag, J. (1986) On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society: Series B*, 48, 259–279.

Best, N., Richardson, S. & Thomson, A. (2005) A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*, 14, 35–59.

Bhattacharya, J., Currie, J. & Haider, S. (2004) Poverty, food insecurity, and nutritional outcomes in children and adults. *Journal of Health Economics*, 23, 839–862.

Copas, J. (1989) Unweighted sum of squares test for proportions. *Journal of the Royal Statistical Society: Series C*, 38 , 71–80.

Deaton, A. (2008) Income, health, and well-being around the world: Evidence from the Gallup world poll. *Journal of Economic Perspectives* 22, 53–72.

Dotto, F., Farcomeni, A., Pittau, M.G. & Zelli, R. (2019) A dynamic inhomogeneous latent state model for measuring material deprivation. *Journal of the Royal Statistical Society: Series A*, 182, 495–516.

Everitt, R.G. (2012) Bayesian parameter estimation for latent Markov random fields and social networks. *Journal of Computational and graphical Statistics*, 21, 940–960.

FAO, IFAD, UNICEF, WFP, & WHO. (2019) The state of food security and nutrition in the world 2018. Building climate resilience for food security and nutrition. Technical report, Food and Agriculture Organization of the United Nations, Rome.

Friel, N. & Pettitt, A. (2004) Likelihood estimation and inference for the autologistic model. *Journal of Computational and Graphical Statistics*, 13, 232–246.

Friel, N., Pettitt, A., Reeves, R. & Wit, E. (2009) Bayesian inference in hidden Markov random fields for binary data defined on large lattices. *Journal of Computational and Graphical Statistics*, 18, 243–261.

Frongillo, E.A., Nguyen, H.T., Smith, M.D. & Coleman-Jensen, A. (2017) Food insecurity is associated with subjective well-being among individuals from 138 countries in the 2014 Gallup world poll. *The Journal of Nutrition*, 147, 680–687.

Gallup organization (2020) Gallup data set details. Available from https://www.gallup.com/services/177797/country-data-set-details.aspx [Accessed 2nd February 2021].

Garcia, A. & Wanner, T. (2017) Gender inequality and food security: Lessons from the genderresponsive work of the international Food Policy Research Institute and the Bill and Melinda Gates foundation. *Food Security*, 9, 1091–1103.

Gelman, A. (2013) Two simple examples for understanding posterior *p*-values whose distributions are far from uniform. *Electronic Journal of Statistics*, 7, 2595–2602.

Green, P.J. & Richardson, S. (2002) Hidden Markov models and disease mapping. *Journal of the American Statistical Association*, 97, 1055–1070.

Lawson, A.B. (2013) *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology*, 2nd edn. Boca Raton, FL: Chapman and Hall/CRC.

Li Donni, P. & Marino, M. (2018) Patterns of poverty among elderly Americans: A latent class Markov model. *Applied Economics Letters*, 25, 791–795.

Lin, Z., Sanders, S.J., Li, M., Sestan, N., State, M.W. & Zhao, H. (2015) A Markov random field-based approach to characterizing human brain development using spatial-temporal transcriptome data. *Annals of Applied Statistics*, 9, 429–451.

Mahadevan, R. & Hoang, V. (2016) Is there a link between poverty and food security? *Social Indicators Research*, 128, 179–199.

Marin, J.M., Mengersen, K.L. & Robert, C.P. (2005) Bayesian modelling and inference on mixture of distributions. In: Dey, D. & Rao, C. (Eds.) *Handbooks of statistics*, Volume 25. Amsterdam: Elsevier Science, pp. 459–507.

Marshall, R.J. (1991) A review of methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society: Series A*, 154, 421–441.

Nord, M., Hooper, M.D. & Hopwood, H. (2008) Household-level income-related food insecurity is less prevalent in Canada than in the United States. *Journal of Hunger & Environmental Nutrition*, 3, 17–35.

Powdthavee, N., Burkhauser, R. & De Neve, J.-E. (2017) Top incomes and human well-being: Evidence from the Gallup world poll. *Journal of Economic Psychology*, 62, 246–257.

Qian, W. & Titterington, D. (1991) Estimation of parameters in hidden Markov models. *Philosophical Transactions of the Royal Society of London: Series A*, 337, 407–428.

Rose, D. (1999) Economic determinants and dietary consequences of food insecurity in the United States. *The Journal of Nutrition*, 129, 517S–520S.

Smith, M.D., Rabbitt, M.P. & Coleman-Jensen, A. (2017) Who are the world's food insecure? New evidence from the Food and Agriculture Organization's Food Insecurity Experience Scale. *World Development*, 93, 402–412.

Spezia, L., Brewer, M.J. & Birkel, C. (2017) An anisotropic and inhomogeneous hidden Markov model for the classification of water quality spatio-temporal series on a national scale: the case of Scotland. *Environmetrics*, 28, e2427.

Spezia, L., Friel, N. & Gimona, A. (2018) Spatial hidden Markov models and species distributions. *Journal of Applied Statistics*, 45, 1595–1615.

Stephens, M. (2000) Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B*, 62, 795–809.

Suryanarayana, M.H. & Silva, D. (2007) Is targeting the poor a penalty on the food insecure? Povery and food insecurity in India. *Journal of Human Development*, 8, 89–107.

Tanner, M.A. & Wong, W.H. (1987) The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528–540.

Vehtari, A., Gelman, A. & Gabry, J. (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27, 1413–1432.

Watanabe, S. (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.

Wei, Z. & Li, H. (2008) A hidden spatial-temporal Markov random field model for network-based analysis of time course gene expression data. *Annals of Applied Statistics*, 2, 408–429.

Zucchini, W., MacDonald, I.L. & Langrock, R. (2017) *Hidden Markov models for time series: An introduction using R*. New York: Chapman and Hall/CRC.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Bartolucci F, Farcomeni A. A hidden Markov space–time model for mapping the dynamics of global access to food. *J R Stat Soc Series A*. 2021;00: 1–21. https://doi.org/10.1111/rssa.12746