

# Hidden Markov modelling of spatio-temporal dynamics of measles in 1750–1850 Finland

Tiia-Maria Pasanen, Jouni Helske & Tarmo Ketola

To cite this article: Tiia-Maria Pasanen, Jouni Helske & Tarmo Ketola (02 Mar 2026): Hidden Markov modelling of spatio-temporal dynamics of measles in 1750–1850 Finland, Journal of Applied Statistics, DOI: [10.1080/02664763.2026.2634794](https://doi.org/10.1080/02664763.2026.2634794)

To link to this article: <https://doi.org/10.1080/02664763.2026.2634794>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



View supplementary material [↗](#)



Published online: 02 Mar 2026.



Submit your article to this journal [↗](#)



Article views: 480



View related articles [↗](#)



View Crossmark data [↗](#)

# Hidden Markov modelling of spatio-temporal dynamics of measles in 1750–1850 Finland

Tiia-Maria Pasanen<sup>a</sup>, Jouni Helske<sup>a,b</sup> and Tarmo Ketola<sup>c,d</sup>

<sup>a</sup>Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland; <sup>b</sup>INVEST Research Flagship Centre, University of Turku, Turku, Finland; <sup>c</sup>Department of Forestry, University of Helsinki, Helsinki, Finland; <sup>d</sup>Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

## ABSTRACT

Real world spatio-temporal datasets, and phenomena related to them, are often challenging to visualise or gain a general overview of. In order to summarise information encompassed in such data, we combine two well known statistical modelling methods. To account for the spatial dimension, we use the intrinsic modification of the conditional autoregression, and incorporate it with the hidden Markov model, allowing the spatial patterns to vary over time. We apply our method to parish register data considering deaths caused by measles in Finland in 1750–1850, and gain novel insight of previously undiscovered infection dynamics. Five distinctive, reoccurring states, describing spatially and temporally differing infection burden and potential routes of spread, are identified. We also find that there is a change in the occurrences of the most typical spatial patterns circa 1812, possibly due to changes in communication networks after major administrative transformations in Finland.

## ARTICLE HISTORY

Received 8 November 2024  
Accepted 6 February 2026



## KEYWORDS


Bayes; epidemiology; hidden Markov model; intrinsic conditional autoregression; measles; spatio-temporal

## 1. Introduction

Describing, visualising, and modelling spatio-temporal panel data is often challenging, particularly in high-dimensional settings with numerous time points and spatial areas. In this paper, we propose a Bayesian hidden Markov model with a state-dependent spatial correlation structure as an unsupervised learning method for decomposing the spatio-temporal variation in the data into relatively few interpretable latent components. We apply our approach by modelling the spatio-temporal occurrence of measles in Finland during the 18th and 19th centuries. We also explore how the large administrative changes – which Finnish society went through due to the annexation of Finland to Russia after the war between Sweden and Russia in 1809 – altered the measles epidemics.

Wars, trade network changes, administrative border changes, and associated human movements are major upheavals in transmission networks of epidemics by affecting the connections between people. Human-to-human transmitted infections depend on the

**CONTACT** Tiia-Maria Pasanen  tiia-maria.h.pasanen@jyu.fi  Department of Mathematics and Statistics, University of Jyväskylä, Seminaarinkatu 15, 40100 Jyväskylä, Finland

 Supplemental data for this article can be accessed online at <http://dx.doi.org/10.1080/02664763.2026.2634794>.

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

human networks, and closely connected populations are expected to indicate more similar infection patterns [22,28]. Therefore, new borders, on the one hand, and opened new connections, on the other, affect the spread of epidemics in the population. Vivid examples of such cases can be found in contemporary data showing, for example, the connection between the opening of trade and the spread of respiratory illnesses, and between increased mobility of people and the spread of HIV epidemics [25,31]. Older data provide evidence, for example, of the effect of trade centres on historical plague epidemics [51].

Throughout the history of mankind, armed conflicts and border changes between countries have been common, affecting the common welfare of people. For centuries, Finland had been part of Sweden, but in 1721, Russia annexed the southeastern parts of Finland, or Old Finland. After the Finnish War in 1808–1809, the rest of Finland became an autonomous grand duchy of Russia. In 1812, Old Finland was united with the new autonomous Finland, and the capital of Finland was moved from Turku on the southwestern coast to Helsinki on the southern coast, significantly affecting markets and trade connections [30]. Being a remarkable transformation, the administrative change, however, left also many things unchanged. For example, the demographic records collected during the Swedish administration were still maintained in a similar way as before, due to which we have the data analysed here.

Finland has a long history in collecting demographic and health-related data, with local parishes systematically maintaining records of baptisms, burials, and causes of death from 1749 onwards. Even older data exist, but from this point on, the systematic data collection was mandatory, though not all collected data have survived to date [33]. Although death diagnostics were symptom-based and determined by priests [49, e.g.], the distinctive features of certain diseases allowed for relatively accurate diagnoses. One of such infections is measles, which we focus on. According to our data, measles accounted for approximately 2% of all deaths during our study period from 1750 to 1850. The population of Finland was roughly 500,000 in 1750, and grew to over 1.6 million by 1850 [44,48]. However, the healthcare system was primitive and virtually non-existent before the 19th century, lagging behind in development compared to the other Nordic countries and Europe until the 20th century [40,41].

Small population size coupled with hundreds of sparsely spaced small towns makes it difficult for epidemics to remain fully endemic [21]. In such conditions of Finland, it is likely that fadeouts and reintroductions dominated the infection dynamics. This implies that changes in the transmission networks – routes and sources – could have strongly affected the infection dynamics, especially considering all the administrative changes affecting Finland during the era. These factors contributing to the sensitivity to transmission – small population, sparsely spaced towns, and low incidence – together with a large number of missing cases also make the Finnish data challenging to analyse.

Conditional autoregressive models (CAR) [7] and their variants, such as intrinsic conditional autoregressive models (ICAR), are frequently employed to describe areal data. Although it is possible to extend CAR models to spatio-temporal settings [13], these interaction models can become computationally intensive with high-dimensional data in which spatial and temporal effects are not separable [23]. Approximate methods such as INLA [38] are computationally efficient alternatives to asymptotically exact Bayesian estimation via Markov chain Monte Carlo (MCMC), but these methods typically impose some restrictions on the model structure and the number of hyperparameters [2]. The potential bias

of these approximations is also challenging to quantify in specific cases, although in many typical settings the approximation bias has been found to be negligible [29,39]. However, visualising and interpreting the results of these interaction models may be complex: with time-varying spatial effects and a large number of time points, a substantial number of static maps are required to convey information about the estimated (or observed) spatial dynamics. This is at odds with our aim to communicate the main features and dominant patterns of the data in a compact and accessible form.

In life course research, hidden Markov models have been suggested as a probabilistic method to condense information of complex, multidomain life courses into a few interpretable latent states [17,18]. In these applications, given a common data generating process, individuals are assumed to be independent, having their own latent trajectories. In contrast, in spatio-temporal panel data, individuals share a spatial dependency structure which should be accounted for.

Hidden Markov models have been used in epidemiological research also before. In Amorós *et al.* [1], hidden Markov models are utilised in a spatio-temporal setting to detect influenza outbreaks, and in Knorr-Held and Richardson [24], they are used to identify meningococcal disease incidences. Both studies use two states to identify the endemic and epidemic, or endemic and hyperendemic, phases of the infections but do not aim to deeply analyse the underlying dynamics of the diseases. Amorós *et al.* [1] include the spatial dependence in the epidemic state as temporally independent ICAR components and omit it in the endemic state, whereas Knorr-Held and Richardson [24] account for the spatial dependency as ICAR in both states but consider it as state and time-invariant. Employing more hidden states may aid in discovering latent phenomena related to the spatial and temporal patterns of epidemics. On the other hand, increasing the number of states increases computational demands, and eventually identifiability and interpretability issues, both in Bayesian and maximum likelihood settings. These discrete state hidden Markov models also fall outside the class of models supported by INLA framework.

By combining the hidden Markov model and the intrinsic conditional autoregression, we are able to acknowledge the spatial and temporal dependencies simultaneously and identify five easily interpretable disease states that altered in Finland in 1750–1850. We gain new insights into the Finnish disease and communication history and the behaviour of measles epidemics in an environment with a limited healthcare and sanitation system. Our study also demonstrates how sparse and scarce data with significant proportion of missing information can be analysed by combining well established methods in a novel way. The model can be applied to other diseases and datasets as well, possibly offering a new tool to learn from corresponding developing areas of modern world.<sup>1</sup>

## 2. Data and methods

In this section, first, we describe the data and their background in a historical and Finnish context. The rest of the chapter progresses from presenting the statistical components needed for our approach to presenting the main model itself. Subsection 2.2 introduces the hidden Markov models and Subsection 2.3 the intrinsic conditional autoregression. Subsection 2.4 represents in detail the main model and the necessary technical assumptions to estimate it, recalling the connection between the practice, given our application, and the statistical theory.

## 2.1. Data

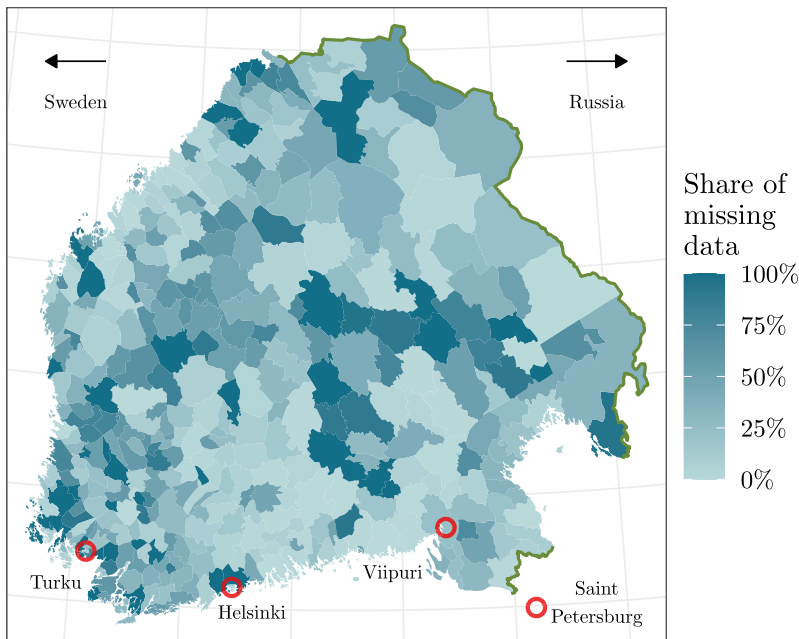
We analyse deaths from measles recorded in the parish registers, using monthly data from January 1750 to December 1850 from 387 different towns located in the southern half of Finland. The original records have count information on the deaths at the town level, but since the counts are generally low (99% of all measles counts in our data are three or fewer) and there are no town-level population sizes available – only nationwide estimates [48] – it is challenging to separate the actual differences between towns from those resulting from the varying population sizes.

To overcome these issues, we summarise the data into a more robust form. The observations we use are dichotomous, indicating whether at least one death caused by measles was recorded in each town. Thus, the data contains zeros for towns and months with no recorded deaths, and ones for those with recorded cases. Corresponding data, covering a different time span, has been used in Pasanen *et al.* [32], which also describes the handling of the data in more detail. Other studies analysing data based on the same registers but covering different areas or time periods, or being aggregated to a different level, include, for example, Ketola *et al.* [22] and Briga *et al.* [10,11]. Pasanen *et al.* [32] also implies that the dichotomisation of the observations results in conclusions corresponding to those of undichotomised data.

Although in general the cause of death records contain errors due to misdiagnoses, measles was likely diagnosed rather accurately because of its distinct visible symptoms. However, the dataset exhibits substantial missingness of records attributable to heterogeneity in record-keeping practices across towns [33] and to the destruction of records in fires and other past events. This kind of missingness is unlikely to depend on the outcome variable itself.

Additionally, there are months during which no deaths for any cause have occurred due to the small population sizes – there were simply too few people for deaths to eventuate monthly in each town. In our data, such month-town observations are always coded as missing ones. In reality, this leads to overestimating the share of missing data. It is more likely that a missing observation between two months with detected cases means no deaths at all than missingness because of record keeping or data collection processes. Such missingness is dependent on the outcome variable and should be taken into consideration. Since the Finnish population is smaller the further back in time we look, the proportion of month-town combinations without any deaths in the older data is larger as well. Nevertheless, as the overall missingness is greater in the older records, it also becomes harder to identify and separate the general missingness of the records from the absence of deaths. Therefore, we assume that before the first observation for each town, the records are missing independently of the outcome (lack of records), while we account for potentially non-random, state-dependent missingness after the first observation in our proposed model.

The study period covers 1212 months, each having observations from 75–272 towns per month, so no month has complete information from all towns. Only one town has complete records, whereas 44 towns are missing all observations. The missingness is approximately 70% at the beginning of the study period, decreasing steadily to about 20% at the end of the time series, with an overall missingness rate approximately 40%. However, when considering only data after the first observation in a town, the missingness



**Figure 1.** The percentage of all missing data across the years 1750–1850 in the towns on the study area, covering roughly the southern half of Finland. The red circles highlight the old capital Turku (until 1812), the new capital Helsinki (since 1812), the major trade city Viipuri which was united with the rest of Finland in 1812, and the then capital of Russia, Saint Petersburg. The borderline coloured with green denotes land border whereas the rest of the study area is bordered with water. The arrows denote the directions of the locations of Sweden and Russia.

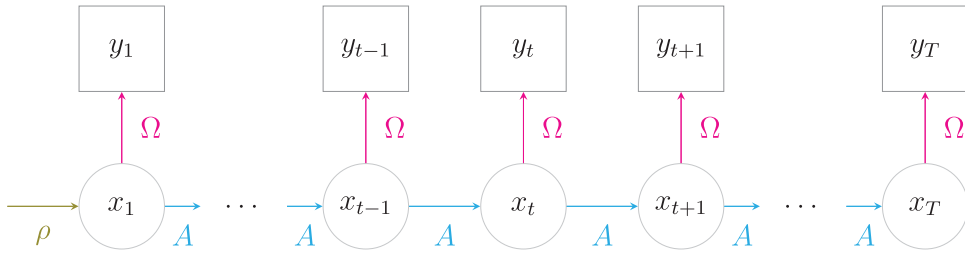
rate is about 30%. The spatial distribution of the overall missingness is illustrated in Figure 1.

In order to present our solution to analysing these data, next, we introduce the concepts of hidden Markov models and intrinsic conditional autoregression. The former compresses the time dependency into a relatively low number of nationwide discrete infection statuses, whereas the latter is responsible for considering the spatial connections.

## 2.2. Hidden Markov models

Discrete time hidden Markov models (HMM) [36] are statistical models for temporal, or other sequentially ordered data, consisting of two essential components. There are, firstly, the hidden stochastic Markov process describing the unobserved, yet often interesting, part of the model, and secondly, the observed process dependent on the hidden component. With the word ‘state’, often appearing in the terminology, we refer to some discrete, prevailing situations related to the phenomenon of interest. For example, there could be two states indicating the presence and absence of a disease, and the observed process would consist of the actual observations of the deaths caused by that disease. Next, we introduce the method in more detail.

The first component of the model is the hidden stochastic process  $x_t$ ,  $t = 1, \dots, T$ , with the state space  $\{1, \dots, S\}$ , i.e. having values from 1 to  $S$ . Usually the temporal dependency



**Figure 2.** Directed graph of a hidden Markov model. The state  $x_1$  in the first time point depends on the initial probabilities  $\rho$ , and the following transitions are determined by the transition probabilities in  $A$ . The observations  $y_t$  depend on the current state  $x_t$  via the emission probabilities in  $\Omega$ , and the current state  $x_t$  in turn depends on the previous state  $x_{t-1}$ .

of the process  $x_t$  encompasses only one previous time point so that it satisfies the first-order Markov property

$$p(x_t | x_1, \dots, x_{t-1}) = p(x_t | x_{t-1}). \quad (1)$$

The hidden process may also meet some other order Markov condition, meaning that the current state depends on as many preceding states as is the degree of the Markov chain, but we do not consider those cases here. For the states  $s \in \{1, \dots, S\}$  we use consecutive numbering as labels for simplicity. Between two preceding time points it is possible to stay at the current state or to switch to another state. The probabilities of different transitions between the states are collected into a transition matrix  $A$ , where each element  $a_{s,s'} = P(x_t = s' | x_{t-1} = s)$ , where  $s, s' \in \{1, \dots, S\}$ . Additionally, initial state probabilities  $\rho_s = P(x_1 = s)$ ,  $s \in \{1, \dots, S\}$ , describe the probabilities to start from each state at the first time point  $t = 1$ .

As a second component, we have the observed process  $y_t$ ,  $t = 1, \dots, T$ , where  $y_t$  depends only on the current state of the hidden process. In consequence, the observations are conditionally independent:

$$p(y_t | y_1, \dots, y_{t-1}, x_1, \dots, x_t) = p(y_t | x_t). \quad (2)$$

Each state  $x_t$  is associated with parameters  $\theta_{x_t}$  of the observational distribution  $p(y_t | x_t)$  and thus we can also write  $p(y_t | x_t) = p(y_t | \theta_{x_t})$ . Of the parameters  $\theta_{x_t}$  some may depend on the value of the state  $x_t$ , while others may be shared between the states. Finally, we can define an emission matrix  $\Omega$ , where  $\omega_{s,t} = p(y_t | \theta_{x_t=s})$ . Now the HMM can be defined by the set  $\{\Omega, A, \rho\}$ .

Often the hidden process and state-dependent parameters, describing the features of particular states, are the main interest of hidden Markov modelling, but since we do not directly observe the hidden process, we have to estimate it based on the observations gained from the non-hidden process. The general idea of the dependencies between the states, parameters and observations in HMMs is visualised in Figure 2.

While various methods have been suggested for estimating the number of states,  $S$ , for example, Pohle *et al.* [34], typically it is treated as known, fixed value when analysing the HMM results. The interpretation ('labelling') of the hidden states is then based on the estimated model parameters  $\{\Omega, A, \rho\}$ , and on the hidden state trajectory  $x_1, \dots, x_T$ ,

given the observations  $y_t$ . As an example, consider HMM consisting of binary observations and two states, with  $\rho = (1, 0)^T$ ,  $a_{1,1} = 0.9$ ,  $a_{2,2} = 0.5$ ,  $\omega_{1,t} = \text{Bernoulli}(0.1)$ , and  $\omega_{2,t} = \text{Bernoulli}(0.8)$ . From these, we see that we always start from state one which is relatively persistent state ( $a_{1,1} = 0.9$ ) emitting mostly zeros, whereas we are more likely to observe ones when the hidden process is in state two. Based on this, if the observations were related, for example, to the occurrence of some disease, we could label the first state as endemic state and the second state as epidemic state. We could also gain further understanding from the estimated state trajectory, from which we could see, for example, that state one is more common during winter, whereas state two occurs only during summer months. This could lead us to conclude that the transmission probability of our disease is higher during summer than winter, and thus further guide us to find reasons for that.

To perform a fully Bayesian estimation of the HMM parameters and the corresponding latent states, we first run an MCMC targeting the marginal posterior of the model parameters,  $p(\Omega, A, \rho|y)$ , where  $y = (y_1, \dots, y_T)$ . For this we need the marginal likelihood  $p(y|\Omega, A, \rho)$ , which can be computed with the forward part of the forward-backward algorithm [6,?], and the references therein]. The latent states  $x = (x_1, \dots, x_T)$  can then be sampled in the post-processing stage given the posterior samples of  $\{\Omega, A, \rho\}$ . This marginalisation approach commonly used with latent variable models allows much more efficient sampling than the approaches where the estimation is based on the joint posterior  $p(x, \Omega, A, \rho)$ , and is also strictly necessary when using efficient Hamiltonian Monte Carlo type of MCMC algorithms which rely on the differentiability of the posterior.

The marginal likelihood approach is also convenient with respect to missing data under MAR. When computing the emission probabilities  $\omega_{s,t}$ , i.e. the conditional probabilities of the observations given the current state and the model parameters, we set  $\omega_{s,t} = 1$  for all  $s$  if the observation  $y_t$  is missing. This means that as we do not gain any new information about the states due to the missing observation at time  $t$ , we essentially skip the contribution of  $\omega_{s,t}$  at the marginal likelihood computation in forward algorithm at that time point [52].

The likelihood and the posterior density of the hidden Markov models typically contain multiple modes. Some of these modes can be due to the label-switching problem [20] but in complex models true local modes can also be present. In maximum likelihood setting the model parameters are often estimated by repeating the estimation several times with varying starting values for the optimisation algorithm, with the best solution declared as a global mode, while ignoring the other modes. In Bayesian setting multimodality of the posterior can pose both computational (convergence of the MCMC algorithms) and interpretability challenges. For example, it is difficult to interpret (label) and visualise the hidden states if the corresponding observational density is multimodal. These issues have to be taken into consideration when using HMMs.

The complexity of the response variables  $y_t$  is not limited to scalars, but they can be multidimensional as well. In our case, they are vectors containing spatially dependent elements. This means that the states, as well as the parameters, should allow for spatial structure.

### **2.3. Intrinsic conditional autoregressive models**

Next, we introduce the intrinsic conditional autoregressive (ICAR) model, which is a special case of the conditional autoregressive (CAR) model [7], especially meant for spatial analysis to depict the amount of dependency between different locations. These models

are based on a spatial division of a larger area, employing a neighbourhood definition for the generated subareas, and combining both of these with a Gaussian distribution in order to quantify the amount of the spatial dependency. This method does not offer information on the direction of the effects, i.e. the causal relationships, but only on the total amount of dependency.

For the spatial division, denote each areal unit or site of a regional entity with  $i = 1, \dots, N$ , and correspondingly its neighbour with  $j = 1, \dots, N$ . To account for the contiguity of the sites, let  $W$  be an  $N \times N$  matrix for the binary neighbourhood so that the elements in it are

$$w_{i,j} = \begin{cases} 1, & \text{when sites } i \text{ and } j \text{ are neighbours, and } i \neq j \\ 0, & \text{otherwise.} \end{cases}$$

Additionally, denote with  $D$  an  $N \times N$  diagonal matrix indicating the amounts of neighbours of each site so that

$$d_{i,j} = \begin{cases} \text{number of neighbours of site } i, & \text{when } i = j \\ 0, & \text{otherwise.} \end{cases}$$

Now, we employ a variable  $\boldsymbol{\varphi} \in \mathbb{R}^N$  to express the intensity of the dependency between each site and its neighbours. In case of the CAR model, we have

$$\boldsymbol{\varphi} \sim N(0, Q^{-1}), \quad (3)$$

where the precision matrix  $Q$  has to be symmetric and positive definite for the distribution to have a proper joint probability density. This can be achieved with the matrices  $W$  and  $D$  by setting

$$Q = \tau D(I - \alpha D^{-1}W) = \tau(D - \alpha W), \quad (4)$$

where  $\tau \in \mathbb{R}^+$  is a precision parameter of the variables  $\varphi_i$ ,  $\alpha \in (0, 1)$  depicts the amount of spatial correlation, and  $I$  is an identity matrix.

The normal distribution of formula (3) can also be written as

$$p(\boldsymbol{\varphi}) = (2\pi)^{N/2} |\tau(D - \alpha W)|^{-1/2} \exp\left(-\frac{\tau}{2} \boldsymbol{\varphi}^\top (D - \alpha W) \boldsymbol{\varphi}\right), \quad (5)$$

where  $\boldsymbol{\varphi}^\top$  denotes the transpose of  $\boldsymbol{\varphi}$  and  $|x|$  stands for the determinant of  $x$ . In case of ICAR model, we assume that  $\alpha = 1$ , which results in improper distribution, which cannot be used as a model for the data, but as a prior distribution for spatial dependency, as is done for example in the Besag York Mollié (BYM) models [8]. The parameters  $\varphi_i$  are not identifiable without further constraints, as adding a constant to all of them does not change their density. A common way to avoid the identifiability problem is to restrict them to sum to zero over the sites.

Some reasons to employ the improper distribution of ICAR instead of the proper of the CAR, are the facts that, firstly, to account for a significant amount of spatial association, the parameter  $\alpha$  most likely has to be close to one, and, secondly, the width of the posterior spatial pattern might be limited in the case of CAR [4]. Additionally, when employing the

ICAR model, the specification  $\alpha = 1$  obviates the need to calculate the determinant in Equation (5) since it becomes constant, thus reducing the computation time even from days to hours [26]. Therefore, in some cases, it might be convenient to choose ICAR in the first place.

### 2.4. Main model: HMM with ICAR

In our approach, we combine the HMM and the spatial ICAR model, introduced above, thus enabling the HMM analysis of spatio-temporal data sets. We apply our method to the previously introduced historical Finnish parish register data to study the dynamics of the Finnish measles epidemics in 1750–1850. More precisely, we model the probability of observing at least one measles induced death in each town at each month.

As mentioned in Subsection 2.1, our data are missing a large number of observations. Under the assumption that they are missing at random, the HMM is a convenient approach to work with such imperfect data sets. Due to the hidden level, the observations do not depend on time, conditioned on the states, which is not the case when modelling dependencies only on the observed level. Now, the likelihood can be calculated without the missing observations by computing  $p(y_t|x_t)$  over only the observed sites at time  $t$  [52]. Thus, missing observations are naturally handled during model estimation, yet the model provides estimates for all sites and time points, with the missing cases influencing the posterior uncertainty of the estimates.

Using simulation experiments, [50] showed that the bias due to not accounting non-ignorable missingness mechanism in HMMs can depend on the stability of the hidden states. Specifically, they found that the biases were smaller in cases of large self-transition probabilities (diagonal values of  $A$ ). Speekenbrink and Visser [42] studied HMMs in a case where missingness depends on the hidden state (and is thus missing not at random, MNAR). We follow their suggestion of jointly modelling  $y_t$ , the main outcome of interest, and binary missingness indicator  $r_t$ , assuming conditional independence of  $y_t$  and  $r_t$  given the hidden state  $x_t$ . Therefore, instead of simple MAR assumption, our model is also compatible with state-dependent MNAR.

We denote each site, i.e. town, with  $i = 1, \dots, N$ , and each time point, i.e. month, with  $t = 1, \dots, T$ . We model the probability of observing at least one death  $y_{i,t}$  caused by measles in site  $i$  and time point  $t$ , and the corresponding missingness indicator  $r_{i,t}$ , as

$$y_{i,t} \sim \text{Bernoulli}(\text{logit}^{-1}(\mu_{x_t} + \lambda_i + \varphi_{i,x_t} + \gamma_t)), \tag{6}$$

$$r_{i,t} \sim \text{Bernoulli}(\text{logit}^{-1}(\xi_{x_t} + \beta_{x_t} t')), \tag{7}$$

where  $t' = (t - 1)/(T - 1)$  and the latent, discrete states

$$x_t|x_{t-1} \sim \text{Categorical}(A_{x_{t-1},\cdot}) \tag{8}$$

represent the prevailing nationwide infection statuses. The notation  $A_{x_{t-1},\cdot}$  means the row  $x_{t-1}$  including all columns of the transition matrix  $A$ . As we are using our method to study the occurrences and absences of deaths, Bernoulli distribution is a natural choice for the model. With other kinds of response variables also the distribution should be changed, for example Poisson distribution for counts. In formula (6), which corresponds to the  $p(y_t|x_t)$  in Equation (2), the logit link function maps the combination of the explanatory terms onto

a probability scale required by the Bernoulli distribution. Also the link function should be changed to an appropriate one when using a different distribution. The four explanatory terms in the model of deaths  $y_{i,t}$  describe the state specific constants, the local constants, the intensity of spatial interaction between the site and its neighbours, and a monthly seasonal variation. The missingness probability is assumed to depend on the previous missingness  $r_{i,t-1}$ , current hidden state  $x_t$ , and the (scaled) calendar time  $t'$ . The formula (8) defines the connection between the underlying states  $x_t$  and the transition matrix  $A$ . To simplify the notation, we occasionally use subscript  $s$  to refer a case  $x_t = s$ , as in  $\mu_s = \mu_{x_t=s}$ ,  $s \in \{1, \dots, S\}$ .

The state specific constant  $\mu_{x_t=s}$ ,  $s \in \{1, \dots, S\}$ , sets a base level for the nationwide probability of observing at least one death in each state  $s$ . In order to reduce the common, yet problematic, multimodality encountered with HMMs, we define the state specific constants so that they are monotonically increasing in a similar way as described in Bürkner and Charpentier [12]. Due to this, our states are ordered based on the general incidence level. In our case, to begin with, we define the constants for the first and last state,  $\mu_1$  and  $\mu_S$ , respectively. We use these two to scale the ones for the other states. We do this by defining a simplex  $\mathbf{m}$  of  $S$  components, i.e. a vector summing to one, with the first term  $m_1$  fixed to zero, and defining the rest of the constants as

$$\mu_s = \mu_1 + (\mu_S - \mu_1) \sum_{i=1}^s m_i \text{ for all } s = 1, \dots, S. \quad (9)$$

As we are using a Bayesian approach, this allows us to employ informative priors for the constants of the first and last state,  $\mu_1$  and  $\mu_S$ .

To allow local deviations from the nationwide base level, we add a site dependent term  $\lambda_i$ , which is constant over the states, i.e. over time. These local terms do not have a spatial dependency structure by definition, but they aim to capture any heterogeneous, time constant features of the sites, for example, differing population densities or communication intensities in our case. As we set the mean of the local constants to zero to improve model identifiability and MCMC sampling efficiency, we also gain a more straightforward interpretation of the state dependent constant  $\mu_s$  as a nationwide average level.

The actual spatial dependence between the sites is introduced in the third term,  $\varphi_{i,x_t}$ . The dependency is achieved via the ICAR structure outlined above, i.e. adapting the formulas (3) and (4), and letting  $\varphi_s \sim N(0, \sigma_{\varphi,s}^2 (D - W)^{-1})$ , where  $s \in \{1, \dots, S\}$ . To employ the ICAR component, we need a definition for the neighbourhood. First of all, our sites are the towns based on the parishes, i.e. they are areas with borders. Here, two regions are considered as neighbours if they share a border. Other definitions, for example, based on the distance from the centres of the sites, could be used as well. In practice, the matrices  $W$  and  $D$  are constructed according to our neighbourhood definition in order to access the covariance of the spatial terms. For identifiability, we use sum-to-zero constraint on the state specific spatial terms  $\varphi_{i,s}$ ,  $s = 1, \dots, S$ .

The final component  $\gamma_t$  in the model is a monthly average deviation additional to all the other components, defined as

$$\gamma_t = \gamma_m, \quad m = 1 + (t - 1) \bmod 12, \quad \sum_{m=1}^{12} \gamma_m = 0.$$

This means that we assume that this seasonal term is non-stochastic in a sense that the monthly effects do not vary over years, and the sum of the monthly effects is zero [14].

To estimate the model we use a Bayesian approach, which on its behalf allows smooth handling of missing data, inclusion of prior information and interpretable uncertainty quantification. This procedure requires setting prior distributions for the unknown parameters, and for the full model they are

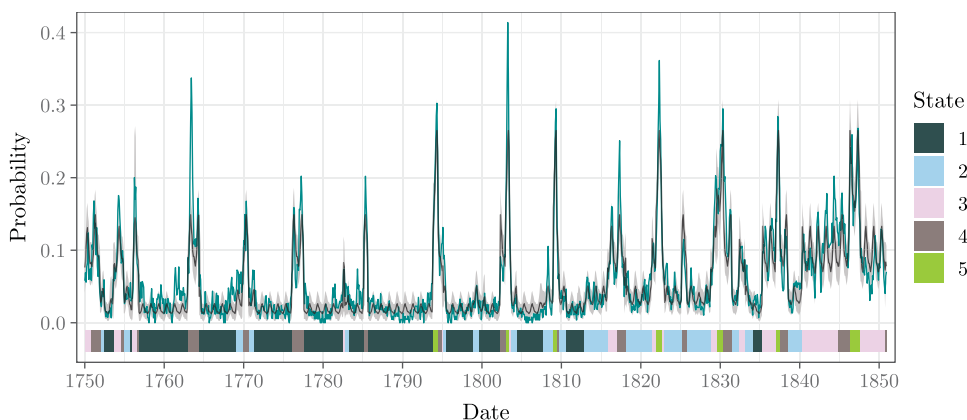
$$\begin{aligned} \mu_1 &\sim N(-4.5, 0.25^2), \\ \mu_S &\sim N(-1.75, 0.5^2)[\mu_1, ], \\ \mathbf{m}_{2:S} &\sim \text{Dirichlet}(5_1, \dots, 5_{S-1}), \\ \boldsymbol{\varphi}_s &\sim N(0, \sigma_{\varphi,s}^2(D - W)^{-1}), \text{ given } \frac{1}{N} \sum_{i=1}^N \varphi_{i,s} = 0, \\ \lambda_i &\sim N(0, \sigma_\lambda^2), \text{ given } \frac{1}{N} \sum_{i=1}^N \lambda_i = 0, \\ \sigma_\lambda &\sim N(0, 1)[0, ], \\ \sigma_\varphi &\sim N(0, 1)[0, ], \\ \boldsymbol{\gamma}_m &\sim N(0, 1), \text{ given } \frac{1}{12} \sum_{m=1}^{12} \gamma_m = 0, \\ \boldsymbol{\rho} &\sim \text{Dirichlet}(1_1, \dots, 1_S), \text{ and} \\ A_{s,\cdot} &\sim \text{Dirichlet}(p_1, \dots, p_S), \text{ where } p_{s'} = \begin{cases} 0.5, & \text{when } s' \neq s \\ 2S, & \text{when } s' = s, \end{cases} \end{aligned}$$

$s = 1, \dots, S, m = 1, \dots, 12, i = 1, \dots, N$ , where  $N(\cdot, \cdot)[z, ]$  denotes a normal distribution truncated at  $z$ . In our case, the number of time points  $T = 1212$  and the number of sites  $N = 387$ . For our final model, we set the number of states to be  $S = 5$ .

These priors are mostly weakly informative and chosen to enhance the computational efficiency [5,27]. In addition, these, together with the initial values for MCMC obtained from pilot runs aimed to locate the region of highest posterior density, are used to reduce the risk of the MCMC sampler wandering and getting trapped around minor modes of the complex posterior of HMMs. As a consequence of ignoring such regions of the posterior, the following results can underestimate the full posterior uncertainty.

### 3. Results

The model is estimated with MCMC using *cmdstanr* [15], which is an R interface [35] for the probabilistic programming language Stan for statistical inference [43]. Posterior samples are drawn using NUTS sampler [9,19] with four chains. Each chain consists of 15,000 iterations, the first 5000 being discarded as warm-up. The computation with parallel chains takes approximately 29 hours in total. The estimation is done on a supercomputer node with four cores of Xeon Gold 6230 2.1 GHz processors and 16 GB of RAM. According



**Figure 3.** The turquoise line denotes the proportion of towns where at least one death caused by measles was observed in that month according to the data. The black line shows the expected value of the posterior predictive distribution of the nationwide proportion of towns with at least one death by measles. The grey area around the black line shows the 95% posterior interval of the predicted average. The colours below the curve indicate the most likely hidden state at that time.

to the MCMC diagnostics of the *cmdstanr* [46], the model converges without any divergences. The  $\hat{R}$  statistics are all below 1.004, and the bulk and tail effective sample sizes are roughly between 1,100 and 73,000, with the smallest ones corresponding to the missingness parameters of  $\zeta$  and  $\beta$  of State 2. The R and Stan codes and the data used for the analysis are available on GitHub (<https://github.com/tihepasa/spatialHMM>).

Figure 3 shows, for each time point  $t$ , the observed proportion of towns with at least one death caused by measles, and the posterior predictive mean and 95% quantiles of this proportion based on our model. While the observed proportion is computed using only towns with non-missing observations at time  $t$ , the model-based estimates include all 387 towns. Specifically, at time  $t$ , for each posterior draw of the model parameters, we first draw a Bernoulli realisation  $\tilde{y}_{i,t}$  with probability  $\text{logit}^{-1}(\mu_s + \lambda_i + \varphi_{i,s} + \gamma_t)$  for all  $i = 1, \dots, 387$ . We then compute the marginal mean of  $\tilde{y}_{i,t}$  over  $i$ , and finally, the posterior predictive mean and quantiles are computed across all posterior draws.

It appears that State 1 is predominantly prevalent before the year 1813, while States 2, 3 and 5 are emphasised during the remainder of the study period. State 5 emerges for the first time in 1793, whereas States 2 and 3 have a few occurrences since the beginning. State 4, on the other hand, appears throughout the whole study period. In general, State 1 is the most likely during 45% of the months, State 2 during 22% of the months, and States 3 and 4 during 16% and 13% of the months, respectively, leaving State 5 prevailing during 5% of the months. The probabilities of the estimated most likely states to be the actual most likely states always range between 44–100%, with mean of 97%, less than 4% having probability smaller than 70%, and only three, randomly distributed in terms of time, falling below 50%.

The spatial dimension of the states can be seen in Figure 4, which presents the estimated local probabilities and the corresponding data of observing at least one death caused by measles in each state. The probabilities are computed in accordance with the model, as in the temporal case above, aggregated over time, and displayed on a logarithmic scale to facilitate easier comparison both between and within the states. The smallest probabilities

**Table 1.** Observed proportions, posterior mean and 95% quantiles of the death and missingness probabilities by most likely hidden state.

State	Death				Missingness			
	Observed	Mean	2.5%	97.5%	Observed	Mean	2.5%	97.5%
1	0.02	0.02	0.00	0.08	0.24	0.24	0.15	0.31
2	0.04	0.04	0.00	0.17	0.16	0.16	0.12	0.26
3	0.10	0.10	0.00	0.33	0.15	0.15	0.13	0.24
4	0.11	0.11	0.02	0.31	0.19	0.20	0.12	0.29
5	0.21	0.21	0.02	0.56	0.13	0.13	0.11	0.16

are associated with State 1, which seems to be common during the periods when the proportion of the towns observing at least one death is low, see Figure 3. In States 1, 2 and 3, the probabilities seem to be larger in the southeastern parts of the country, and as moving from State 1 to the others, the increased probabilities spread wider covering almost the whole study area in State 3. In State 4, the highest probabilities are located in the northern regions, and in State 5, they are concentrated in the southwestern half of the area.

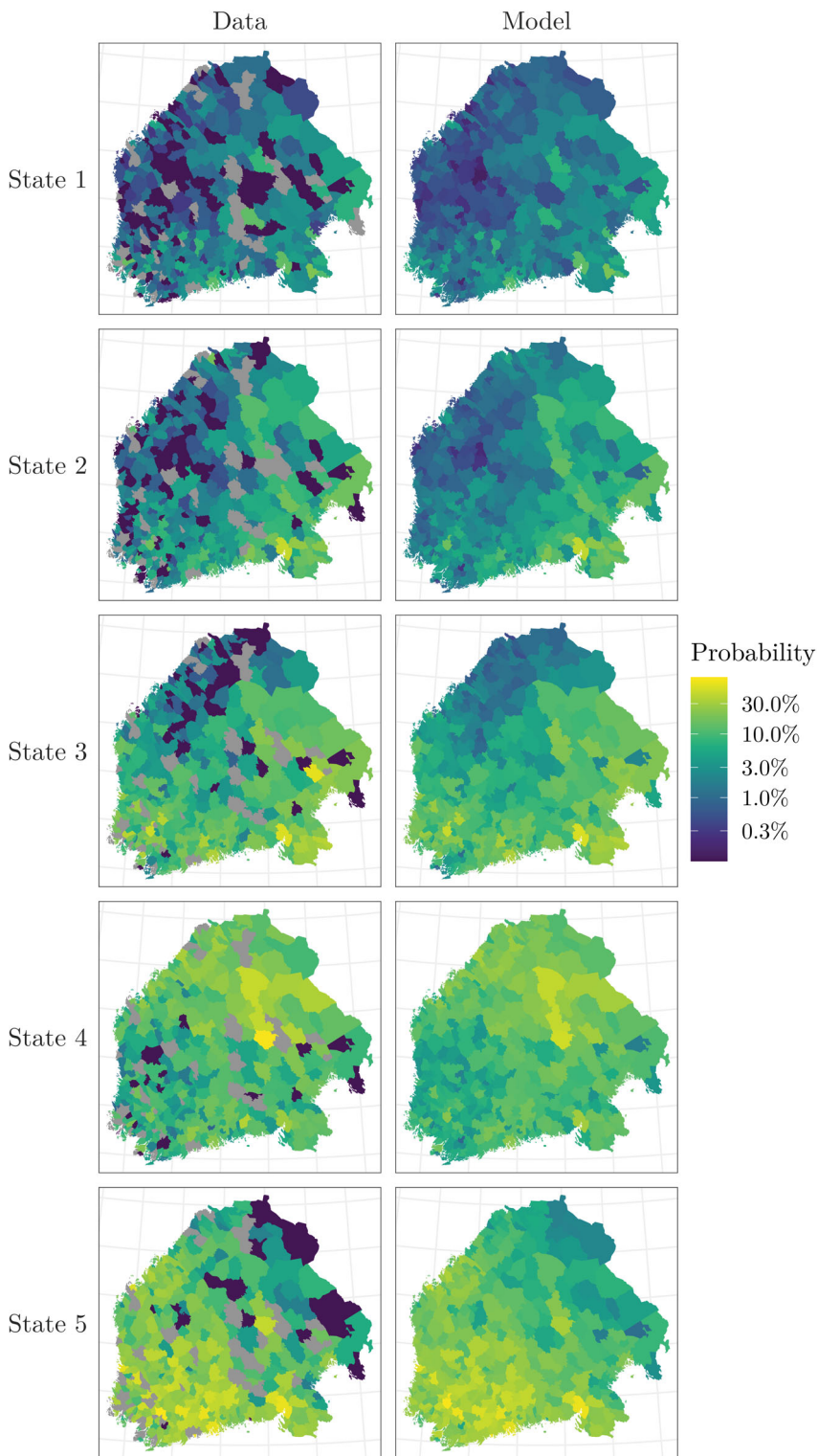
The dark blue areas in the data column of Figure 4 indicate that there are some towns where no deaths caused by measles were recorded, though it seems unlikely that any town would have been isolated enough to avoid the disease completely. Possible explanations for the lack of measles observations in these towns could be misdiagnoses or varying record-keeping practices [33].

We estimated the probability of an observation to be missing in each state over the time, and the results are shown in Figure 5. The missingness seems to be the most likely in State 5 and the least likely in State 1. In the beginning of our study period the differences between states are clear, whereas in the end the probabilities overlap. In concordance with the observed missingness, the estimated probability of missingness in each state also decreases with time.

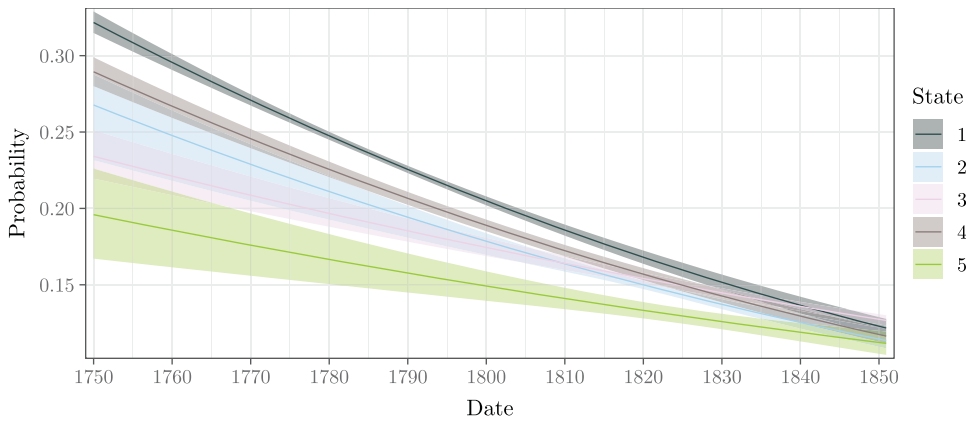
Conditioning on the most likely hidden state at each time point, we also summarise the average death and missingness probabilities over towns and time points, and the corresponding observed proportions of death and missingness indicators  $y$  and  $r$ , in Table 1. We see that on average our model performs well.

According to the transition probabilities, shown in Table 2, it is most likely to remain in the current state. If we disregard this possibility, the most probable transitions are from State 1 to State 2, from State 2 to State 1, from State 3 to States 2 and 4, from 4 to State 2, and from State 5 to State 4. It should be remembered that these probabilities describe the overall transition probabilities instead of the likelinesses of the fast, abrupt monthly transitions. The estimates of the initial state probabilities  $\rho_s$ , along with other scalar parameters, can be found from Table 3. State 3 appears to be the starting point for the state trajectory in these data. However, the posterior distributions of the initial probabilities are quite wide. This is attributable to the fact that there is only one state trajectory to estimate, and an estimation based on a single observation does not provide much information.

The scalar constants  $\mu_s$ , controlling the base levels of the probabilities to observe at least one death caused by measles in each state, are all negative in logit-scale and differ between the states distinctively, see Table 3. When these are transformed into a probability scale, the base probabilities range from 1% in State 1 to 16% in State 5.



**Figure 4.** On the left column are the proportion of observations of at least one death caused by measles in each town as over the time points when each hidden state was most likely, and on the right column are the corresponding estimated probabilities  $\text{logit}^{-1}(\mu_s + \lambda_i + \varphi_{i,s} + \gamma_t)$ , also averaged over the the time points where state  $s$  was most likely. The rows represent different states. Grey area indicates a site with only missing observations in months corresponding to that state. Note that the colour scale is logarithmic.



**Figure 5.** Estimated probability of an observation to be missing in each state over the time. The lines represent the posterior means and the shaded areas 95% posterior intervals.

**Table 2.** Posterior means and 95% posterior intervals of the transition probabilities in the transition matrix  $A$ . Rows represent the states to transfer from and the columns the states to transfer to.

	State 1	State 2	State 3	State 4	State 5
State 1	0.97 (0.95, 0.98)	0.01 (0.00, 0.03)	0.01 (0.00, 0.02)	0.01 (0.00, 0.02)	0.00 (0.00, 0.01)
State 2	0.04 (0.02, 0.07)	0.92 (0.89, 0.95)	0.02 (0.01, 0.04)	0.01 (0.00, 0.03)	0.00 (0.00, 0.02)
State 3	0.01 (0.00, 0.02)	0.03 (0.01, 0.06)	0.91 (0.87, 0.95)	0.03 (0.01, 0.06)	0.02 (0.01, 0.05)
State 4	0.03 (0.01, 0.06)	0.05 (0.02, 0.09)	0.02 (0.00, 0.04)	0.89 (0.84, 0.94)	0.02 (0.00, 0.04)
State 5	0.01 (0.00, 0.03)	0.01 (0.00, 0.04)	0.04 (0.01, 0.10)	0.07 (0.02, 0.14)	0.87 (0.79, 0.94)

The local constants  $\lambda_i$ , independent of the state, are shown in Figure 6. As expected, there are no visible spatial patterns in these parameters, aimed to capture spatially independent differences of the sites. The deviation  $\sigma_\lambda$  of the constants  $\lambda_i$  is 0.58 with posterior interval [0.53, 0.65], see also Table 3, suggesting that there is some non-spatial heterogeneity between the towns. For instance, in State 5, the 5% and 95% quantiles for the posterior mean of the local probabilities  $\text{logit}^{-1}(\mu_S + \lambda_i)$  are (0.07, 0.33) with median 0.17.

The spatial dependence in the spread of measles is accounted for by the spatial terms  $\varphi_{i,s}$ . The deviation parameter  $\sigma_\varphi$  varies between states, from posterior mean of 1.4 [1.21, 1.63] in state 1 to 0.6 [0.47, 0.73] in state 4, see also Table 3. Combined with the information on the most likely states in each time point, we can deduce that the deviation of the spatial components  $\varphi_{i,s}$  is greatest roughly before 1808. The state-specific spatial terms  $\varphi_{i,s}$  are shown in Figure 7. The spatial patterns are similar to those of the estimated probabilities to observe at least one death caused by measles in Figure 4.

To capture the yearly temporal variation, the model includes a seasonal component  $\gamma_t$ , which is shown in Figure 8. These monthly terms sum to zero over a year, i.e. they depict an average monthly variation additional to the other components of the model. The effect of the month increases from January to peak in May and decreases after that until December.

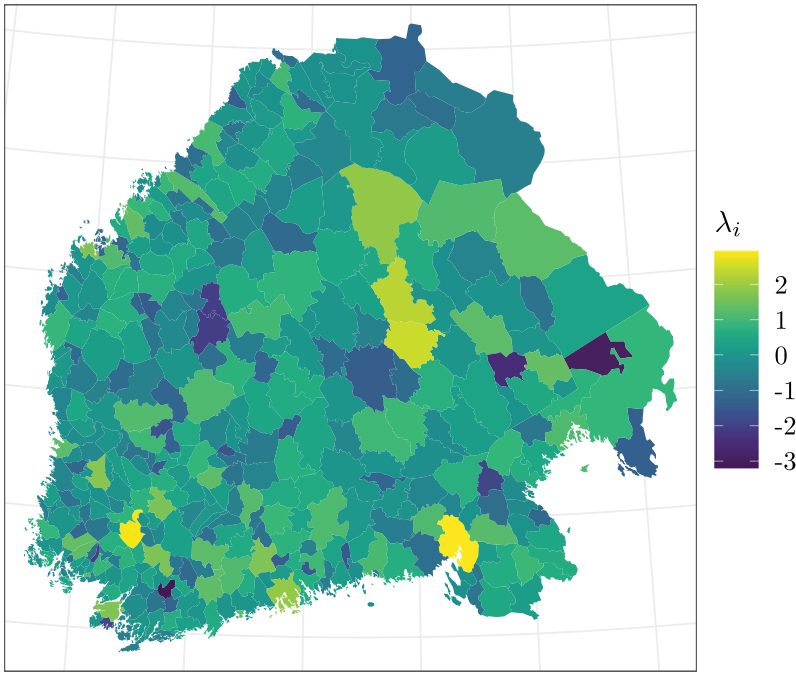
**Table 3.** Posterior means and standard deviations, 95% posterior intervals, bulk and tail effective sample sizes (ess) and  $\widehat{R}$  statistics of the initial probabilities  $\rho_s$ , the state specific constants  $\mu_s$ , the deviation of local constants  $\sigma_{\lambda_s}$ , and the deviation parameter  $\sigma_{\phi}$  of the ICAR component.

	mean	sd	2.5%	97.5%	ess bulk	ess tail	$\widehat{R}$
$\rho_1$	0.17	0.14	0.00	0.52	47223	21092	1.00
$\rho_2$	0.17	0.14	0.00	0.53	28095	23330	1.00
$\rho_3$	0.33	0.18	0.05	0.72	32695	23143	1.00
$\rho_4$	0.17	0.14	0.01	0.52	52046	23485	1.00
$\rho_5$	0.17	0.14	0.01	0.52	54981	24539	1.00
$\mu_1$	-4.63	0.05	-4.73	-4.53	8688	11541	1.00
$\mu_2$	-3.93	0.05	-4.02	-3.84	9585	15357	1.00
$\mu_3$	-2.74	0.03	-2.80	-2.67	20500	28782	1.00
$\mu_4$	-2.39	0.03	-2.45	-2.34	28311	31093	1.00
$\mu_5$	-1.66	0.03	-1.72	-1.59	21989	28406	1.00
$\sigma_{\phi_1}$	1.41	0.11	1.21	1.63	8132	15750	1.00
$\sigma_{\phi_2}$	1.02	0.09	0.84	1.21	3875	6601	1.00
$\sigma_{\phi_3}$	0.80	0.07	0.68	0.94	4994	9495	1.00
$\sigma_{\phi_4}$	0.59	0.07	0.47	0.73	3823	5619	1.00
$\sigma_{\phi_5}$	0.75	0.07	0.63	0.89	5432	11167	1.00
$\sigma_{\lambda_s}$	0.58	0.03	0.53	0.65	11795	20730	1.00
$\xi_1$	-0.75	0.02	-0.78	-0.71	10165	18381	1.00
$\xi_2$	-1.00	0.07	-1.17	-0.91	1858	1125	1.00
$\xi_3$	-1.19	0.04	-1.27	-1.10	4733	2167	1.00
$\xi_4$	-0.90	0.02	-0.94	-0.85	36124	28992	1.00
$\xi_5$	-1.42	0.10	-1.61	-1.23	20976	25885	1.00
$\beta_1$	1.23	0.04	-1.32	-1.15	11799	19076	1.00
$\beta_2$	-1.05	0.09	-1.19	-0.83	1924	1148	1.00
$\beta_3$	-0.74	0.05	-0.85	-0.65	4993	2215	1.00
$\beta_4$	-1.13	0.04	-1.21	-1.05	36627	28818	1.00
$\beta_5$	-0.66	0.13	-0.90	-0.41	22940	27358	1.00

The effect is positive, thereby increasing the probability of observing at least one death caused by measles, from March to July, and negative or decreasing the risk otherwise. A somewhat similar pattern, with a peak in May, was also identified in [32] for measles during 1820–1850 in Finland.

While our HMM was estimated as homogeneous in terms of the transition matrix  $A$ , Figure 3 suggests that the probability of observing deaths by measles changed around 1810. Before this, the hidden State 1 with the lowest baseline probabilities is the most prominent, after which the pattern of the hidden states appears to change considerably, both in terms of the baseline probability (transitioning from State 1 to State 2) and more frequent occurrences in the west coast (State 3). Therefore, as an additional analysis, we aimed to identify the most probable point in time where the spatio-temporal dynamics of measles changed regimes. For this purpose, we estimated a two-state left-to-right HMM, where the observations consisted of a thousand posterior samples of the state trajectories of our main model. All trajectories were modelled jointly so that they shared the same model parameters (state-specific categorical emission probabilities and the single transition probability from State 1 to State 2), but each trajectory had a separate hidden state process.

From the sampled state trajectories of this left-to-right model, we computed the distribution of the most probable change point, which was identified as November 1812 (with a 95% posterior interval from October 1812 to February 1813). Table 4 displays the estimated emission matrix of this model. From this, we can conclude that the major changes were the increased probabilities of States 2, 3, and 5 and the decreased probability of State



**Figure 6.** The posterior means of the local constants  $\lambda_{i,S}$  by state.

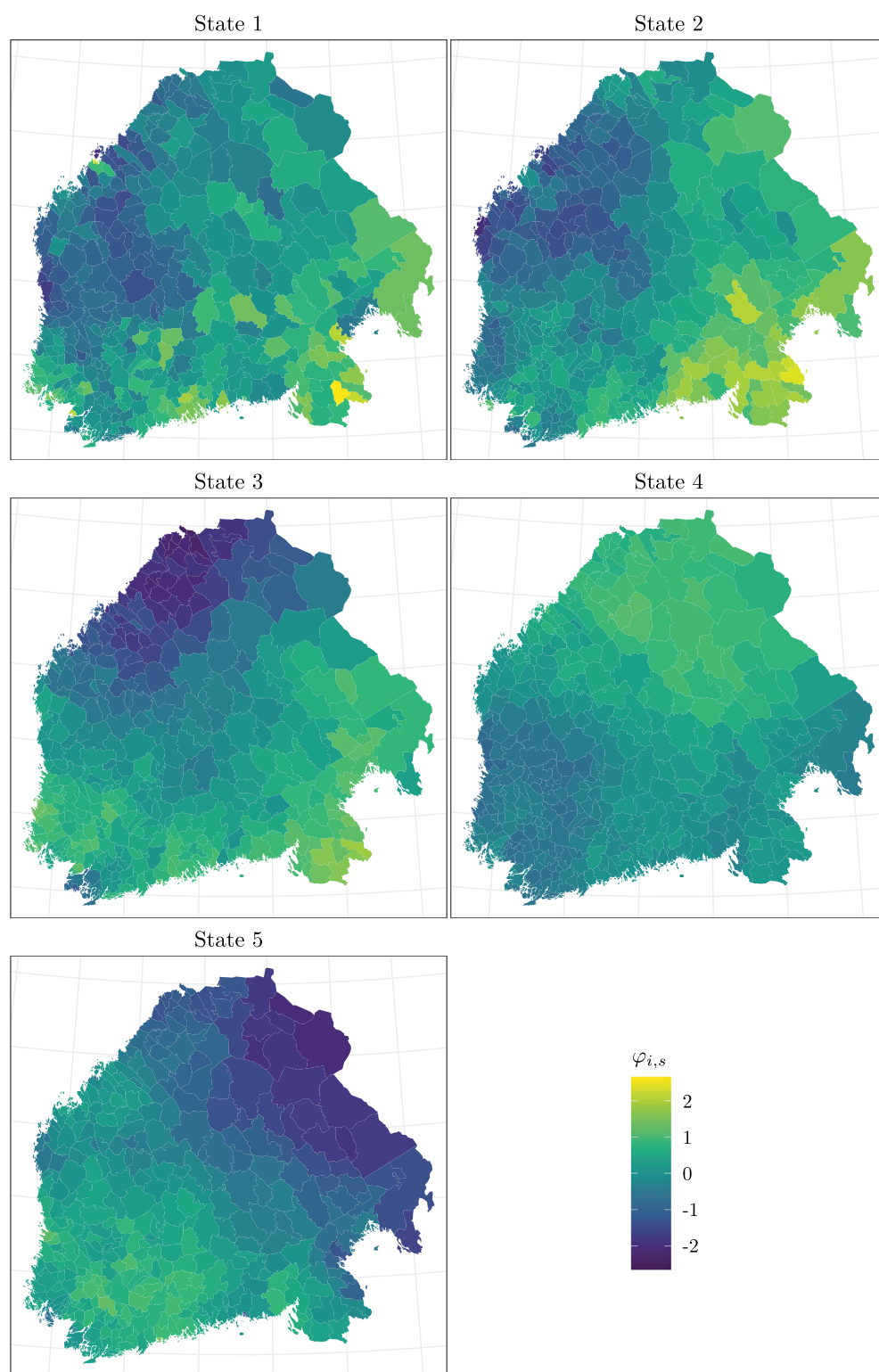
**Table 4.** The emission matrix of the change point model including the posterior means and 95% posterior intervals of the probabilities of being in each state before and after the change point.

	State 1	State 2	State 3	State 4	State 5
Before	0.69 (0.69, 0.69)	0.12 (0.12, 0.12)	0.05 (0.05, 0.05)	0.12 (0.12, 0.12)	0.03 (0.02, 0.03)
After	0.04 (0.04, 0.04)	0.39 (0.39, 0.39)	0.34 (0.34, 0.34)	0.14 (0.14, 0.14)	0.09 (0.09, 0.9)

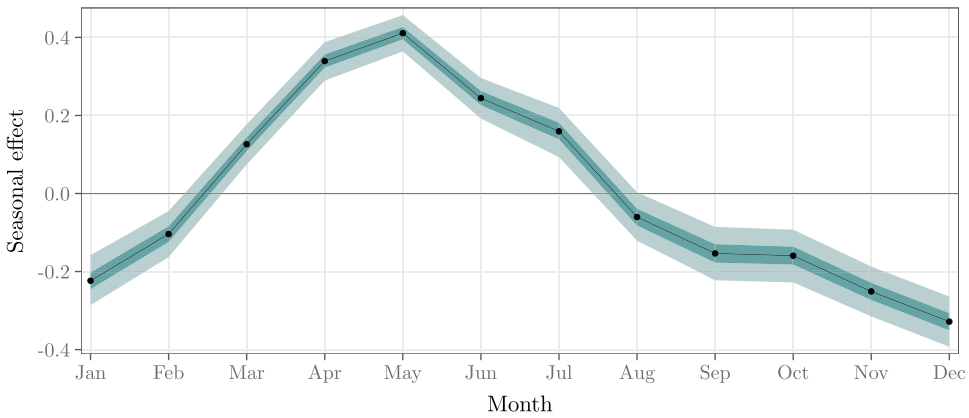
1 after November 1812. This finding aligns well with the associated administrative changes that led to transferring the capital towards the east, and to altering the trade connections [30].

Determining the number of states in HMM is not a trivial task, and there is no definitive procedure for that, even though some suggestions exist [34, e.g.]. In our case, we utilised five states, although a corresponding model was also estimated with four and six states. With six states there were several modes and convergence issues not only between but also within the chains. The model with four states, on the other hand, passed the convergence criteria but was not chosen here. Given that the states exhibited clear differences when there were five of them, we opted for the model with a greater number of hidden states. Presumably, this more detailed model also depicts the underlying phenomena more realistically in this context, since the epidemics are not discrete but continuous processes both in space and time.

Additionally, we compared the four-state and five-state versions of the main model, as well as a simpler model which has a state-independent  $\sigma_\varphi$  and which does not have an



**Figure 7.** Posterior means of the spatial terms  $\varphi_{i,s}$  by state.



**Figure 8.** The black dots and line show the posterior mean of the monthly seasonal term  $\gamma_t$ . The dark and pale turquoise areas represent the 50% and 95% posterior intervals, respectively.

**Table 5.** Mean and standard error of pairwise ELPD differences over ten heldout samples for different models. Positive value means that the left-hand side model performed better. The simpler models are specified in Section 2 in Supplementary material.

Comparison	ELPD <sub>diff</sub>	SE <sub>diff</sub>
Simpler S = 5 vs. Simpler S = 4	2.7	0.9
Main S = 5 vs. Main S = 4	3.2	0.9
Main S = 5 vs. Simpler S = 5	1.2	0.6

explicit missingness model (see Section 2 in Supplementary material) using expected log predictive density (ELPD). This measures the goodness of the entire predictive distribution [45]. The commonly used approximate leave-one-out cross-validation resulted in high Pareto  $\hat{k}$  values [45], likely due to the multilevel structure with town-specific parameters and mixture nature of the observation probabilities. Therefore, we created ten replications of the original data, randomly leaving out additional 1% (2838) observations in each. We then estimated candidate models and computed the pointwise log-likelihoods for the held-out data. Table 5 shows the average and standard error of pairwise ELPD differences over the replicates when comparing different models. These results suggest a preference for the five-state models as well as for our main model. However, the average ELPD difference between the two five-state models is relatively small, which is to be expected given their very similar performance for the death-occurrence outcome.

For comparison, the supplementary material contains figures (Supplementary Figures 1–6) and tables (Supplementary Tables 1–2) for the four-state model, corresponding to the ones represented here for the five-state model in Figures 3–8 and Tables 2 and 3. Based on Figures 3 and 4 and Supplementary Figures 1 and 2, it seems that States 1 and 2 of the five-state model have merged into State 1 of the four-state model, but otherwise the states correspond to each other. Also in the four-state model, a change in the dynamics towards the end of the time series is visible (see Supplementary Figure 1). The figures

and tables representing the results of the simpler models with four and five states are also included in Supplementary material.

#### 4. Discussion

In this study, we demonstrated the usage of the hidden Markov models in spatio-temporal analysis. Our aim was to gain information on the historical Finnish measles epidemics as a spatial and temporal process by modelling the probability of observing at least one death caused by measles in each town and month. The proposed HMM approach allowed us to summarise the information of data that are otherwise complicated and challenging to understand.

Using reported fatalities of measles in Finland throughout 1750–1850, and the Bayesian hidden Markov model with state-dependent spatial correlation structure, we identified five reoccurring epidemic states in Finland. The epidemics were characterised by two states of low burden of infection (States 1 and 2), and three states of higher infectious burden (States 3, 4 and 5), covering different parts of Finland (see Figure 4).

Our analyses revealed also a distinctive change in the epidemic dynamics between November 1812 and February 1813, matching well with the changes caused by the war between Sweden and Russia over the area of Finland in 1808–1809, and the imposed large administrative changes, including the transfer of the capital from Turku to Helsinki, in 1812. These transformations strengthened significantly the eastern influence and trade connections [30]. The most notable change in the estimated state trajectory after the annexation of Finland by Russia was the replacement of State 1 with a generally higher death observation probability of State 2 as a non-epidemic state. Also the likelihood of States 3 and 5 increased, reflecting the higher infection burden at southern Finland and Karelian Isthmus, or south-western Finland, respectively. The view of the strengthened eastern influence is further supported by the increase of the infection burden especially in the eastern towns during the low epidemic states (Figure 4), possibly indicating persistent infectious pressure from the nearby, then capital of Russia, Saint Petersburg (see Figure 1). Overall, infections became more common across whole Finland after the war.

The dynamics of measles and other contagious infections are sensitive to population size [21]. Due to the clear abrupt change in infection dynamics, we can perhaps safely rule out the role of the gradual population increase [48] here as the main driver of the change of the epidemics. Further examination considering particularly the spatial effect of the population sizes remains as a future work due to the missing local population size information. Especially sparsely populated areas without endemic measles infections follow the changes of adjacent larger populations [16,21,22,37]. Epidemics are affected by transport networks, geographical obstacles and borders [47]. Finland and Sweden are separated by a large geographical barrier: The Gulf of Bothnia is 80–150 kilometres wide and 700 kilometres long, and obviously restricting the movement of people – and infections. Sea on the other side and the Russian border on the other side seem to have kept the infectious burden smaller in pre-war Finland compared with the post-war era when contacts to Russia and elsewhere increased strongly [30]. During the history of nations, the borders have changed due to wars and the directions of transport and trade have shifted frequently, but the dynamics of

epidemics due to such regime shifts have not been closely followed. Luckily, in our case, the administrative changes did not modify the data collection procedures of parishes so that we had data of the deaths available. This allowed us to study more closely the dynamics of epidemics due to changing transmission networks.

When it comes to the local transmission routes inside Finland, the definition of the proximity between the towns plays a crucial role. Other relevant definitions for the neighbourhood could be road or water connections instead of or in addition to the border sharing we used. As during longer time periods the areal structure is rarely static due to wars and other environmental changes, it would be interesting to extend the model into a case in which the spatial structure and neighbours can vary in time. This would require some further specifications for the state-specific spatial coefficients, as allowing several spatial structures within one state could result in having not only  $S$ , but the number of different structures times  $S$ , states. To set the specifications and try this kind of approach in practice, one should have further information about the spatial structure and the changes in it, but technically it could be implemented with our model. Additionally, within a fixed spatial structure one could estimate if some neighbouring towns did not interact at all so that there were actually discontinuities, as in Balocchi and Jensen [3].

As mentioned in Section 1, there are several ways to model and estimate spatio-temporal connections. It would be natural to consider the phenomena as a continuous process instead of merely inspecting a collection of discrete states. One attractive way for this kind of more detailed analysis would be to fit a model without any hidden states, using random walks so that  $y_{i,t} \sim \text{Bernoulli}(\mu_t + \lambda_i + \varphi_{i,t} + \gamma_t)$ , where  $\mu_t \sim N(\mu_{t-1}, \sigma_\mu^2)$ , and  $\varphi_t \sim N(\varphi_{t-1}, \sigma_\varphi(D - W)^{-1})$ . In this case the spatio-temporal parameter  $\varphi_{i,t}$  may be seen as an interaction parameter of the temporal and spatial terms, as described for example in Clayton and Bernardinelli [13]. Given the large number of time points, spatial units, and iterations needed for the convergence of the MCMC, this kind of model is computationally infeasible to estimate in our case using MCMC, although approximate solutions using INLA would have likely been obtainable. However, the downside of such models is that due to the parameter profusion, the results would not be as interpretable or offer such a straightforward insight as our hidden Markov model.

We allowed for simplified temporal dependency structure for the model via the parameters that depend on the state. Our method is especially convenient when there is a reason to suspect a presence of recurring patterns in the underlying phenomena. In other words, as we allow all potential transitions between our states so that it is possible to return to any of the previously occurred states later in time, also the latent process consists of repeated patterns. This enables finding the spatio-temporal epidemic dynamics and transmission routes illustrated by the repeated peak and fade-out patterns. In addition to the state-specific constants  $\mu_s$  and spatial terms  $\varphi_{i,t}$  and the corresponding deviation parameters  $\sigma_\varphi$ , also the local constant  $\lambda_i$ , their standard deviations, and the seasonal term  $\gamma_t$  could depend on state if seen appropriate. Our attempts to include the state dependent local constants and seasonal terms led to some identifiability issues. The same was also true for an alternative model in which, instead of the state-dependent  $\mu_s$ , the model included a temporal trend  $\mu_t$  which evolved continuously over time (in which case the states describe only changes in the spatial patterns). Thus, the inclusion of these further complexities in our model remains as a future work.

Our data have a large proportion of missing observations due to different reasons, for example, register keeping practices and data preservation processes. The small population sizes cause additional missingness because for the towns the months without any registered deaths are coded as missing even though it is possible that in reality no one died, but we have no means to know that with the information we have currently available. As a result, we aggregated the data and modelled the missingness to enhance the reliability of our analysis. In this way, we were also able to gain information on the probability of missingness in different states during the hundred-year study period.

While our focus was on binary data, other observational distributions could be used and mixed, for example, a joint model of Poisson and Gaussian responses. There are previous indications of the profitability of the dichotomisation considering the last 30 years of our data [32]. The model may also be extended to a case with exogenous covariates for the observational distributions or for the transition probabilities.

## Note

1. This text was originally published as a preprint (<https://arxiv.org/abs/2405.16885v1>).

## Acknowledgments

The authors wish to acknowledge CSC – IT Centre for Science, Finland, for computational resources. The authors also thank Virpi Lummaa and Harri Högmander. This research was partially supported by the INVEST Research Flagship Centre.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research was supported by the Research Council of Finland (grant numbers 331817, 355153, 345546, and 278751), Emil Aaltonen Foundation, the Finnish Cultural Foundation.

## References

- [1] R. Amorós, D. Conesa, A. López-Quílez, and M.A. Martínez-Beneito, *A spatio-temporal hierarchical Markov switching model for the early detection of influenza outbreaks*, *Stoch. Environ. Res. Risk Assess.* 34 (2020), pp. 275–292. <https://doi.org/10.1007/s00477-020-01773-5>.
- [2] H. Bakka, H. Rue, G.A. Fuglstad, A. Riebler, D. Bolin, J. Illian, E. Krainski, D. Simpson, and F. Lindgren, *Spatial modeling with R-INLA: A review*, *WIREs Comput. Stat.* 10 (2018), pp. e1443 <https://doi.org/10.1002/wics.1443>.
- [3] C. Balocchi and S.T. Jensen, *Spatial modeling of trends in crime over time in Philadelphia*, *Ann. Appl. Stat.* 13 (2019), pp. 2235–2259. <https://doi.org/10.1214/19-aos1280>.
- [4] S. Banerjee, B. Carlin, and A. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed., Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 2015. <https://doi.org/10.1201/b17115-10>
- [5] K.M. Banner, K.M. Irvine, and T.J. Rodhouse, *The use of Bayesian priors in Ecology: The good, the bad and the not great*, *Methods Ecol. Evol.* 11 (2020), pp. 882–889. <https://doi.org/10.1111/2041-210X.13407>.
- [6] L.E. Baum and T. Petrie, *Statistical inference for probabilistic functions of finite state Markov chains*, *Ann. Math. Stat.* 37 (1966), pp. 1554–1563. <https://doi.org/10.1214/aoms/1177699147>.

- [7] J. Besag, *Spatial interaction and the statistical analysis of lattice systems*, J. R. Stat. Soc. Ser. B (Methodol.) 36 (1974), pp. 192–225. <https://doi.org/10.1111/j.2517-6161.1974.tb00999.x>.
- [8] J. Besag, J. York, and A. Mollié, *Bayesian image restoration, with two applications in spatial statistics*, Ann. Inst. Stat. Math. 43 (1991), pp. 1–20. <https://doi.org/10.1007/BF00116466>.
- [9] M. Betancourt, *A conceptual introduction to Hamiltonian Monte Carlo*, preprint (July 16, 2018), arXiv [accessed May 3, 2024]. <https://doi.org/10.48550/arXiv.1701.02434>.
- [10] M. Briga, T. Ketola, and V. Lummaa, *The epidemic dynamics of three childhood infections and the impact of first vaccination in 18th and 19th century Finland*, preprint (October 31, 2022), medRxiv [accessed May 25, 2023]. <https://doi.org/10.1101/2022.10.30.22281707>.
- [11] M. Briga, S. Ukonaho, J.E. Pettay, R.J. Taylor, T. Ketola, and V. Lummaa, *The seasonality of three childhood infections in a pre-industrial society without schools*, preprint (October 18, 2021), medRxiv [accessed September 29, 2023]. <https://doi.org/10.1101/2021.10.08.21264734>.
- [12] P.C. Bürkner and E. Charpentier, *Modelling monotonic effects of ordinal predictors in Bayesian regression models*, Br. J. Math. Stat. Psychol. 73 (2020), pp. 420–451. <https://doi.org/10.1111/bmsp.12195>.
- [13] D. Clayton and L. Bernardinelli, *Bayesian methods for mapping disease risk*, in *Geographical and Environmental Epidemiology: Methods for Small Area Studies*, P. Elliott, J. Cuzick, D. English, and R. Stern, ed., Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780192622358.003.0018>.
- [14] J. Durbin and S.J. Koopman, *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, 2012. <https://doi.org/10.1093/acprof:oso/9780199641178.001.0001>.
- [15] J. Gabry and R. Češnovar, *cmdstanr: R Interface to 'CmdStan'* (2022). Available at <https://mc-stan.org/cmdstanr/>, <https://discourse.mc-stan.org>.
- [16] B. Grenfell and J. Harwood, *(Meta)population dynamics of infectious diseases*, Trends Ecol. Evol. 12 (1997), pp. 395–399. [https://doi.org/10.1016/S0169-5347\(97\)01174-9](https://doi.org/10.1016/S0169-5347(97)01174-9).
- [17] S. Helske and J. Helske, *Mixture hidden Markov models for sequence data: The seqHMM package in R*, J. Stat. Softw. 88 (2019), pp. 1–32. <https://doi.org/10.18637/jss.v088.i03>.
- [18] S. Helske, J. Helske, and M. Eerola, *Combining sequence analysis and hidden Markov models in the analysis of complex life sequence data*, in *Sequence Analysis and Related Approaches. Life Course Research and Social Policies*, Vol. 10, G. Ritschard and M. Studer, eds., Springer International Publishing, Cham, 2018, pp. 185–200. [https://doi.org/10.1007/978-3-319-95420-2\\_11](https://doi.org/10.1007/978-3-319-95420-2_11).
- [19] M.D. Hoffman and A. Gelman, *The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo*, J. Mach. Learn. Res. 15 (2014), pp. 1593–1623.
- [20] A. Jasra, C.C. Holmes, and D.A. Stephens, *Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling*, Stat. Sci. 20 (2005), pp. 50–67. <https://doi.org/10.1214/088342305000000016>.
- [21] M.J. Keeling and B.T. Grenfell, *Disease extinction and community size: Modeling the persistence of measles*, Science 275 (1997), pp. 65–67. <https://doi.org/10.1126/science.275.5296.65>.
- [22] T. Ketola, M. Briga, T. Honkola, and V. Lummaa, *Town population size and structuring into villages and households drive infectious disease risks in pre-healthcare Finland*, Proc. R. Soc. B Biol. Sci. 288 (2021), pp.20210356 <https://doi.org/10.1098/rspb.2021.0356>.
- [23] L. Knorr-Held, *Bayesian modelling of inseparable space-time variation in disease risk*, Stat. Med. 19 (2000), pp. 2555–2567. [https://doi.org/10.1002/1097-0258\(20000915/30\)19:17/18;2555::AID-SIM587;3.0.CO;2-RAJ23](https://doi.org/10.1002/1097-0258(20000915/30)19:17/18;2555::AID-SIM587;3.0.CO;2-RAJ23).
- [24] L. Knorr-Held and S. Richardson, *A hierarchical model for space-time surveillance data on meningococcal disease incidence*, J. R. Stat. Soc. Ser. C: Appl. Stat. 52 (2003), pp. 169–183. <https://doi.org/10.1111/1467-9876.00396>.
- [25] F. Lin, X. Wang, and M. Zhou, *How trade affects pandemics? Evidence from severe acute respiratory syndromes in 2003*, World Econ. 45 (2022), pp. 2270–2283. <https://doi.org/10.1111/twec.13127>.
- [26] M. Morris, K. Wheeler-Martin, D. Simpson, S.J. Mooney, A. Gelman, and C. DiMaggio, *Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan*, Spat. Spatiotemporal. Epidemiol. 31 (2019), pp.100301 <https://doi.org/10.1016/j.sste.2019.100301>.

- [27] C. Muth, Z. Oravecz, and J. Gabry, *User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan*, *Quant. Methods Psychol.* 14 (2018), pp. 99–119. <https://doi.org/10.20982/tqmp.14.2.p099>.
- [28] A. Nitsch, V. Lummaa, T. Ketola, T. Honkola, O. Vesakoski, and M. Briga, *The spatial distribution of pertussis, but not measles or smallpox, in pre-industrial Finland matches dialects*, *iScience* 28 (2025), pp.112530<https://doi.org/10.1016/j.isci.2025.112530>.
- [29] H. Ogden, *On asymptotic validity of naive inference with an approximate likelihood*, *Biometrika* 104 (2017), pp. 153–164. <https://doi.org/10.1093/biomet/asx002>.
- [30] J. Ojala and A. Rähkä, *Navigation Acts and the integration of North Baltic shipping in the early nineteenth century*, *Int. J. Marit. Hist.* 29 (2017), pp. 26–43. <https://doi.org/10.1177/0843871416678166>.
- [31] E. Oster, *Routes of infection: Exports and HIV incidence in Sub-Saharan Africa*, *J. Eur. Econ. Assoc.* 10 (2012), pp. 1025–1058. <https://doi.org/10.1111/j.1542-4774.2012.01075.x>.
- [32] T.M. Pasanen, J. Helske, H. Högmänder, and T. Ketola, *Spatio-temporal modeling of co-dynamics of smallpox, measles, and pertussis in pre-healthcare Finland*, *PeerJ* 12 (2024), pp. e18155<https://doi.org/10.7717/peerj.18155>.
- [33] K. Pitkänen, *The reliability of the registration of births and deaths in Finland in the eighteenth and nineteenth centuries: Some examples*, *Scand. Econ. Hist. Rev.* 25 (1977), pp. 138–159. <https://doi.org/10.1080/03585522.1977.10407878>.
- [34] J. Pohle, R. Langrock, F.M. Van Beest, and N.M. Schmidt, *Selecting the number of states in hidden Markov models: Pragmatic solutions illustrated using animal movement*, *J. Agric. Biol. Environ. Stat.* 22 (2017), pp. 270–293. <https://doi.org/10.1007/s13253-017-0283-8>.
- [35] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2023. <https://www.R-project.org/>.
- [36] L.R. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition*, *Proc. IEEE* 77 (1989), pp. 257–286. <https://doi.org/10.1109/5.18626>.
- [37] P. Rohani, C. Green, N. Mantilla-Beniers, and B. Grenfell, *Ecological interference between fatal diseases*, *Nature* 422 (2003), pp. 885–888. <https://doi.org/10.1038/nature01542>.
- [38] H. Rue, S. Martino, and N. Chopin, *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations*, *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 71 (2009), pp. 319–392. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>.
- [39] H. Rue, A. Riebler, S.H. Sørbye, J.B. Illian, D.P. Simpson, and F.K. Lindgren, *Bayesian computing with INLA: A review*, *Annu. Rev. Stat. Appl.* 4 (2017), pp. 395–421. <https://doi.org/10.1146/annurev-statistics-060116-054045>.
- [40] T. Saarivirta, D. Consoli, and P. Dhondt, *Suomen terveydenhuoltojärjestelmän ja sairaaloiden kehittyminen, Vaatimattomista oloista modernin terveydenhuollon eturintamaan*. *Kasvatus & Aika* 3 (2009). Available at <https://journal.fi/kasvatusjajaika/article/view/68129>.
- [41] T. Saarivirta, D. Consoli, and P. Dhondt, *The evolution of the Finnish health-care system early 19th century and onwards*, *Int. J. Bus. Soc. Sci.* 3 (2012), pp. 243–257.
- [42] M. Speekenbrink and I. Visser, *Ignorable and non-ignorable missing data in hidden Markov models*, preprint (2021) arXiv 2109.02770. <https://doi.org/10.48550/arXiv.2109.02770>.
- [43] Stan Development Team, *Stan modeling language users guide and reference manual, version 2.34* (2024). Available at <https://mc-stan.org>.
- [44] Tilastollinen päätoimisto, *Pääpiirteet Suomen väestötilastosta vuosina 1750–1890: 1: Väestön tila*, SVT: Suomen virallinen tilasto VI. Väkiluvun-tilastoa 29. Table 1 A (1899).
- [45] A. Vehtari, A. Gelman, and J. Gabry, *Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC*, *Stat. Comput.* 27 (2017), pp. 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>.
- [46] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.C. Bürkner, *Rank-normalization, folding, and localization: An improved  $\hat{R}$  for assessing convergence of MCMC (with discussion)*, *Bayes. Anal.* 16 (2021), pp. 667–718. <https://doi.org/10.1214/20-BA1221>.
- [47] A. Vora, D.S. Burke, and D.A.T. Cummings, *The impact of a physical geographic barrier on the dynamics of measles*, *Epidemiol. Infect.* 136 (2008), pp. 713–720. <https://doi.org/10.1017/S0950268807009193>.

- [48] M. Voutilainen, J. Helske, and H. Högmander, *A Bayesian reconstruction of a historical population in Finland, 1647–1850*, *Demography* 57 (2020), pp. 1171–1192. <https://doi.org/10.1007/s13524-020-00889-1>.
- [49] H. Vuorinen, *Suomalainen tautinimistö ennen bakteriologista vallankumousta*, *Hippokrates: Suomen lääketieteen historian seuran vuosikirja* 16 (1999), pp. 33–61.
- [50] H.W. Yeh, W. Chan, and E. Symanski, *Intermittent missing observations in discrete-time hidden Markov models*, *Commun. Stat. Simul. Comput.* 41 (2012), pp. 167–181. [10.1080/03610918.2011.581778](https://doi.org/10.1080/03610918.2011.581778).
- [51] R.P. Yue, H.F. Lee, and C.Y. Wu, *Trade routes and plague transmission in pre-industrial Europe*, *Sci. Rep.* 7 (2017), pp. 12973. <https://doi.org/10.1038/s41598-017-13481-2>.
- [52] W. Zucchini and I. MacDonald, *Hidden Markov Models for Time Series: An Introduction Using R*, Chapman and Hall/CRC, New York, 2009. <https://doi.org/10.1201/9781420010893>.