JYU DISSERTATIONS 864

Tiia-Maria Pasanen

Bayesian Spatio-Temporal Modeling of Areal Data

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF MATHEMATICS
AND SCIENCE

JYU DISSERTATIONS 864

Tiia-Maria Pasanen

Bayesian Spatio-Temporal Modeling of Areal Data

Esitetään Jyväskylän yliopiston matemaattis-luonnontieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Agora-rakennuksen auditoriossa 2
tammikuun 11. päivänä 2025 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Mathematics and Science of the University of Jyväskylä,
in building Agora, auditorium 2, on January 11, 2025 at 12 o'clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2025

Editors
Jouni Helske
Department of Mathematics and Statistics, University of Jyväskylä
Ville Korkiakangas
Open Science Centre, University of Jyväskylä

ISBN 978-952-86-0446-4 (PDF)
URN:ISBN:978-952-86-0446-4
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-952-86-0446-4

ABSTRACT

Pasanen, Tiia-Maria
Bayesian spatio-temporal modeling of areal data
Jyväskylä: University of Jyväskylä, 2025, 52 p. (+included articles)
(JYU Dissertations
ISSN 2489-9003; 864)
ISBN 978-952-86-0446-4 (PDF)

Data may be considered spatio-temporal when they describe both location and time related to a phenomenon. Spatio-temporal statistical methodology aims to find and quantify the dependencies within and between these two dimensions. This dissertation focuses on methods for areal data and discrete time structure, introducing models tailored for the needs of the applications and estimated with Bayesian framework.

The models are applied to study the Finnish grain markets during the 1860s famine, the spread of infectious diseases in Finland during the 18^{th} and 19^{th} centuries, and the adaptation of Finnish fathers to parental leave policies during 2009–2017. Spatiality is covered, for example, by conditional and simultaneous autoregressions, and time, for example, by random walk and autoregressive model, error correction model, and hidden Markov model. Their combinations also allow for the interaction of space and time. The methods are constructed to enable application-driven analysis without setting additional limits on it. As a consequence, the structures to model and the models are complicated. Nevertheless, the results are interpretable.

Keywords: autoregression, Bayesian model, conditional autoregression, error correction model, hidden Markov model, simultaneous autoregression, spatial data, spatio-temporal model, time series

TIIVISTELMÄ

Pasanen, Tiia-Maria
Alueellisen aineiston spatio-temporaalinen Bayes-mallintaminen
Jyväskylä: University of Jyväskylä, 2025, 52 s. (+artikkelit)
(JYU Dissertations
ISSN 2489-9003; 864)
ISBN 978-952-86-0446-4 (PDF)

Aineistoa voidaan pitää spatio-temporaalisena, kun se kuvaa sekä ilmiöön liittyvää sijaintia että aikaa. Spatio-temporaalisten tilastollisten menetelmien tarkoituksena on etsiä ja kvantifioida näiden ulottuvuuksien sisäisiä ja välisiä riippuvuuksia. Tämä väitöskirja keskittyy alueellisen aineiston ja diskreetin aikarakenteen menetelmiin esitellen malleja, jotka on yksilöity sovellusten tarpeisiin ja estimoitu bayesiläisittäin.

Malleilla tutkitaan Suomen 1860-luvun nälänhädän aikaisten viljamarkkinoiden toimintaa, tartuntatautien leviämistä Suomessa 1700- ja 1800-luvuilla sekä suomalaisten isien sopeutumista vanhempainvapaapolitiikkaan vuosina 2009–2017. Spatiaalisuus otetaan huomioon esimerkiksi ehdollisen ja simultaanisen autoregression avulla, kun taas aika huomioidaan esimerkiksi satunnaiskulun ja autoregressiivisen mallin, virheenkorjausmallin sekä piilo-Markov-mallin avulla. Näiden rakenteiden yhdistelmät mahdollistavat myös paikan ja ajan interaktion. Menetelmät on pyritty rakentamaan siten, että analyysi on sovelluslähtöistä, eivätkä mentelmät rajoita sitä. Tämän seurauksena mallinnettavat rakenteet ja mallit ovat monimutkaisia. Tulokset ovat silti helposti tulkittavissa.

Avainsanat: aikasarja, autoregressio, Bayes-malli, ehdollinen autoregressio, piilo-Markov-malli, simultaaninen autoregressio, spatiaalinen data, spatio-temporaalinen malli, virheenkorjausmalli

Author

Tiia-Maria Pasanen
Department of Mathematics and Statistics
University of Jyväskylä
Finland

Supervisors

Dr. Jouni Helske
INVEST Research Flagship Centre
University of Turku
Finland

Dr. Harri Högmander
Department of Mathematics and Statistics
University of Jyväskylä
Finland

Docent Tarmo Ketola
Department of Forestry
University of Helsinki
Finland

Dr. Miikka Voutilainen
Department of History and Ethnology
University of Jyväskylä
Finland

Reviewers

Assistant Professor Paula Moraga
Division of Computer, Electrical and Mathematical Sciences and Engineering
King Abdullah University of Science and Technology
Saudi Arabia

Professor Mikko Sillanpää
Department of Mathematical Sciences
University of Oulu
Finland

Opponent

Professor Katja Ickstadt
Department of Statistics
TU Dortmund University
Germany

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Jouni Helske, Harri Högmander, Tarmo Ketola, and Miikka Voutilainen for sharing their knowledge, guiding me, and giving me the opportunity to complete a PhD under their supervision. All of you have taught me a lot and showed new aspects of the world. Jouni and Harri deserve special recognition for showing me patiently the way through countless statistical and other problems over the years.

I also wish to thank Tarmo and Miikka for funding through the Finnish Cultural Foundation (Host, pathogen & environment—disease triangle in pre-health care Finland) and Jouni for funding through the Research Council of Finland grant 331817 (Towards well-informed decisions: Predicting long-term effects of policy reforms on life trajectories (PREDLIFE)). For funding I would also like to thank Emil Aaltonen Foundation and the Department of Mathematics and Statistics. I also want to acknowledge CSC, IT Center for Science in Finland, for computational resources.

I wish to express my gratitude for Assistant Professor Paula Moraga and Professor Mikko Sillanpää for being pre-examiners, and for Professor Katja Ickstadt for serving as the opponent. Thank you for your time and your comments.

Most of the research of this dissertation is conducted with my supervisors, but there are also other people who have played an important role in this project: Satu Helske, Simon Chapman, and Giovanni Giuliani. Thank you for the collaboration and your enthusiasm.

I am grateful for my fellow PhD students, my colleagues, and the personnel at the Department of Mathematics and Statistics for all their support and for making me feel welcome. I also wish to thank Sara Taskinen for commenting the dissertation manuscript, and Jukka Nyblom for numerous discussions.

Finally, I wish to sincerely thank my family and my friends for their extensive support and understanding during this project. No words are enough to express your importance. I would not be here without you.

Jyväskylä, October 2024
Tiia-Maria Pasanen

LIST OF FIGURES

LIST OF INCLUDED ARTICLES

I Pasanen, T.-M., Voutilainen, M., Helske, J., and Högmander, H. A Bayesian spatio-temporal analysis of markets during the Finnish 1860s famine. *Journal of the Royal Statistical Society Series C: Applied Statistics, Volume 71, Issue 5, Pages 1282–1302*, 2022. DOI: https://doi.org/10.1111/rssc.12577.

II Pasanen, T.-M., Helske, J., Högmander, H., and Ketola, T. Spatio-temporal modeling of co-dynamics of smallpox, measles, and pertussis in pre-healthcare Finland. *PeerJ 12:e18155*, 2024. DOI: https://doi.org/10.7717/peerj.18155.

III Pasanen, T.-M., Helske, J., and Ketola, T. Hidden Markov modelling of spatio-temporal dynamics of measles in 1750–1850 Finland. *Preprint*, 2024. DOI: https://doi.org/10.48550/arXiv.2405.16885.

IV Pasanen T.-M., Helske, S., Giuliani, G. A., Chapman, S. N., and Helske, J. Adaptation to paternal leave policies in Finnish municipalities: changing gender norms and cross-border policy legacies. *Preprint*, 2024. DOI: https://doi.org/10.31235/osf.io/k27yw.

The author of this dissertation has actively contributed to the research of the joint Articles I–IV. Together with the co-authors, the author has contributed to the development of the methods, the formulation of the research questions, the experiments, and the models, and the writing of Articles I–IV. The author has been responsible for the implementation of the models and simulations, and for the production of the visualizations and tables for Articles I–IV, with the exception of the Party Manifesto analysis and the justification for excluding election votes for the Swedish People's Party in the main analysis in Article IV.

CONTENTS

1 INTRODUCTION

In everyday life, we often encounter phenomena, depending on both time and location simultaneously. Possibly one of the most obvious ones is the weather. On areas close to each other there are similar wider patterns, but also local differences from those occur. Additionally, the weather changes a lot in time, sometimes quickly. The same kind of structures can be seen in epidemiology where infectious diseases spread through social contacts so that epidemics appear and fade away. In econometric settings, the prices of products may vary in time, for example, due to seasons, and between different stores, towns, or countries. These are merely a few examples. What all these phenomena have in common is that they typically consist of complex dynamics which we might not know but would like to find out. Sometimes we need to be able to explain things, for example, the reasons for epidemics. At times, the prediction is a more important aspect, as is the case with weather forecasts or when finding a balance between the supply and demand of products.

In statistics, such phenomena are called spatio-temporal due to their relation to both space and time. The methodology applied to study them is accordingly called spatio-temporal. The statistical tools in this dissertation are aimed for the exploration rather than the prediction of the phenomena. Many of the approaches applied here include complex models which are computationally expensive and slow. The execution time possibly lasting for days is often impractical or even unbearable. Nevertheless, the historical applications have their own features, being part of the past, and the slowness of research considering people who died hundreds of years ago is fortunately merely an inconvenience in this case.

Article I of this dissertation focuses on the Finnish markets during the 1860s famine. They have earlier been studied with the error correction model, introduced in section 2.2, by Ó Gráda (2001). Compared with Article I, in the work of Ó Gráda the data are more aggregated, the famine peak in 1867–1868 is treated as a dummy effect, and the analysis is done with several separate error correction models that are compared with each other using a reference location. However, in the current study, there are 80 administrative districts instead of the eight Ó Gráda used, thus the comparisons and testing related to the general error cor-

rection model would become exhaustive. As a response to the challenges, the traditional error correction model is generalized to cover dozens of spatially dependent time series simultaneously without any reference locations. The error correction terms are allowed to depend on time which relieves us of the need to predefine any specific periods for the famine peaks. The results reveal novel details of Finnish markets during the 1860s.

As for Articles II and III, the application field lies in epidemics. Diseases are clearly a matter of time and location, as have been painfully tangibly witnessed during the COVID-19 pandemic of recent years, which have been excessively studied in a short time (e.g, Giuliani et al., 2020; Jia et al., 2020). There are a lot of data considering this new infection, partly due to the large number of inhabitants populating the earth. The data are more reliable and organized than centuries ago, and quickly available since globalization has changed the speed at which not only people and infections but also information can travel. The data collected into Finnish church archives during the 18^{th} and 19^{th} centuries are still not completely digitized, and even if they were, they would not be as reliable or comprehensive as, for example, most of the COVID-19 data. The amount of the past data is also smaller due to the smaller population. Despite the drastic differences between then and now, there have been and still are sparse populations in which tracking of the routes of spread is not straightforward.

The Finnish historical archive data of the 18^{th} and 19^{th} centuries, on which the data analyzed in Articles II and III are based, have been studied earlier using a different methodology, aggregation level, and subsets of the data (e.g., Briga et al., 2021, 2022; Ketola et al., 2021). The novelty of Article II is studying the co-dynamics of several diseases utilizing sparse and scarce data. Article III, in turn, introduces a new way to extract information from this kind of spatio-temporal data consisting of a considerable number of locations and times. It intertwines the spatial dependencies with the temporal dependencies via intrinsic conditional autoregression, discussed in section 3.1, and hidden Markov model, discussed in section 4.3, thus offering a smooth way to illustrate the main patterns of the phenomenon in space and time.

In Article IV, the interest is to study the differences between Finnish municipalities using the parental leave quota of fathers during 2009–2017. These data are contemporary and can be seen as reliable and comprehensive in contrast to the other datasets used in the previous articles. The novelty of this research is to apply spatio-temporal methodology to learn if there are spatial patterns in parental leave uptakes and if they evolve over time or can be explained with demographic background information.

The rest of this summary offers the basics to understand the included articles and is structured as follows. In the second chapter, the general frameworks of spatial and temporal data are introduced since the spatio-temporal data are a combination of both these dimensions. Also, as historicity plays an important role in the applications, some important characteristics of such data are discussed. The third chapter describes some statistical methods for spatial modeling, namely the conditional autoregression and its derivative, the intrinsic conditional autore-

gression, as well as the spatial autoregressive model. The fourth chapter focuses on temporal models such as the autoregression and its generalization. Also, the error correction and hidden Markov models are covered. In the fifth chapter, spatial and temporal models are combined for simultaneous usage. The sixth chapter covers a brief description of the Bayesian approach in statistics. The seventh chapter consists of the summaries of the articles included in this dissertation. Finally, everything is brought together in a discussion in the last chapter.

2 SPATIO-TEMPORAL PHENOMENA AND DATA

This chapter introduces the concepts of spatio-temporal phenomena and data, as well as some common mathematical notations related to them. We go through separately spatial and temporal phenomena, since their combination results in the spatio-temporal nature we are interested in. Whereas this chapter introduces the data, the processes generating such data are introduced in chapters 3, 4, and 5. As the articles of this dissertation are highly involved with past data, the end of the chapter also describes some aspects that are often present in historical datasets.

2.1 Spatial data

In this spatial context, space and spatial are terms referring to locations, for example, geographical regions on a map or locations of nucleotides in DNA. The essential thing with this kind of space or regional entity is that it can be divided into smaller subregions or locations that are connected to each other by some account. The division of the space into these smaller parts is also sometimes called a tessellation.

The interest in spatial analysis often is whether there are dependencies between the locations such that the phenomenon behaves similarly in places close to each other. For example, the weather is more likely alike in nearby areas than those far from each other due to physical factors, a prevalence of an infection in one municipality might resemble that of neighboring municipalities because of social and biological connections, or the adjacent sequences of nucleotides or genes in DNA might have biological connections. In addition to elucidating the structure of the phenomenon, the spatial dependencies found can also be utilized to predict the behavior of a phenomenon in locations where it has not been observed, for example, due to missing data.

For a formal definition, denote each location or site of the complete space with $i = 1, \ldots, N$, where N is the total number of the locations. The observations

concerning each location i are denoted with y_i. In order to consider and identify the dependencies between the locations, their neighbors have to be acknowledged. The neighborhood of the location i, i.e., the collection of the neighboring locations, is denoted with J_i, and each neighbor of the location i is denoted with $j = 1, \ldots, N_i$, where N_i is the number of the neighbors of the location i. Sometimes the location i is also referred to as the focal location to distinguish it from the neighbors.

Naturally, also the concept of neighborhood has to be defined for spatial approaches. Different kinds of definitions may be used depending on the situation, maybe the most common ones being shared border and distance between centroids of the locations. These both have their advances and drawbacks. Shared borders or short distances do not guarantee a communication route between the locations in geographical settings, nor do unshared borders or long distances guarantee the lack of those routes. During historical times, waterways were popular connections despite the possible long distances, whereas short distances along difficult terrain most likely did not result in any contact. A neighborhood definition based on road networks might be useful with regard to phenomena that are related to the interaction of people. Distance, taking into consideration the landforms, could be more suitable for phenomena such as weather that are not based on human communication.

The neighborhood or the closeness of the locations is additionally described with the order of the neighbors. The first order neighbors are the ones having a common borderline or being otherwise the closest ones, the second order neighbors are the first order neighbors of the first order neighbors, and so forth, see Figure 1 for an illustration. Usually, a location cannot be its own neighbor. Sometimes it requires considering neighbors of several orders to capture the true spatial dependencies. In this dissertation, all the articles utilize the first order neighborhood based on border sharing.

2.2 Time series

A time series is a sequence of repeated observations of one variable in time. Some examples of time series are a monthly mortality rate in a municipality, a daily price of a certain product, and an hourly temperature in a location. An example illustration of a time series can be seen in Figure 2 which represents the numbers of deaths in one Finnish municipality during 1750–1850.

The time resolution of the observations may be dense, as several measurements in a second, or more sparse, as weekly, monthly, or yearly measurements. The intervals between the observations are usually the same length but also other considerations are possible. The structure of time series can be modeled with discrete or continuous time frameworks, depending on which of these the time is seen in the application. This dissertation focuses on discrete time series with constant observation intervals. Often with time series analysis, an interesting ques-

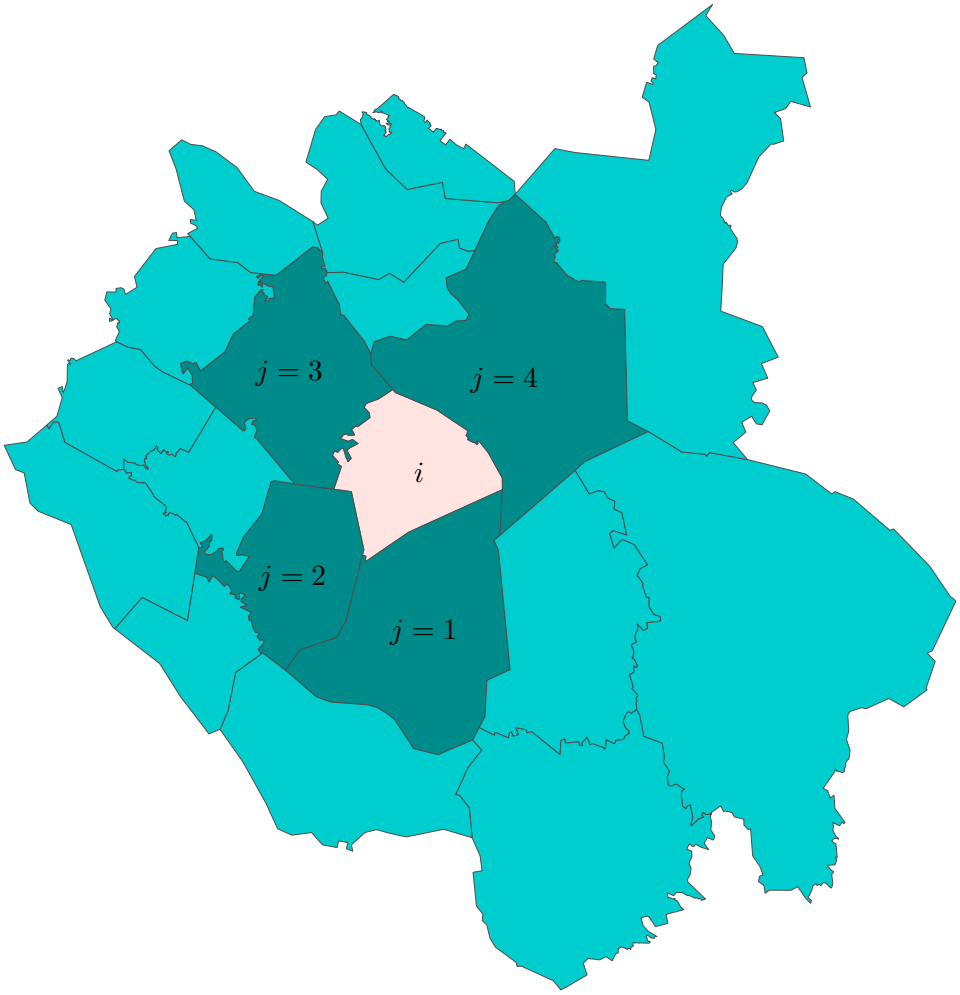

FIGURE 1 An example of spatial division. The focal location, denoted with i, is in the middle. The four first order neighbors of it are colored with dark turquoise. The rest of the locations, colored with light turquoise, are the second order neighbors of the focal location.

tion is whether there are some temporal dependency structures within the time series, for example, a long-term trend, or some shorter-term dependencies such as daily, weekly, monthly, or yearly periodicities.

Time series are commonly denoted with observations y_t, where $t = 1, \ldots, T$ indicates the time point of the observation y, and T is the total number of the time points. If the observations indicate dependency, it is called autocorrelation, which refers to correlating with oneself. The observation may correlate only with the observation of the closest time point before (or after), or the effect might reach over longer times. This is somewhat comparable with the order of the neighborhood in the previous section.

Naturally, there can be multiple time series simultaneously instead of a plain one. Those time series might consider different variables being measured from one statistical unit using the same set-up and covering the same time span, or consider the same thing being measured from different statistical units at the same time. The combination of both of these perspectives is also possible. The

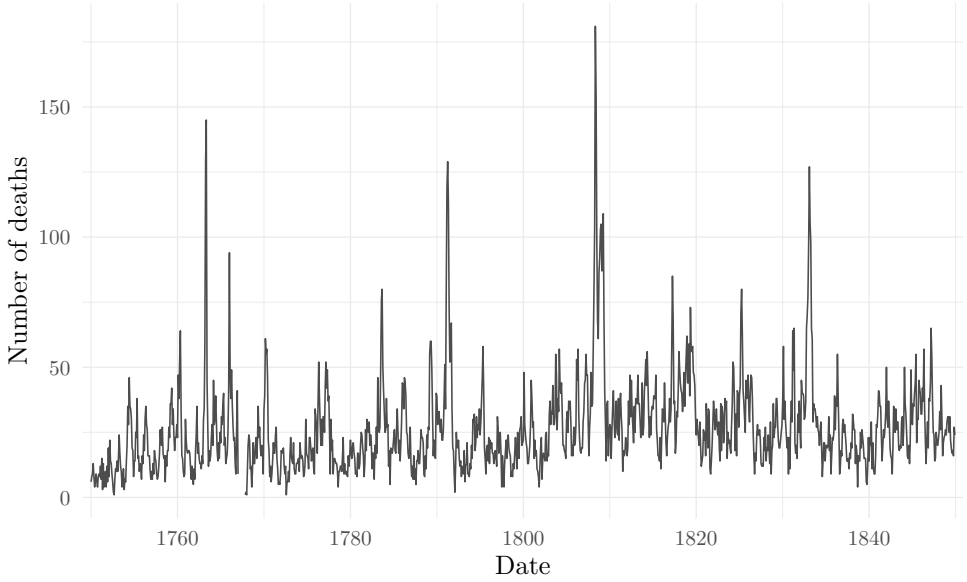

FIGURE 2 An example of time series. The black line represents the number of monthly deaths in the area of Kuopio during 1750–1850 according to the Finnish church book archive data used in Article III.

statistical units can be, for example, people, stores, cities, or countries. In these cases of several time series, not only the dependencies within a single time series but also those shared between the time series are usually considered interesting. If the statistical units happen to be spatial locations, we have spatio-temporal data, and a spatio-temporal model should be considered for the analysis.

2.3 Special features of historical data

While almost all data can be seen as historical in some sense, here we discuss information dating back centuries. These data have been collected a long time ago by people with the knowledge and skills of that time. Not everyone had the ability to write things down due to a lack of skills, opportunity, or proper equipment. Furthermore, those with the necessary prerequisites most likely did not see the world the same way people do today due to the huge increase in the amount of information. For example, the people determining the causes of death in Finland during the 18^{th} and 19^{th} centuries were usually priests with little or no medical training, not to mention the prevailing lack of modern medicine (see, e.g., Vuorinen, 1999). In other words, some of the records may not be valid in the light of current knowledge. This is especially unfortunate when the data is related to a phenomenon that does not exist anymore or has changed dramatically. For example, smallpox has been successfully eradicated, thus data describing its spread cannot be gained from the modern world.

Additionally, the data were usually recorded for current purposes, such as collecting demographic information for the needs of the ruler and for improving the kingdom or empire (see, e.g., Suomen Sukututkimusseura, 2024). Thus, the

data were not collected nor organized for any kind of modeling or more general distant future use as, for example, the epidemiological research performed in Articles II and III. Had the people known the objectives waiting ahead, they would most likely have paid attention and care treating the data.

Another aspect of historical data is their incomplete nature, which likely is a permanent feature. Of course, not even the contemporary data should be regarded as comprehensive without careful consideration, but the exclusions made in the old days might be unrecoverable today. The Finnish church archives mentioned above do not cover the whole population but exclude, for example, those not known being members of any specific parish (see, e.g., Pitkänen, 1977). The excluded people might be only a small proportion of the total population in this particular case, but not necessarily always. In some lucky instances, the gaps following the exclusions might be filled using other complementary data sources. In general, the imperfect nature of historical data should not be forgotten.

While the collection and recording procedures of the data were somewhat unreliable at times, the preservation process came with its own shortcomings. Losing archives in fires or other unfortunate conditions occurred once in a while. On some occasions, the data were rewritten or modified during their preservation, leading to additional errors. For example, unclear handwriting, changes in currencies, or clerical errors may have caused trouble. Since carbon copies rarely exist, there are few means to discover the flaws.

Nowadays, the digitization of the data brings several sources of information into daylight, enabling and streamlining their use. Until all archives are available in digital format, the digitized data are selected. In the case of the Finnish church archives, the records have been digitized by volunteers (Suomen Sukututkimusseura, 2024), often amateur genealogists researching their own family histories. Consequently, the digital archives cover only those original sources that have been considered by someone. The digitization is still a work in progress, and until it is finished, the data are not complete despite their decent representativeness.

Missing data are a feature commonly present in data regardless of the historical or contemporary nature of them. In a historical context, the percentage of missing observations can be exhaustive, and the inability to collect any appropriate new data or correct the information only emphasizes the problem. For example, in the Finnish church book data used in Article III, the percentage of missing observations is at worst roughly 70% at the beginning of the 1750s, and while it decreases steadily over time, it is still around 20% at best in the 1840s, as may be seen from Figure 3. Such a level of missing data limits the selection of applicable methods.

Nevertheless, sometimes there is no option but to find a way to work with these historic, incomplete, and occasionally excessively missing data. The expertise of a professional familiar with the research context, combined with statistical methods capable to function despite the missing information, are important factors when overcoming the problems stated above. Integrating information from several origins might help not only with the omitted aspects of the phenomenon

but also with data missing for other reasons. All in all, the premises set by historic data raise a need for the use of complex yet flexible statistical methods, such as hierarchical, multilevel, or Bayesian models. This is even more pronounced when analyzing simultaneously several aspects of data, for example, spatial and temporal dimensions.

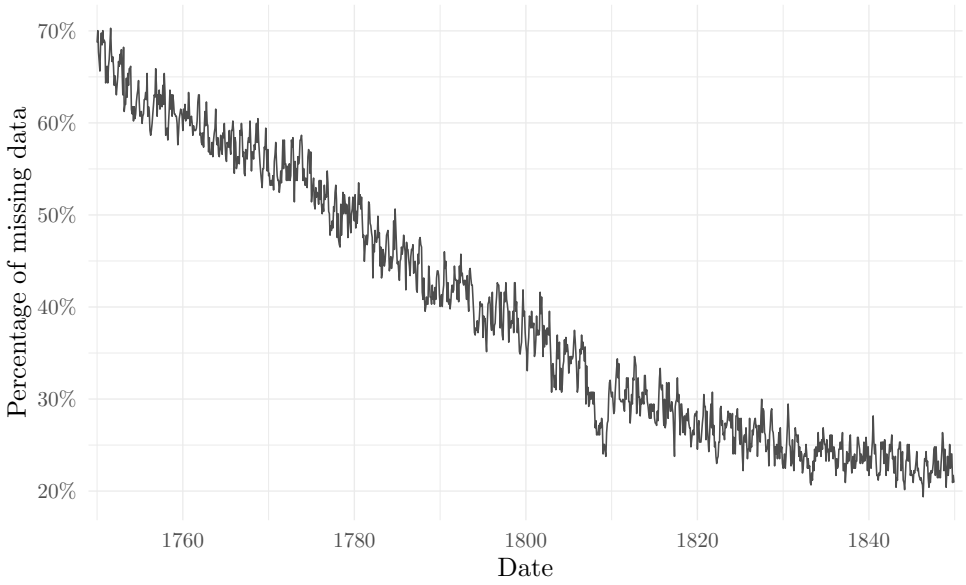

FIGURE 3 The percentage of missing data during 1750–1850 in the Finnish church book archive data used in Article III.

3 SPATIAL MODELS

Spatial phenomena and data often require special consideration of dependencies between the locations. There are several approaches to that, and this chapter introduces the basic theory of conditional autoregression and its intrinsic modification, as well as simultaneous autoregression, which all are spatial models. In order to account for the spatial dependency of locations, their adjacency has to be defined. As mentioned in section 2.1, the neighborhood may be defined in various ways. In what follows, two locations are assumed to be neighbors if they share a borderline.

3.1 Conditional autoregression and its modifications

The conditional autoregressive (CAR) model (Besag, 1974) is a statistical method for accounting spatial dependency between adjacent areas. It is a special case of Markov random field and based on the idea of using multiple conditionally independent univariate distributions to gain a multivariate joint distribution for modeling the correlation between several variables (Held and Rue, 2010; Banerjee et al., 2015). The method quantifies the amount of dependency without specifying the direction of the effect so that no causal conclusions can be made based only on the CAR model.

As in the previous chapter, denote each location with $i = 1, \ldots, N$, and its neighbor with corresponding indexing $j = 1, \ldots, N$. As a location cannot be its own neighbor, the maximum number of neighbors of any individual location is $N - 1$. The neighborhood of each location i is denoted with J_i. Now, based on the neighborhood definition, construct an $N \times N$ proximity or adjacency matrix, W, so that it is symmetric and its elements are

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

In other words, the elements $w_{i,j}$ indicate that the two locations, i and j, are neigh-

bors. Instead of the plain binary indicators of neighborhood, the adjacency matrix could also hold some other weights for each pair of locations. The numbers of neighbors of each location i, denoted with N_i, in turn, are collected in another $N \times N$ matrix, D. The elements of this diagonal matrix are

$$d_{i,j} = \begin{cases} N_i, \text{ when } i = j, \\ 0, \text{ otherwise.} \end{cases}$$

To account for the actual intensity of the spatial dependency between the locations, we employ an N-dimensional random variable $\boldsymbol{\varphi}$. For these spatial variables, we assign a Gaussian distribution according to the most common approach when defining a CAR model. Alternatively, the Gaussian approach is applicable, for example, with an autologistic model using binary variables with logistic link function, or an auto-Poisson model for count variables using log-linear link function (see, e.g., Held and Rue, 2010; Banerjee et al., 2015). In Cressie (1993), also, an autogamma approach is brought up.

The full conditional distribution of the Gaussian spatial variable $\boldsymbol{\varphi}$ is

$$\varphi_i | \boldsymbol{\varphi}_{-i} \sim \mathrm{N}\left(\alpha \sum_{j=1}^{N} \tilde{w}_{i,j} \varphi_j, \, \tilde{\sigma}_i^2 \right), \tag{1}$$

where $\boldsymbol{\varphi}_{-i}$ denotes a vector without the element considering the location i, and the parameter $\alpha \in (0, \, 1)$ describes the amount of spatial dependency. The notations with tilde represent scaled variables $\tilde{w}_{i,j} = w_{i,j}/N_i$ and $\tilde{\sigma}_i^2 = \sigma_i^2/N_i$. Thus, each variable φ_i follows a Gaussian distribution conditioned on the corresponding variables of other locations, or in other words, φ_j, where $j \neq i$. The expected value is a constant proportion α of the weighted sum of the corresponding variables of other locations, and the variance is an unknown parameter. The unscaled variance σ_i^2 is often fixed to be constant over the locations i, which is also done here. Equation 1 defines a Gaussian Markov random field (see, e.g., Rue and Held, 2005), offering a smooth spatial distribution for the variables φ_i.

The joint distribution based on the conditional distributions can be derived using Brook's Lemma (Brook, 1964), which states that if $p(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in R^N$ in the support of $p(\boldsymbol{x})$, then for any two variables $\boldsymbol{x}$ and $\boldsymbol{y}$ in the support we have

$$\begin{aligned} \frac{p(\boldsymbol{x})}{p(\boldsymbol{y})} &= \prod_{i=1}^{N} \frac{p(x_i | y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_N)}{p(y_i | y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_N)} \\ &= \prod_{i=1}^{N} \frac{p(x_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_N)}{p(y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_N)}. \end{aligned} \tag{2}$$

If we now let $\boldsymbol{x} = \boldsymbol{\varphi}$ and $\boldsymbol{y} = \mathbf{0} = (0_1, \ldots, 0_N)^\top$, and we apply Brook's Lemma,

we get

$$
\begin{aligned}
\frac{p(\boldsymbol{\varphi})}{p(\mathbf{0})} &= \prod_{i=1}^{N} \frac{\exp\left(-\frac{1}{2\tilde{\sigma}_i^2\pi}\left(\varphi_i - \alpha\sum_{j<i}\tilde{w}_{i,j}\varphi_j - \alpha\sum_{i<j}\tilde{w}_{i,j}0_j\right)^2\right)}{\exp\left(-\frac{1}{2\tilde{\sigma}_i^2\pi}\left(0_i - \alpha\sum_{j<i}\tilde{w}_{i,j}\varphi_j - \alpha\sum_{i<j}\tilde{w}_{i,j}0_j\right)^2\right)} \\
&= \prod_{i=1}^{N}\exp\left(-\frac{1}{2\tilde{\sigma}_i^2\pi}\left(\varphi_i - \alpha\sum_{j<i}\tilde{w}_{i,j}\varphi_j\right)^2 + \frac{1}{2\tilde{\sigma}_i^2\pi}\left(\alpha\sum_{j<i}\tilde{w}_{i,j}\varphi_j\right)^2\right) \\
&= \exp\left(-\sum_{i=1}^{N}\frac{1}{2\tilde{\sigma}_i^2\pi}\left(\left(\varphi_i - \alpha\sum_{j<i}\tilde{w}_{i,j}\varphi_j\right)^2 - \left(\alpha\sum_{j<i}\tilde{w}_{i,j}\varphi_j\right)^2\right)\right) \\
&= \exp\left(-\sum_{i=1}^{N}\frac{1}{2\tilde{\sigma}_i^2\pi}\left(\varphi_i^2 - 2\varphi_i\alpha\sum_{j<i}\tilde{w}_{i,j}\varphi_j\right)\right) \\
&= \exp\left(-\sum_{i=1}^{N}\frac{1}{2\tilde{\sigma}_i^2\pi}\varphi_i^2 + \alpha\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{2\tilde{\sigma}_i^2\pi}\varphi_i\tilde{w}_{i,j}\varphi_j\right) \\
&= \exp\left(-\frac{1}{2\pi}(\boldsymbol{\varphi}-\mathbf{0})^\top\tilde{\Sigma}^{-1}(I-\alpha\tilde{W})(\boldsymbol{\varphi}-\mathbf{0})\right),
\end{aligned}
\tag{3}
$$

where $\tilde{\Sigma}$ is a diagonal matrix with elements $\tilde{\sigma}_i^2$ on its diagonal, the entries of the matrix $\tilde{W}$ are the scaled variables $\tilde{w}_{i,j}$, and I denotes an $N \times N$ identity matrix. The fifth equality follows from the symmetry assumption that $\tilde{w}_{i,j}/\tilde{\sigma}_i^2 = \tilde{w}_{j,i}/\tilde{\sigma}_j^2$. Now, the joint distribution of the last line can also be written as

$$
\boldsymbol{\varphi} \sim \mathrm{N}(\mathbf{0},\ \tilde{\Sigma}(I-\alpha\tilde{W})^{-1}). \tag{4}
$$

A more general representation is

$$
\boldsymbol{\varphi} \sim \mathrm{N}(\mathbf{0},\ Q^{-1}), \tag{5}
$$

where Q denotes the precision matrix which has to be symmetric and positive definite in order to the distribution to have a proper joint probability density. Above, these features have been achieved with the weights and deviations scaled by the number of neighbors of each location. A convenient way to express the precision matrix under our conditions is to use the matrices W and D and set

$$
Q = \tau D(I - \alpha D^{-1}W) = \tau(D - \alpha W), \tag{6}
$$

where $\tau = \frac{1}{\sigma^2} \in \mathbb{R}^+$ is a precision parameter of the variables φ_i. Using the definition for Q as in Equation 6, the probability density function of the Gaussian distribution of Equation 5 can be written as

$$
p(\boldsymbol{\varphi}) = (2\pi)^{N/2}\left|(\tau(D-\alpha W))^{-1}\right|^{1/2}\exp\Big(-\frac{\tau}{2}\boldsymbol{\varphi}^\top(D-\alpha W)\boldsymbol{\varphi}\Big). \tag{7}
$$

Sometimes we need to account for a significant amount of spatial correlation or allow wide spatial patterns in the posterior distribution. In these cases, the parameter α most likely draws close to one (Banerjee et al., 2015). If the parameter α in Equation 6 is set to 1, the model is called an intrinsic conditional autoregressive (ICAR, sometimes also IAR) model. On the one hand, this results in an improper distribution, which cannot be used as a model for the data, but which, on the other hand, may be utilized as a prior distribution for spatial dependency, as is done for example in the Besag York Mollié (BYM) models (Besag et al., 1991).

The ICAR model is often represented in a pairwise form, which is especially profitable in the computational sense (see, e.g., Morris et al., 2019). The pairwise formula can be derived from Equation 7 with some algebra represented next. Keeping in mind that the parameter $\alpha = 1$ now that we are discussing the ICAR model, we concentrate on the modification of the component inside the exponential function. In what follows, the matrix form is broken down into sum representation and then the terms are rearranged according to the properties of the matrices W and D:

$$\begin{aligned}
-\frac{\tau}{2}\boldsymbol{\varphi}^\top (D - W)\,\boldsymbol{\varphi} &= -\frac{\tau}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\varphi_i\left(d_{i,j} - w_{i,j}\right)\varphi_j \\
&= -\frac{\tau}{2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\varphi_i d_{i,j}\varphi_j - \sum_{i=1}^{N}\sum_{j=1}^{N}\varphi_i w_{i,j}\varphi_j\right) \\
&= -\frac{\tau}{2}\left(\sum_{i=1}^{N}\varphi_i^2 d_{i,i} - \sum_{i=1}^{N}\sum_{j\in J_i} 2\varphi_i\varphi_j\right) \\
&= -\frac{\tau}{2}\left(\sum_{i=1}^{N}\sum_{j\in J_i}\left(\varphi_i^2 + \varphi_j^2\right) - \sum_{i=1}^{N}\sum_{j\in J_i} 2\varphi_i\varphi_j\right) \\
&= -\frac{\tau}{2}\sum_{i=1}^{N}\sum_{j\in J_i}\left(\varphi_i^2 - 2\varphi_i\varphi_j + \varphi_j^2\right) \\
&= -\frac{\tau}{2}\sum_{i=1}^{N}\sum_{j\in J_i}\left(\varphi_i - \varphi_j\right)^2.
\end{aligned} \tag{8}$$

In the third equality, the first sum follows from the fact that the matrix D is diagonal and thus $d_{i,j} \neq 0$ only when $i = j$. The second sum, on behalf, is the consequence of the matrix W being symmetric and its entries being either zeros or ones depending on whether the locations i and j are adjacent or not. The fourth equation relies on the fact that the diagonal elements $d_{i,i} = N_i$ state the numbers of the neighbors, and the desired number of repetitions of each variable φ_i can be achieved by going through the whole neighborhood J_i.

Now, the intrinsic version of Equation 7 takes the form

$$p(\boldsymbol{\varphi}) = (2\pi)^{N/2}\left|\left(\tau(D - W)\right)^{-1}\right|^{1/2}\exp\left(-\frac{\tau}{2}\sum_{i=1}^{N}\sum_{j\in J_i}\left(\varphi_i - \varphi_j\right)^2\right), \tag{9}$$

which is also called the pairwise form. When looking at this pairwise formula, one can see that adding any constant to all the variables φ_i does not change their density. This causes identification problems, which, however, can be overcome using some constraints. One of the most common remedies is to restrict the variables to sum to zero over all the locations.

ICAR is not the only modification of CAR, but the constraints and assumptions placed on the original CAR model can lead to several different kinds of alternatives. Some of these derivatives are briefly introduced and applied in Vicente et al. (2020).

3.2 Simultaneous autoregression

Another model for accounting for the spatial dependency of adjacent areas is the simultaneous autoregressive (SAR) model which was originally introduced by Whittle (1954). SAR models are related to the CAR models in the sense that any SAR model can be presented as a CAR model, while the reverse is not true. In what follows, we go through the basic theory of SAR models based on Ord (1975) and Banerjee et al. (2015).

Consider the notation of the sites i and their neighbors j as with the CAR model above. Denote the spatially structured variable of length N with $\boldsymbol{y}$, and its each element with y_i, $i = 1, \ldots, N$. Assign a weight $b_{i,j}$ to quantify the relation of the adjacent variables y_i and y_j. This relation does not have to be symmetric, but in general, we allow the inequality $b_{i,j} \neq b_{j,i}$. The $N \times N$ matrix collecting all the weights is denoted with B.

Now, let

$$y_i = \sum_{j \in J_i} b_{i,j} y_j + \varepsilon_i, \tag{10}$$

where the independent error terms have a Gaussian distribution: $\varepsilon_i \sim \mathrm{N}(0,\, \sigma_i^2)$. In other words, the spatial variable y_i of the location i is the weighted sum of the corresponding variables of its neighboring locations accompanied with some white noise error. In matrix form Equation 10 can be written as

$$\begin{aligned} & \boldsymbol{y} = B\boldsymbol{y} + \boldsymbol{\varepsilon} \\ \iff \quad & (I - B)\boldsymbol{y} = \boldsymbol{\varepsilon} \\ \iff \quad & \boldsymbol{y} = (I - B)^{-1}\boldsymbol{\varepsilon}, \end{aligned} \tag{11}$$

where I denotes an identity matrix of an appropriate size, here $N \times N$, and the N-length vector of errors is again Gaussian, that is $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0},\, \Sigma)$ where

$$\Sigma_{i,j} = \begin{cases} \sigma_i^2, \text{ when } i = j, \\ 0, \text{ otherwise.} \end{cases}$$

The variances σ_i^2 are unknown.

Utilizing the above notation, Equation 11 can be rewritten as a multivariate Gaussian distribution as

$$\boldsymbol{y} \sim \mathrm{N}\Big(\mathbf{0},\ (I-B)^{-1}\Sigma((I-B)^{-1})^{\top}\Big). \tag{12}$$

The notation of Equation 12 is also equivalent with Equation 10. Furthermore, if we assume that the variance σ_i^2 is the same for all the locations i, meaning that $\Sigma = \sigma^2 I$, the distribution simplifies to

$$\boldsymbol{y} \sim \mathrm{N}\Big(\mathbf{0},\ \sigma^2(I-B)^{-1}((I-B)^{-1})^{\top}\Big). \tag{13}$$

Since the relation between the locations i and j does not have to be symmetric, also the matrix $I - B$ may be non-symmetric. Nevertheless, in order for the distribution in Equation 13 to be proper, the matrix $I - B$ has to fulfill some conditions, more specifically, it has to be full rank, i.e., non-singular. In literature, there are two common alternatives to guarantee this propriety.

In the first option, we may set the matrix $B = \rho W$, where ρ is a spatial autoregression parameter which has to belong to the interval $(\frac{1}{\lambda_1}, \frac{1}{\lambda_N})$, where λ_1 and λ_N are the smallest and the largest eigenvalues of the matrix W, respectively. The matrix W, in turn, is defined as an adjacency matrix as with CAR in section 3.1. As a result, the matrix $I - B$ is non-singular, and thus the distribution in Equation 13 is proper.

In the other common alternative we, again, set $B = \rho W$, but instead of finding eigenvalues and scaling with them, we normalize the adjacency matrix W with its rowsums. The elements of the normalized matrix, denoted with $\tilde{W}$, are

$$\tilde{w}_{i,j} = \frac{w_{i,j}}{\sum_{j=1}^{N} w_{i,j}}. \tag{14}$$

This leads the normalized adjacency matrix $\tilde{W}$ being non-symmetric. Due to the normalization, the eigenvalues of $\tilde{W}$ are between -1 and 1, so that it is natural to set $B = \tilde{\rho}W$, where $\tilde{\rho} \in (-1,\ 1)$ is called a spatial autocorrelation parameter. Again, the resulting matrix $I - B$ is non-singular. Note the similarity between this normalization and the one used with the CAR in the previous section.

As mentioned at the beginning of this section, SAR models are special cases of CAR models. The connection holds when both definitions yield proper distributions and the covariance of a SAR model equals the covariance of a CAR model. That is, using the notations of this and the preceding section:

$$Q^{-1} = (I-B)^{-1}\Sigma((I-B)^{-1})^{\top}. \tag{15}$$

However, it is noteworthy that the proper models employing the spatial dependency coefficients, α in the case of CAR and ρ in the case of SAR, do not necessarily imply the same dependency (Wall, 2004). Also, the CAR model uses a symmetric spatial dependency matrix whereas the SAR model works with asymmetric ones as well. This might seem beneficial for the SAR model, but actually it can lead to identifiability issues due to the estimation of the covariance matrix $(I-B)^{-1}\Sigma((I-B)^{-1})^{\top}$. Further comparisons between the models can be found, for example, from Cressie (1993) and Banerjee et al. (2015).

3.3 Spatiality accounted by explanatory variables

Sometimes we do not want to or we are not able to assign any specific pre-defined dependency structure for the spatial units, maybe due to irregular dependencies or due to the heterogeneous nature of the units. In these cases, the space can be acknowledged, for example, by including spatial explanatory variables. This leads to modeling the response with some covariates, which, in turn, can be done in numerous different ways.

One possible practice is to employ regression terms to tie the response together with the covariates. Denote the response variable with y_i and the explanatory variable with x_i, where $i = 1, \ldots, N$ stands for the location. Now, the neighborhood J_i of the location i can be utilized, for example, by saying that the probability of y_i depends on the neighboring values x_j via independent regression coefficients per neighbor pair and some function f so that

$$p(y_i) = f(\sum_{j \in J_i} \beta_{i,j} x_j). \tag{16}$$

Instead of individual neighbors, one can study the relation between the whole neighborhood and the response by setting the coefficients constants over the neighbors, that is $\beta_{i,j} = \beta_i$, in Equation 16. This might be convenient when the number of parameters to estimate is getting large, i.e., there are many locations and neighbors, or when there is a limited amount of information in the data.

This kind of approach is utilized, for example, in Article II. There the response indicates the local confirmed occurrence of a disease, and this occurrence is explained by local occurrences of that and other diseases. The model takes into consideration only the observations of the focal location and its closest neighbors, which serve as components for the spatial dependency. Some of the parameters to estimate are local and some are shared between the locations. The approach resembles mixed, i.e., multilevel, models (see, e.g., Gelman and Hill, 2007).

4 TEMPORAL MODELS

Time series are often seen as noisy observations of stochastic processes, and the analysis of them aims to find the underlying processes, or the actual phenomena, and to separate them from the errors. This chapter introduces some statistical methods for time series analysis relevant to this dissertation. First, we go through the temporal autoregression for one time series and then broaden the concept for multiple time series. Furthermore, we discuss the error correction model for analyzing several (econometric) time series. Finally, we introduce the hidden Markov model which makes a clear distinction between the observed and latent levels.

4.1 Autoregression and random walk

Autoregression as a term implies that something is regressed against itself. In spatial context, the local variable is modeled with or conditioned on its corresponding neighboring values, whereas in temporal context the temporal variable is modeled using the same variable from other time points. Usually in the literature Yule (1927) is credited for being the first to introduce the autoregressive model in temporal context (Klein, 1997). The theory represented in this chapter is based on Hamilton (1994) and Hyndman and Athanasopoulos (2021).

As in section 2.2, denote the variable of interest with y_t, where $t = 1, \ldots, T$ indicates the time point of the observation and T is the number of time points. The autoregressive process of order p, shortened often with AR(p), for one variable is

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t, \tag{17}$$

where the constant c is an intercept, the variable ϕ_i stands for an unknown model parameter reflecting the strength of temporal dependency, and the errors ε_t are independently and identically distributed white noise. We assume that the errors are Gaussian with zero-mean, $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. Thus, the order p indicates the num-

ber of previous values needed to account for the temporal dependency, and the expected value of each observation y_t consists of a constant c and the weighted sum of p preceding observations.

Depending on the values of the constant c and the autoregressive coefficients ϕ_i, the process has different interpretations. If we consider an AR(1) process, depending only on the previous observation, and the constant as well as the autoregressive coefficient are zero, the result reduces to pure white noise. Instead, if the autoregression parameter equals one while the constant is zero, the process is called a random walk. If the autoregression parameter is one and the constant differs from zero, we have a random walk with a drift.

The Gaussian random walk may also be presented in the form

$$y_t \sim \mathrm{N}(y_{t-1},\, \sigma^2). \tag{18}$$

As well as the general autoregression, also the random walk introduces a time series with large scale variation in the case of large deviation parameter σ^2 and small scale in the case of small deviation parameter.

The variable of interest, y_t, does not have to be one dimensional. In fact, the autoregressive process is directly expandable to a multivariate case, which is known as vector autoregression (VAR). Now, assume again that we have a sequence of observations from T consecutive time points and let the observations at each time point t be an N-dimensional vector $\boldsymbol{y}_t$. Straightforwardly, Equation 17 takes the form

$$\boldsymbol{y}_t = \boldsymbol{c} + \sum_{i=1}^{p} \Phi_i \boldsymbol{y}_{t-i} + \boldsymbol{\varepsilon}_t, \tag{19}$$

where the symbols in bold, $\boldsymbol{c}$ and $\boldsymbol{\varepsilon}$, are as before except that they are now N-dimensional vectors. The autoregressive coefficients form an $N \times N$ matrix Φ_i. The errors follow a Gaussian distribution: $\varepsilon \sim \mathrm{N}(\mathbf{0},\, \Sigma)$, where Σ is the covariance matrix. The errors are still assumed to be independent of time.

If one seeks a way to include one or more seasonal or cyclic components or other corresponding elements in the model, they can, for example, simply insert those as additional terms in Equation 17 or Equation 19. Seasonality may be accounted for, e.g., by dummy variables, indicating each seasonal moment of time, or by weighted lagged effects. Assume that we have a seasonal effect consisting of K components, for example, a monthly effect repeating every year would have $K = 12$ components. Denote each of these components with $k = 1, \ldots, K$, and the current component of time point t with k_t. Combined with Equation 17, the model with seasonal indicator functions without lagged effects would be

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \theta_{k_t} + \varepsilon_t, \tag{20}$$

where the unknown parameter θ_{k_t} is the seasonal effect. The alternative with lagged effects allowing different numbers of lags for the seasonal and non-seasonal effects, in turn, would be

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i'=1}^{p'} \theta_{k_{t-i'+1}} + \varepsilon_t, \tag{21}$$

where p is the number of non-seasonal lagged effects and p' is the number of seasonal lagged effects including the current effect. For a model with monthly effect and $p = p' = 1$ the expected value of the current observation consists of the constant, the previous observation, and the seasonal effect of the current month, which corresponds to Equation 20 with $p = 1$. There are several choices the user may make, thus these are merely examples of the inclusion of seasonal components. More information about seasonal effects in temporal models can be found, for example, in Metcalfe and Cowpertwait (2009), Box et al. (2016), and Hyndman and Athanasopoulos (2021).

4.2 Error correction model

The vector autoregression mentioned above is frequently used in econometrics in an applied form called the error correction model (ECM), introducing an error correction term to the VAR model. This chapter follows Alogoskoufis and Smith (1991), Kennedy (2003), and Lütkepohl (2005) in representing the model theory. To start with, we assume that we have at least two time series which are cointegrated, that is, they share a common underlying long-run relationship or trend. The term 'error' refers to a deviation from the long-run equilibrium, which, in turn, affects the short-run relationships. The error correction model describes a stochastic process of a dependent time series returning to a common equilibrium after changes in the other time series. The model is convenient for estimating the short-term and long-term dependencies between the time series.

With two sequences of observations x_t and y_t, where $t = 1, \ldots, T$, the general ECM may be written as

$$y_t - y_{t-1} = \alpha + \beta(x_t - x_{t-1}) + \gamma(y_{t-1} - x_{t-1}) + \varepsilon_t, \tag{22}$$

where α denotes the trend shared between the time series, β stands for the short-run effect of changes in the reference series x on y, and γ captures the long-term effect of the difference between the time series. The errors are commonly assumed to be Gaussian with mean zero and some unknown variance.

In the econometric context, the short-term and long-term coefficients are normally expected to be limited: $0 < \beta < 1$ and $-1 < \gamma < 0$. The short-term coefficient β reflects the momentary shared movements between the time series, whereas the coefficient γ is the error correction term adjusting the long-term equilibrium. In other words, the error correction term presents the share of error that is eradicated between consecutive time points, and the closer it is to -1 the larger the proportion of the difference fading away between time points t and $t - 1$ and the faster the return to the equilibrium.

Whereas the case of one equation is called the error correction model, the case of several equations is called the vector error correction model (VECM). This model arises when there are more than two cointegrated time series, i.e., the observations are vectors $\boldsymbol{y}_t$ instead of single scalars y_t and x_t (see, e.g, Juselius, 2006).

As the number of time series grows, the VECMs may become complex and the preliminary inspection of the time series and their cointegration relationships before fitting the model can turn out laborious (see, e.g., Juselius, 2006; Kennedy, 2003).

4.3 Hidden Markov model

Discrete time hidden Markov models (HMM) are stochastic models with two levels separating the observed and unobserved dynamics of a phenomenon consisting of consecutive events. The observed process is a sequence of discrete observations which depend on the unobserved process. The unobserved process, in turn, is a latent or hidden Markov process with discrete states reflecting the prevailing position of the phenomenon of interest. This chapter follows the theory represented in Rabiner (1989), Zucchini and MacDonald (2009), and Visser and Speekenbrink (2022) unless locally otherwise stated.

Denote the hidden states with s_t, where $t = 1, \ldots, T$. The possible values of the states s_t are denoted with $k = 1, \ldots, S$ for simplicity, even though they in practice are some descriptive situations of reality which might not even have any innate order. The hidden states form a Markov chain, commonly fulfilling the first-order Markov property so that any state depends only on the previous one. Thus, the probability of a current state conditioned on the preceding states is

$$p(s_t|s_1, \ldots, s_{t-1}) = p(s_t|s_{t-1}). \tag{23}$$

Of course, the process may alternatively satisfy a Markov condition of some other degree so that the states depend on as many preceding states as is the degree of the Markov chain. We do not consider those cases here, but additional information can be found, for example, in Zucchini and MacDonald (2009).

Depending on the application, it is possible to stay in the current state or move to another one. The probabilities of different transitions between the states are denoted with $a_{k,k'} = P(s_t = k'|s_{t-1} = k)$ where k and k' are labels for the states $1, \ldots, S$. The transition probabilities are collected in an $S \times S$ transition matrix A. For the possible states of the first time point, we have to assign separate probabilities since there is no transition to the first state but it is the starting point. These initial probabilities are commonly denoted with $\rho_k = P(s_1 = k)$, where $k = 1, \ldots, S$.

To return to the observational level, denote the observations with y_t, where $t = 1, \ldots, T$, and let them depend only on the current state s_t of the hidden process. Due to this, the observations are conditionally independent, which can be formulated to an equation

$$p(y_t|y_1, \ldots, y_{t-1}, s_1, \ldots, s_t) = p(y_t|s_t). \tag{24}$$

This observational distribution $p(y_t|s_t)$ and its parameters θ_{s_t}, associated with the states s_t, are the link between the observed process and the hidden layer. One

may also write $p(y_t|s_t) = p(y_t|\theta_{s_t})$. Some of the parameters θ_{s_t} may depend on the value of the state s_t, while others may be shared between the states. The probabilities of the states s_t to generate the observations y_t with the parameters θ_{s_t} are denoted with $\omega_{k,t} = p(y_t|\theta_{s_t})$, and they are collected in an $S \times T$ emission matrix Ω.

Finally, the HMM may be defined by the set $\{\Omega, A, \rho\}$. Additionally, Equation 23 and Equation 24 describe the dependencies within the model. The general structure of an HMM is illustrated in Figure 4. Usually, the points of interest in hidden Markov modeling are the states and parameters included in the hidden layer and describing some unknown features of the real world. Being unobserved, they have to be estimated based on the observations of the non-hidden level. The phenomena might be complex, so, fortunately, the HMM methodology is generalizable from the case of a single times series to multidimensional data.

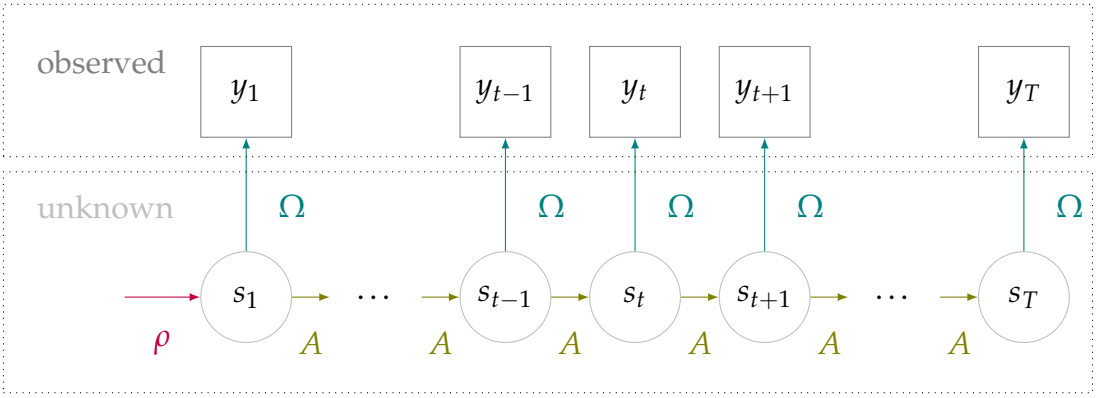

FIGURE 4 An example of a directed acyclic graph of a hidden Markov model.

The full joint distribution of the observations and the hidden states consists of the initial probabilities, the transition probabilities, and the emission probabilities and it can be written as

$$p(y_1, \ldots, y_T, s_1, \ldots, s_T|\Omega, A, \rho) = p(s_1|\rho)p(y_1|s_1, \Omega) \prod_{t=2}^{\top} p(s_t|s_{t-1}, A)p(y_t|s_t, \Omega). \tag{25}$$

The sum of these joint distributions over all possible sequences of states, in turn, is the likelihood of the model, i.e., the marginal distribution of the observations as a function of the model parameters. Formally, this can be written as

$$L(y_1, \ldots, y_T|\Omega, A, \rho) = \sum_{s_1, \ldots, s_T \in \mathcal{S}} p(y_1, \ldots, y_T, s_1, \ldots, s_T|\Omega, A, \rho), \tag{26}$$

where $\mathcal{S}$ denotes the set of all possible hidden processes. A convenient way to calculate the likelihood is to use the forward variables of the forward-backward algorithm (Baum and Petrie, 1966; Rabiner, 1989, and the references therein). Next, we describe the forward and backward variables briefly. As the parameters Ω, A, and ρ are considered known in the following algorithms, according to the common practice, we do not write them explicitly as conditions in the equations below either.

First, define a forward variable $\alpha_t(k)$ representing the joint probability of having the observed sequence $y_1, \ldots, y_t$ until time t and being then in the state k at time point t:

$$\alpha_t(k) = p(y_1, \ldots, y_t, s_t = k). \tag{27}$$

This joint distribution is marginalized over the hidden state path leading to being in state k at time t. By the definition and the dependency between the observations and the states, for the first time point $t = 1$, we have that

$$\alpha_1(k) = p(s_1 = k)p(y_1|s_1 = k). \tag{28}$$

Values at the following time points, in turn, may be calculated recursively as

$$\begin{aligned}\alpha_t(k) &= \sum_{i=1}^{S} p(y_1, \ldots, y_{t-1}, y_t, s_{t-1} = i, s_t = k) \\ &= \sum_{i=1}^{S} p(y_1, \ldots, y_{t-1}|s_{t-1} = i)p(y_t, s_t|s_{t-1} = i) \\ &= \sum_{i=1}^{S} p(s_t = k|s_{t-1} = i)p(y_t|s_t = i)\alpha_{t-1}(i).\end{aligned} \tag{29}$$

The second and third lines follow from the properties of the observation to depend only on the current state and of the current state to depend only on the previous state. Using the forward variables, the complete likelihood is

$$p(y_1, \ldots, y_T) = \sum_{i=1}^{S} \alpha_T(i). \tag{30}$$

Since these probabilities tend to zero fast, especially with large values of T, the estimation may suffer from numerical instability. However, this can be relieved by calculating the scaled forward variables

$$\tilde{\alpha}_t(k) = \frac{\alpha_t(k)}{\sum_{i=1}^{S} \alpha_t(i)} \tag{31}$$

and utilizing them to achieve the likelihood.

The other part, the backward variable $\beta_t(k)$, describing the probability to observe the sequence $y_{t+1}, \ldots, y_T$ after the time point t on the condition of being in state k at time point t, is defined as

$$\beta_t(k) = p(y_{t+1}, \ldots, y_T|s_t = k). \tag{32}$$

This conditional probability is marginalized over all the hidden state paths. Now, instead of starting from the first time point we begin the estimation from the last time point T where we have

$$\beta_T(k) = 1 \text{ for all } k = 1, \ldots, S. \tag{33}$$

The preceding values may be calculated recursively going backwards:

$$\begin{aligned}\beta_t(k) &= \sum_{i=1}^{S} p(y_{t+1}, y_{t+2}, \ldots, y_T, s_t = k, s_{t+1} = i) \\ &= \sum_{i=1}^{S} p(y_{t+1}, s_{t+1} = i | s_t = k) p(y_{t+2}, \ldots, y_T | s_{t+1} = i) \\ &= \sum_{i=1}^{S} p(s_{t+1} = i | s_t = k) p(y_{t+1} | s_{t+1} = i) \beta_{t+1}(i).\end{aligned} \tag{34}$$

As well as the forward variables, for numerical reasons, also the backward variables are often scaled. It is done using scaling factors based on the scaled forward variables $\tilde{\alpha}_t(k)$ in Equation 31. Further details can be found, for example, in Visser and Speekenbrink (2022). Instead of scaling, the numerical instability can be alleviated by making the calculations on a logarithmic scale, which on the other hand is computationally expensive since it requires transforming the probabilities back to their natural scale with exponential function.

Together the forward and backward probabilities can be used to calculate the probabilities to be in each state at each time point given the observations. The statewise conditional probabilities are defined as

$$p(s_t = k) = \frac{\alpha_t(k)\beta_t(k)}{L(y_1, \ldots, y_T)}. \tag{35}$$

This is useful when making inferences based on the model, especially considering the interest lies in the hidden layer.

Calculating the likelihood by marginalizing over the paths of the hidden states is convenient when it comes to missing data. If the observations are missing at random, i.e., the missingness does not depend on the value of the observations, the emission probabilities for the missing values are set to one, $\omega_{k,t} = 1$ for all k. This corresponds to omitting the effect of the emission probability on the computation of the marginal likelihood at the time point of the missing observation since no new information based on the data is gained then. For example, Visser and Speekenbrink (2022) discusses more the treatment of missing data in the case of HMMs.

One of the first questions while applying a HMM is the number of hidden states S to use. Sometimes the decision might be obvious due to the phenomenon or the limitations of the data, but often not—especially because it is the hidden layer we are focusing on here. Several methods have been suggested for estimating the number of states, for example, in Pohle et al. (2017), the expert knowledge, the validation of the model via model checking, the model selection criteria, and the computational aspects are recommended to be considered jointly. Nevertheless, the choice seems quite subjective regardless of the selection method, and usually, the number of states is treated as a fixed, known value during the analysis. This, however, affects the way the results are and should be interpreted. A specific number of states may lead to results that direct us to certain interpretations and explaining specifically them, whereas using a different number of states

can result in a different kind of result, possibly hiding the actual phenomenon or revealing an alternative theory.

After choosing the number of states, one has to define which transitions between the states are possible. Often there is no reason to set any limits, but sometimes additional information can be employed to relieve the estimation process. HMMs can also be used specifically in cases where this kind of restrictive information is to be utilized. Whereas AR models are good in describing smooth transitions in the time series, HMMs are able to allow abrupt changes. One example of this is the change point model, in which the order of transitions is known and only the length of stay in each state is estimated. This, and other HMMs in which the states are ordered so that transitions returning to any preceding state are not allowed, are called left-to-right models, which are also visualized in Figure 5. In Article III, a left-to-right model with two states is used to find a single change point.

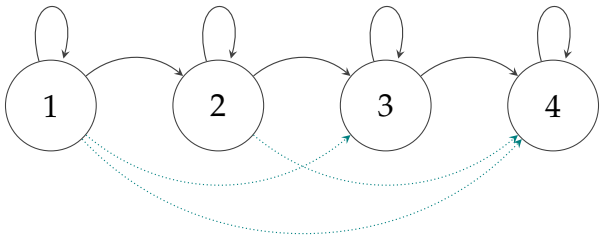

FIGURE 5 The possible transitions in a four-state left-to-right HMM. By ignoring the transitions denoted with dotted blue arrows, the model corresponds to a change point model in which all states are visited.

When it comes to the estimation of the hidden Markov models, there typically occurs multimodality that complicates the analysis. Sometimes the multiple modes might be a result of the label-switching problem (Jasra et al., 2005), and the situation may be relieved, for example, by using specific algorithms or employing additional constraints suitable for the specific case. In complex models, the multiple modes might not be related to merely the difficulties in estimation, but they actually are true local modes. With the maximum likelihood approach the model parameters may be estimated several times using different starting values for the optimization algorithm, and the best solution is declared also as the global mode. In the Bayesian setting, the multimodality of the posterior distribution can hamper the convergence of the estimation algorithms and the interpretation of the results.

4.4 Time-dependent and monotonic effects

When we have an explanatory variable in our model, a coefficient dependent on time can be assigned to that to describe the changes over time. In that way,

we may study the temporal association between the response and the covariate. Naturally, the same can be done for several background variables simultaneously if needed.

The temporal coefficients may be structured in several ways. There does not have to be any specific dependency structure between the time points, but the coefficients can depend on time but be independent of each other. Then again, we may also say that $\beta_t \sim \mathrm{N}(0,\ \sigma^2)$, which results in conditionally independent coefficients given the deviation σ. The coefficients may also construct, for example, a random walk or other autoregressive component as described in section 4.1. In the case of the random walk, for time points $t = 2, \ldots, T$, we have

$$\beta_t \sim \mathrm{N}(\beta_{t-1},\ \sigma^2), \tag{36}$$

where β_t is the temporal coefficient and σ is its standard deviation. The first time point has to be considered separately. In the Bayesian context, a prior may be assigned to that. The autoregressive processes are used, for example, in Article I. Another possible structure is monotonicity, which is used in Articles III and IV.

A monotonic effect is increasing or decreasing over time, but it cannot alter between those. Monotonic effects are discussed, for example, in Bürkner and Charpentier (2020). In Article IV, the effects are allowed to increase or decrease and in Article III they are restricted to increase in order to alleviate the label-switching problem mentioned in section 4.3.

One way to construct a monotonic effect is to use a simplex $\boldsymbol{b}$ length of the number of the time points, T. The first element of the simplex, b_1, is fixed to zero and the rest of the elements sum up to one. The actual coefficients form another vector, $\boldsymbol{\beta} \in \mathbb{R}^T$. The coefficients of the first and last time points, β_1 and β_T, respectively, are defined separately since they are needed for the construction of the ones between them. In Bayesian applications, the first and last coefficients may be defined by assigning priors to them. Now, using the simplex and the predefined coefficients, we can set

$$\beta_t = \beta_1 + (\beta_T - \beta_1) \sum_{i=1}^{t} b_i, \tag{37}$$

where $t = 2, \ldots, T-1$, which leads to the monotonicity.

The monotonicity enables abrupt changes in the level of the effect. This is visible, for example, in Article IV, in which a clear decrease of an effect is visible after a legislative reformation related to the application is made.

5 SPATIO-TEMPORAL MODELS

Spatial models employ structure for a multidimensional variable at one time point, whereas temporal models join the variables over time. If we have data covering both spatial and temporal dimensions, we also have to consider the dependencies within both dimensions. This can be achieved with spatio-temporal models. They are constructed as combinations of spatial and temporal models, so that time and space can be acknowledged simultaneously. Next, we introduce the fusion of the intrinsic conditional regression, discussed in section 3.1, with the temporal random walk model, discussed in section 4.1, we describe how adding spatial structure into the error correction model, familiar from section 4.2, may lead to a simultaneous autoregression introduced in section 3.2, and how the hidden Markov model from section 4.3 can make use of spatial components.

5.1 Random walk ICAR

A multivariate random walk model combined with ICAR structure at the individual time points is a convenient distribution to use as a prior in the Bayesian framework. It specifies a set of local temporal models with a structure such that in locations close to each other, the phenomenon is likely to behave similarly.

For the spatio-temporal parameters $\varphi_{i,t}$, where $i = 1, \ldots, N$ and $t = 1, \ldots, T$, we set

$$\boldsymbol{\varphi}_t \sim \mathrm{N}(\boldsymbol{\varphi}_{t-1},\ Q^{-1}), \tag{38}$$

where $Q = \tau(D - W)$ is the ICAR precision matrix similar to the one in Equation 5 and Equation 6. By introducing the time dependency into the general ICAR model, in Equation 9 we may replace the parameters $\boldsymbol{\varphi}$ with $(\boldsymbol{\varphi}_t - \boldsymbol{\varphi}_{t-1})$, and $\varphi_{i,t}$

with $(\varphi_{i,t} - \varphi_{i,t-1})$ for all i and correspondingly for all j, which leads to

$$p(\boldsymbol{\varphi}) = (2\pi)^{N/2} \left|(\tau(D - W))^{-1}\right|^{1/2} \times \exp\left(-\frac{\tau}{2} \sum_{i=1}^{N} \sum_{j \in J_i} \left(\varphi_{i,t} - \varphi_{i,t-1} - \varphi_{j,t} + \varphi_{j,t-1}\right)^2\right). \tag{39}$$

This may be seen as an interaction of the independent explanatory components of time and space, and it can be represented as a Kronecker product as in Knorr-Held (2000). More details about the interactions can be found from Clayton (1996). This model becomes computationally expensive as the number of time points increases since the number of the spatial random variables depends on the length of the time series, thus it cannot be straightforwardly applied in all situations.

5.2 Extension of the general error correction model

The original error correction model consists of several time series with shared dependencies which are not spatial by definition. Thus, the spatiality has to be considered separately when needed. The following presents one approach to that, also applied in I.

First, we denote the observations with $y_{i,t}$, where $i = 1, \ldots, N$ stands for the location and $t = 1, \ldots, T$ the time point. Additionally, we employ some neighborhood definition, for example, the border sharing, and denote the neighboring location with j, corresponding to the location indexing used with i. Since there may be several neighbors for one location, the general ECM in Equation 22 is adapted to consider them at once using pairwise short-term and long-term coefficients, $\beta_{i,j}$ and $\gamma_{i,j}$, respectively, by

$$\begin{aligned} y_{i,t} - y_{i,t-1} &= \alpha + \sum_{j \in J_i} \beta_{i,j}(y_{j,t} - y_{j,t-1}) + \sum_{j \in J_i} \gamma_{i,j}(y_{i,t-1} - y_{j,t-1}) + \varepsilon_{i,t} \\ \iff \quad y_{i,t} &= \alpha + y_{i,t-1} + \sum_{j \in J_i} \left(\beta_{i,j} y_{j,t} + \gamma_{i,j} y_{i,t-1} - (\beta_{i,j} + \gamma_{i,j}) y_{j,t-1}\right) + \varepsilon_{i,t}, \end{aligned} \tag{40}$$

where $t = 2, \ldots, T$, the set J_i is the collection of the neighbors of the location i, and the errors $\varepsilon_{i,t} \sim \mathrm{N}(0,\ \sigma^2)$.

If we estimated N separate models, that is, one for each location, they might not be comparable nor would their results be necessarily compatible with each other, even if the whole neighborhood was included at once as above. To avoid estimating the individual models and choosing a reference location to be able to compare them and analyze the behavior of the phenomenon in different locations, the models are estimated simultaneously. Formally, for $t > 1$ we have

$$\boldsymbol{y}_t = \mathbf{1}\alpha + B\boldsymbol{y}_t + D\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{41}$$

where $\mathbf{1}$ denotes an N-length vector of ones, the entries of the $N \times N$ matrix B are

$$b_{i,j} = \begin{cases} \beta_{i,j}, \text{ when locations } i \text{ and } j \text{ are neighbors, and } i \neq j, \\ 0, \text{ otherwise,} \end{cases}$$

and the entries of the $N \times N$ matrix D are

$$d_{i,j} = \begin{cases} -(\beta_{i,j} + \gamma_{i,j}), \text{ when locations } i \text{ and } j \text{ are neighbors, and } i \neq j, \\ 1 + \sum_{k \in J_i} \gamma_{i,k}, \text{ when } i = j, \\ 0, \text{ otherwise.} \end{cases}$$

As before, the errors $\varepsilon_{i,t}$ follow a normal distribution.

Now, Equation 41 can alternatively be written as a Gaussian distribution. Using the above notations, say

$$\boldsymbol{y}_t \sim \mathrm{N}((I - B)^{-1}(\mathbf{1}\alpha + D\boldsymbol{y}_{t-1}),\ \sigma^2(I - B)^{-1}((I - B)^{-1})^\top). \tag{42}$$

The formula resembles a SAR model, in which the spatiality is accounted for by the spatial dependency parameter ρ as in section 3.2 and the predefined weight matrix W, that is, by ρW. If this is set equal to our matrix B, the connection between Equation 42 and Equation 12 is evident. The major difference lies in the fact that with this ECM representation, the elements of the matrix B differing from zero are unknown parameters interpreted as the short-term coefficients of the ECM instead of the dependency parameter and preset weights.

5.3 Spatiality in hidden Markov model

As were not the temporal models before, neither are the hidden Markov models limited by multidimensional responses. The simplest way to join the dimensions of time and space within an HMM is to assign some spatial dependency structure for the parameters, observations, or both. The strength of dependency may vary between the states or be shared between them. In III the states are described with maps rather than single values. The maps may be interpreted and summarized in the way appropriate for the application, of course.

A convenient approach is to assign an observational distribution for the responses $y_{i,t}$ so that there are some spatially dependent parameters. Then, for example, a CAR, ICAR, or SAR model can be applied to those variables. In this framework we assume the states and transitions are shared between the locations instead of having unique paths of hidden states for each location. The temporal dependency is present via the transitions on the hidden layer, but it can be further acknowledged by restricting some transitions between the states or, for example, adding an explicit temporal structure, as a random walk, for the transitions. Thus the temporal structure guides the transitions between the states of all locations simultaneously while the spatial structure holds together the other dimension by

smoothing the differences between adjacent locations. For example, in Article III, a unique spatial random effect is added into each state and an ICAR prior assigned for each of those.

6 BAYESIAN MODELING

All the articles in this dissertation utilize Bayesian estimation methods. This enables the application of complex hierarchical models with a large number of unknown parameters and dependencies. Also, prior knowledge about the studied phenomenon can be exploited in the modeling procedure. The whole analysis is driven by the application rather than by the methods and their possible limitations. The uncertainty which is always present in statistical analysis is inherent in the posterior distributions of individual parameters and, for example, in the distributions of the predictions based on the posterior distributions. The results can be seen as intuitive and easily interpretable in many cases, as is with all the included Articles I–IV. In what follows, the ideas of Bayesian modeling in the general case are shortly reviewed.

Bayesian modeling is an alternative approach to the more traditional maximum likelihood estimation of the model parameters. The method can be seen to consist of three phases. First, one has to have a prior belief of the situation. Secondly, there has to be some kind of evidence or data considering the situation. This evidence is used to update the current belief of the prevailing situation. Thirdly, posterior conclusions are drawn based on the priors and the evidence. The approach relies strongly on the Bayesian way to perceive probability and see the uncertainties as distributions.

The first published Bayesian study was carried out by Bayes (1763), whose name the methodology also carries. According to the Bayes' rule for conditional probabilities, the posterior distribution is

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{43}$$

where $p(\theta|y)$ is the so-called posterior distribution of the unknown parameters given the data. The likelihood, or the data conditioned on the unknown parameters, is denoted with $p(y|\theta)$. The prior the applier assigns for the unknown parameters is, in turn, $p(\theta)$. The denominator $p(y)$ corresponds to the evidence normalizing the whole posterior distribution.

The shape of the posterior distribution is not affected by the normalizing

factor $p(y)$ which can be seen as a constant since the observations y are known. Thus the posterior distribution in Equation 43 is proportional to the numerator:

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{44}$$

The posterior is often calculated in this unnormalized form to reduce the computational burden.

In addition to the parameter estimation, one aim of the inference is to make predictions. According to the Bayes' rule, the prior predictions, or the marginal distribution of y, based only on the predefined model without considering the data, are

$$\begin{aligned} p(y) &= \int p(y, \theta)\, d\theta \\ &= \int p(y|\theta)p(\theta)\, d\theta. \end{aligned} \tag{45}$$

The posterior predictive distribution for some new data $\hat{y}$, given the data already observed and assuming the new and old data, $\hat{y}$ and y, are independent on the condition of the parameters θ, is

$$\begin{aligned} p(\hat{y}|y) &= \int p(\hat{y}, \theta|y)\, d\theta \\ &= \int p(\hat{y}|\theta, y)p(\theta|y)\, d\theta \\ &= \int p(\hat{y}|\theta)p(\theta|y)\, d\theta. \end{aligned} \tag{46}$$

The unknown observables may be predicted with these distributions. The predictions as well as the parameter estimates come in the form of distributions, which describe the probabilities of certain values and the uncertainty related to them and their estimation.

The posterior, especially the normalizing constant, has rarely a closed form representation in the real world applications, which poses computational challenges. Instead of the exact closed form answer, the posterior probabilities are estimated with some approximations. The estimation can be done, for example, via Markov Chain Monte Carlo (MCMC) methods that describe the posterior distribution by drawing samples of it using Markov chains converging to the target distribution. These methods require often extensive computational resources. This is why, despite the early origin of the first Bayesian work, the methodology has become more popular only alongside the development of computers and computation methods, especially since the 1980s when the MCMC algorithm of Gibbs sampler was applied in Geman and Geman (1984). Luckily, since the 1990s and 2000s statistical programs such as Bugs (Bayesian inference using Gibbs sampling, see, e.g., Lunn et al. (2012)) and Stan (Stan Development Team, 2024) have made the methodology more approachable and easier to use (Gelman et al., 2013), providing a new powerful statistical tool. In all of the articles included in this dissertation, Stan is used for the estimation.

Bayesian modeling is one statistical tool among others, and it should be used with care, since it is not the best choice for every problem, whereas appropriately applied, it enables things otherwise unachievable. All in all, it widens the

selection of available methods so that the analysis can more easily be made based on the application and phenomenon instead of being limited by the methodology. For a thorough discussion about Bayesian statistics, its theory, and applications, see, for example, Gelman et al. (2013), or for a more introductory perspective, van Oijen (2020).

7 SUMMARIES AND CONTRIBUTIONS

This chapter summarizes the articles this dissertation consists of. All four studies include Bayesian spatio-temporal models applied to real world problems. The first three articles focus on historical phenomena whereas the fourth one has a more contemporary topic.

7.1 Article I: Bayesian spatial error correction model

In Article I, the error correction model represented in section 4.2 is applied to rye price data from 1860s Finland to investigate the market integration during the great Finnish famine. The motivation of the study lies in the theory that there were connections between the towns such that areas suffering from the famine the least would trade the grain with towns suffering the most, simultaneously alleviating the famine. In further thought, these connections might reflect the general communication routes of the people. Additionally, the circumstances at the time resemble the ones in current developing countries, thus the methodology and results could be utilized to gain information and find solutions to alleviate the current food management problems.

In practice, the traditional error correction model is generalized as described in section 5.2 to fit the application in hand. Additionally, the constant α and the error correction terms $\gamma_{i,j}$ in Equation 41 are allowed to depend on time via AR(1) process and neighbor-pair-specific random walks, respectively. The trend is localized with multiplicative weights $\lambda_i \in \mathbb{R}$ having a mean 1 which enables both the identification of the product $\lambda_i \alpha_t$ and the interpretation of the trend as a nationwide pattern. The model is estimated with a Bayesian approach, which is not the most common way with ECMs.

To concretize the results and visualize the connection routes, a further shock simulation based on the estimated model was made. In the simulation, the potential prices were predicted given the model parameters at a certain time point and a preset common initial price level so that in one town there was an artificial

price increase whose effect the simulation embodied. This way the connections to the nearby towns could be visualized. The simulation also illustrated how our temporally adjusted SAR model implicitly allowed the possibility to affect the current prices of the second order neighbors. Even though our model included only the observations of the first order neighbors in the explanatory component, their association with the previous time point enabled the spatially lagged effect.

All in all, the results reflect connections between the towns, differences and similarities between different kinds of towns, offering novel information. As a result of the data and methodology used, the study provides a more detailed view of the Finnish markets of that time than the previous research has done.

7.2 Article II: Bayesian spatial co-dynamics of diseases

In Article II, the application is the spread and co-dynamics of three infectious diseases in Finnish parishes in 1820–1850. The method borrows information from adjacent areas and other diseases to find dynamics within and between the infections in the pre-healthcare society. The diseases include pertussis, measles, and the afterwards eradicated smallpox. As in Article I, also in this study the communication routes to be discovered play an essential role in terms of the phenomenon.

The application of a multidimensional model to these specific data can be considered a success. There are several sources of uncertainty associated with the data from the initial collection to the preservation process, and in essence, the data concern a sparse and small population, not to mention the quarter of data that are missing. Yet, the phenomenon—the behavior of the infections—is undeniably a process requiring an analysis with sufficiently detailed spatial and temporal division. The compromises made in order to have enough information for spatio-temporal modeling led to a Bernoulli distributed response describing if at least one death caused by a specific disease was recorded at each parish during each month. The monthly time resolution is also an accommodation which in this specific case is dense enough to describe the phenomenon.

The response is explained with seasonal effect and locally adjusted nationwide trends similar to the one used in Article I as well as the binary observations related to the three diseases. In this case, all the diseases have separate but correlated random walks as their incidence trends. The interdependency allows for identifying the co-dynamics between the infections. In addition to the local adjustments, the spatiality is included via the explanatory variables indicating whether some of the diseases had caused at least one death in the focal town or in its neighboring towns in the previous month. The covariates consist of neighborhood-specific means instead of individual values of neighbors since there is not enough information in the data to identify such detailed effects in this case. Some of the model parameters are shared between months or towns, and some are only local or temporal. In reality, the phenomenon is much more com-

plicated than the simplifying assumptions made about the temporal and spatial structures as well as about the background variables. Nevertheless, the results seem reasonable, describing the relations between infections in a new way and offering some novel information about the Finnish history of diseases.

7.3 Article III: Spatial hidden Markov model

Article III continues on the epidemiological path of Article II. The data are based on the same Finnish church book archives, but, this time, the focus is on measles. The time span is extended to cover years 1750–1850, which leads to an increasing amount of missing data as can be seen from Figure 3. As the data are sparse and scarce, the observations are aggregated and simplified to indicate if at least one death caused by measles is observed in each town during each month. The study aims to find a moderate number of different states describing the prevailing probability of infection simultaneously in all locations and to discover how those states alter in time. Thus each state can be illustrated as a map and the order of different states with time series and transition probabilities. This allows for describing multidimensional data in a simpler way than could be done, for example, with animations covering the complete data.

Despite the cumbersome premises of the data, a hidden Markov model is applicable for the analysis. The spatial dimension of the dependencies is covered with the intrinsic conditional autoregression that is integrated with the HMM. At each state, we explain the probability of our response with a state-specific constant, local spatially independent constants, local spatially dependent variables, and a monthly seasonal term. For the spatially dependent parameters, we assign the ICAR structure to control their dependency. The temporal aspect is inherent in the state-dependent variables changing along the transitions.

Additionally, a left-to-right model is estimated based on the state trajectories gained from the original HMM. This is done to find a possible change point in which the communication routes would have changed dramatically. The results depend on the number of states and further analysis accompanied with application expertise is required to draw any strong conclusions. However, our approach sheds light on some of the obscured corners of historical knowledge.

7.4 Article IV: Spatio-temporal paternal leave uptakes

In Article IV, a spatio-temporal Bayesian model is applied to study the differences and dependencies between municipalities in the parental leave uptakes of fathers. The data describing the uptake of quota of paternal leaves in Finland on a yearly level during 2009–2017 come from Statistics Finland. The data are reliable and comprehensive.

The probability of a father to use his quota of parental leave is explained with temporal constant, spatially dependent variables, and time-dependent covariates. The constant describing the nationwide trend of the probability is constructed as a random walk. The spatially dependent variables are assigned an ICAR structure and the covariates have time-dependent regression coefficients shared between the municipalities.

An especially interesting background factor to study is the family value policy in the municipality. According to the results, the association between the used parental leave quota of fathers and the municipality policies supporting outsourcing the care responsibilities from the family decreases over time. Particularly, after the legislative change in 2013, the support does not seem to have a distinctive effect. Overall, the results are aligned with the previous research while bringing new, areal information about parental leave uptakes of Finnish fathers.

8 DISCUSSION

This dissertation consists of four articles on Bayesian spatio-temporal modeling of historical and other phenomena. The methods applied are combinations of spatial models, such as SAR and ICAR, and temporal models, such as the error correction model, autoregressive models, and hidden Markov model. With the methodology, the studies are able to offer novel information on the application topics.

Articles I, II, and III study the pre-healthcare Finland in terms of market integration and epidemiology. In this context, the data are associated with errors on many levels. The collection and preservation procedures are not completely reliable, and the population the data describe differs from contemporary society, making many comparisons cumbersome and inadequate. Due to small population sizes and possibly slower pace of life back in the day, the data are also sparse and scarce, which is why extracting the essential information from the data is not straightforward. Despite all this, the statistical methods used overcome the obstacles set by the data.

Article IV finds the dependencies between Finnish municipalities during 2009–2017 in the uptake of paternal leaves. The study covers the values reflecting the political attitudes toward outsourcing the caretaking responsibilities of families and their connection with the paternal leave uptakes. It also introduces the spatial dimension to this specific sociological application. Both of these are novelties in this context.

Potential issues for future research include, for example, other neighborhood definitions, alternative ways to work with missing information, and finding new models or estimation methods. As mentioned in Article I, the usage of other than first order neighbors would be interesting. Whereas with the historical data, we have used the shared border as a definition for the neighborhood, other definitions might be more relevant if proper professional knowledge could be harnessed for it. Alternatively, an approach similar to the one used in Balocchi and Jensen (2019) could be utilized, since along with the rest of the parameters, they estimate also the possibility of neighboring locations to share a barrier instead of a potential communication route. This seems a method worth trying

until better information, such as waterway or road network, is available. In Article II, the analysis is based on complete cases, even though considering the whole neighborhood instead of individual neighbors allowed using some more observations. Finding another working solution for the missing observations remains. New models or estimation methods could alleviate the issues in Article III as the usage of random walks instead of hidden Markov models was infeasible with the current number of parameters.

The conditions in Finland during the 18th and 19th centuries correspond in many respects to those of developing countries of the modern world. It motivates us to study these historical phenomena—markets reflecting the food management and deaths reflecting the spread of diseases—in sparsely populated agrarian culture without the contemporary advantages of present civilization. The more recent analysis of parental leaves is in essence related to the equality between people, the balance that has been sought for a long time at an ever-increasing pace. The Finnish affluent society can be seen as one of the forerunners of equal parenthood and the research of the effects of reforms of laws and practices can provide critical information not only for the Finnish but also for many other people. Regardless of the differing contexts, all these studies aim to find the connections between people—markets, infections, and adopted influences driving the behavior and decisions of people. The statistical methodology needed for these analyses always depends on the situation, and this dissertation presents some ways to describe the phenomena in which space and time are intertwined.

REFERENCES

Alogoskoufis, G. and Smith, R. (1991). On error correction models: Specification, interpretation, estimation. *Journal of Economic Surveys*, 5(1):97–128. DOI: https://doi.org/10.1111/j.1467-6419.1991.tb00128.x.

Balocchi, C. and Jensen, S. T. (2019). Spatial modeling of trends in crime over time in Philadelphia. *The Annals of Applied Statistics*, 13(4):2235–2259. DOI: https://doi.org/10.1214/19-AOAS1280.

Banerjee, S., Carlin, B., and Gelfand, A. (2015). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2 edition. DOI: https://doi.org/10.1201/b17115.

Baum, L. E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563. DOI: https://doi.org/10.1214/aoms/1177699147.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418. DOI: https://doi.org/10.1098/rstl.1763.0053.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225. DOI: https://doi.org/10.1111/j.2517-6161.1974.tb00999.x.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–20. DOI: https://doi.org/10.1007/BF00116466.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 5 edition.

Briga, M., Ketola, T., and Lummaa, V. (2022). The epidemic dynamics of three childhood infections and the impact of first vaccination in 18th and 19th century Finland. *medRxiv*. DOI: https://doi.org/10.1101/2022.10.30.22281707. [Preprint]. Posted October 31, 2022 [accessed August 25, 2024].

Briga, M., Ukonaho, S., Pettay, J. E., Taylor, R. J., Ketola, T., and Lummaa, V. (2021). The seasonality of three childhood infections in a pre-industrial society without schools. *medRxiv*. DOI: https://doi.org/10.1101/2021.10.08.21264734. [Preprint]. Posted October 18, 2021 [accessed August 25, 2024].

Brook, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, 51(3/4):481–483. DOI: https://doi.org/10.2307/2334154.

Bürkner, P.-C. and Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, 73(3):420–451. DOI: https://doi.org/10.1111/bmsp.12195.

Clayton, D. G. (1996). Generalized linear mixed models. In Gilks, W. R., Richardson, S., and Spiegelhalter, D., editors, *Markov Chain Monte Carlo in Practice*, volume 1, pages 275–302. Chapman and Hall/CRC. DOI: https://doi.org/10.1201/b14835-21.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley. DOI: https://doi.org/10.1002/9781119115151.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. Chapman & Hall/CRC, 3 edition. DOI: https://doi.org/10.1201/b16018.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511790942.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741. DOI: https://doi.org/10.1109/TPAMI.1984.4767596.

Giuliani, D., Dickson, M. M., Espa, G., and Santi, F. (2020). Modelling and predicting the spatio-temporal spread of COVID-19 in Italy. *BMC Infectious Diseases*, 20:1–10. DOI: https://doi.org/10.1186/s12879-020-05415-7.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press. DOI: https://doi.org/10.1515/9780691218632.

Held, L. and Rue, H. (2010). Conditional and intrinsic autoregressions. *Handbook of Spatial Statistics*, pages 201–216. DOI: https://doi.org/10.1201/9781420072884.

Hyndman, R. J. and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. OTexts: Melbourne, Australia, 3 edition. https://otexts.com/fpp3/. Accessed on 24.7.2024.

Jasra, A., Holmes, C. C., and Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, 20(1):50 – 67. DOI: https://doi.org/10.1214/088342305000000016.

Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., and Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, 582(7812):389–394. DOI: https://doi.org/10.1038/s41586-020-2284-y.

Juselius, K. (2006). *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press, USA. DOI: https://doi.org/10.1093/oso/9780199285662.001.0001.

Kennedy, P. (2003). *A Guide to Econometrics*. John Wiley & Sons, 5 edition.

Ketola, T., Briga, M., Honkola, T., and Lummaa, V. (2021). Town population size and structuring into villages and households drive infectious disease risks in pre-healthcare Finland. *Proceedings of the Royal Society B: Biological Sciences*, 288(1949):20210356. DOI: https://doi.org/10.1098/rspb.2021.0356.

Klein, J. L. (1997). *Statistical Visions in Time: A History of Time Series Analysis, 1662-1938*. Cambridge University Press.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19(17-18):2555–2567. DOI: https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-

Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Chapman Hall, London, 1 edition. DOI: https://doi.org/10.1201/b13613.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media. DOI: https://doi.org/10.1007/978-3-540-27752-1.

Metcalfe, A. V. and Cowpertwait, P. S. (2009). *Introductory Time Series with R*. Springer. DOI: https://doi.org/10.1007/978-0-387-88698-5.

Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S. J., Gelman, A., and DiMaggio, C. (2019). Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in Stan. *Spatial and Spatio-Temporal Epidemiology*, 31:100301. DOI: https://doi.org/10.1016/j.sste.2019.100301.

Ó Gráda, C. (2001). Markets and famines: Evidence from nineteenth-century Finland. *Economic Development and Cultural Change*, 49(3):575–590. DOI: https://doi.org/10.1086/452516.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349):120–126. DOI: https://doi.org/10.2307/2285387.

Pitkänen, K. (1977). The reliability of the registration of births and deaths in Finland in the eighteenth and nineteenth centuries: Some examples. *Scandinavian Economic History Review*, 25(2):138–159. DOI: https://doi.org/10.1080/03585522.1977.10407878.

Pohle, J., Langrock, R., Van Beest, F. M., and Schmidt, N. M. (2017). Selecting the number of states in hidden Markov models: Pragmatic solutions illustrated

using animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, 22:270–293. DOI: https://doi.org/10.1007/s13253-017-0283-8.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286. DOI: https://doi.org/10.1109/5.18626.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC. DOI: https://doi.org/10.1201/9780203492024.

Stan Development Team (2024). Stan modeling language users guide and reference manual. Version 2.35. https://mc-stan.org/.

Suomen Sukututkimusseura (2024). HisKi. https://www.genealogia.fi/hiski/. Posted May 31, 2024 [accessed September 17, 2024].

van Oijen, M. (2020). *Bayesian Compendium*. Springer. DOI: https://doi.org/10.1007/978-3-030-55897-0.

Vicente, G., Goicoa, T., and Ugarte, M. D. (2020). Bayesian inference in multivariate spatio-temporal areal models using INLA: Analysis of gender-based violence in small areas. *Stochastic Environmental Research and Risk Assessment*, 34:1421–1440. DOI: https://doi.org/10.1007/s00477-020-01808-x.

Visser, I. and Speekenbrink, M. (2022). *Mixture and Hidden Markov Models With R*. Use R! Springer. DOI: https://doi.org/10.1007/978-3-031-01440-6.

Vuorinen, H. S. (1999). Suomalainen tautinimistö ennen bakteriologista vallankumousta. *Hippokrates: Suomen lääketieteen historian seuran vuosikirja*, 16:33–61.

Wall, M. M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of statistical planning and inference*, 121(2):311–324. DOI: https://doi.org/10.1016/S0378-3758(03)00111-3.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, pages 434–449. DOI: https://doi.org/10.2307/2332724.

Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A*, 226(636-646):267–298. DOI: https://doi.org/10.1098/rsta.1927.0007.

Zucchini, W. and MacDonald, I. (2009). *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman and Hall/CRC. DOI: https://doi.org/10.1201/9781420010893.

ORIGINAL PAPERS

I

A BAYESIAN SPATIO-TEMPORAL ANALYSIS OF MARKETS DURING THE FINNISH 1860S FAMINE

by

Pasanen, T.-M., Voutilainen, M., Helske, J., and Högmander, H. 2022

Journal of the Royal Statistical Society Series C: Applied Statistics, Volume 71, Issue 5, Pages 1282–1302

DOI: https://doi.org/10.1111/rssc.12577

Received: 20 October 2021 | Accepted: 20 May 2022

DOI: 10.1111/rssc.12577

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY
Series C Applied Statistics

A Bayesian spatio-temporal analysis of markets during the Finnish 1860s famine

Tiia-Maria Pasanen[1] | **Miikka Voutilainen**[2] | **Jouni Helske**[1] | **Harri Högmander**[1]

[1]Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland

[2]Department of History and Ethnology, University of Jyväskylä, Jyväskylä, Finland

Correspondence
Tiia-Maria Pasanen, Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland.
Email: tiia-maria.h.pasanen@jyu.fi

Abstract

We develop a Bayesian spatio-temporal model to study pre-industrial grain market integration during the Finnish famine of the 1860s. Our model takes into account several problematic features often present when analysing multiple spatially interdependent time series. For example, compared with the error correction methodology commonly applied in econometrics, our approach allows simultaneous modelling of multiple interdependent time series avoiding cumbersome statistical testing needed to predetermine the market leader as a point of reference. Furthermore, introducing a flexible spatio-temporal structure enables analysing detailed regional and temporal dynamics of the market mechanisms. Applying the proposed method, we detected spatially asymmetric 'price ripples' that spread out from the shock origin. We corroborated the existing literature on the speedier adjustment to emerging price differentials during the famine, but we observed this principally in urban markets. This hastened return to long-run equilibrium means faster and longer travel of price shocks, implying prolonged out-of-equilibrium dynamics, proliferated influence of market shocks, and, importantly, a wider spread of famine conditions.

KEYWORDS

Bayesian statistics, error correction model, Finnish famine, market integration, spatio-temporal model

1 | INTRODUCTION

Analysing the co-movement of prices in multiple spatially connected regions is essential in gaining understanding of the behaviour, structure and efficacy of the markets (e.g. Fackler & Tastan, 2008; Shin, 2010). However, statistical modelling of such spatio-temporal phenomena can be challenging for various reasons.

Error correction models (ECMs) (Alogoskoufis & Smith, 1991; Engle & Granger, 1987) are a popular tool for studying the short-term co-movement and the long-run out-of-equilibrium dynamics in economics and other domains (e.g. Li et al., 2006, 2013; Møller & Sharp, 2014; Nkang et al., 2006), but traditional ECM assumes a single exogenous variable affecting another dependent variable, making it unsuitable for the context of joint modelling of spatio-temporal markets. While vector error correction models (VECMs) (see, e.g. Juselius, 2006) allow joint modelling of multiple interdependent time series, their use entails burdensome, error-prone and potentially biasing testing for appropriate series in each co-integrated relationship, with additional difficulties due to potentially time-varying relationships (Giles & Giles, 1993; Giles & Godwin, 2012; Gonzalo & Lee, 1998; Hjalmarsson & Österholm, 2010).

In this paper, we study the integration of Finnish rye markets in the 1860s, during which Finland suffered the last peacetime famine in Western Europe. Same markets have been studied earlier in Ó Gráda (2001) using aggregated provincial data and a simple ECM approach where one province was fixed as a market leader and where the famine peak in 1867–1868 was accounted via a dummy variable. However, the use of aggregated data loses information about the smaller scale price variation and the concept of a market leader is somewhat artificial in poorly integrated markets which may contain multiple spatial equilibria (e.g. Chilosi et al., 2013; Studer, 2008; Voutilainen et al., 2020), especially during famines (e.g. Shin, 2010), leaving market behaviour under such environments greatly understudied. Instead, we construct a Bayesian spatio-temporal model which jointly models all regional price series without a need for a predetermined point of reference. Furthermore, as the famine was a protracted socio-economic process punctuated by failing harvests throughout the 1860s, we use time-varying coefficients which require no pre-determined period for the famine and allow a smoothly shifting relationship between the spatial markets. While our model bears some similarity to SAR models (Ord, 1975), it provides a flexible and efficient estimation of the spatial correlation structures without a predefined spatial weight matrix. In addition, the model inherently separates the short-term and long-term regional relationships.

By allowing complicated feedback between the regions, we are able to study the overall stability of the market system. We observe that the out-of-equilibrium dynamics are contingent on the overall spatial organization of the markets. Speedier error correction facilitates transmission of price shocks, and regional feedback keeps the overall price system longer periods out of equilibrium. In essence, our findings cast doubt on whether speedier error correction between any pair of markets can solely be used to infer the behaviour of the market system as a whole due to the complex short-term and long-term interdependence within the markets on local and system-wide level. We also document the existence of 'price ripples' (Devereux, 1988; Hunter & Ogasawara, 2019; Seaman & Holt, 1980) from a shock epicentre but observe a spatially asymmetric adjustment to price shocks. The majority of the increased spatial price transmission happened in urban–rural trade with some evidence that urban prices were increasingly influenced by the price level in the surrounding rural regions during the famine.

The paper is structured as follows. In Section 2, we provide background for the Finnish famine of the 1860s and discuss the data used in this study. In Section 3, we extend the traditional ECM to a spatio-temporal setting. In Section 4, we implement the model to study the spatial price behaviour during the Finnish 1860s famine, leaving it for Section 5 to discuss the results and model extensions, and to conclude.

2 | BACKGROUND AND DATA

The role of markets in safeguarding or depleting food security has remained in analytical focus, as recent famines (Maxwell et al., 2020) are characterized by a failure of markets to deliver access to food at affordable prices (Andrée et al., 2020; Devereux, 2009; Howe & Devereux, 2004). While there is little disagreement on the theoretical benefits of a well-integrated market system (e.g. van Campenhout, 2007; Matz et al., 2015; Persson, 1999; Svanidze et al., 2019, see however, e.g. Deng, 2007), a distinction remains between the long-run gains and adverse behaviour of markets over shorter periods.

By common account, markets can influence famines in three ways: markets may (1) alleviate local food security problems through profit-seeking arbitrage, (2) balkanize and stop transmitting price information needed to signal outside traders of localized food shortages and (3) make things worse by exporting food from a region of scarcity (see, e.g. Devereux, 1988, 2009; Ó Gráda, 2005, 2015). Most famines experienced during the past 100 years have been marked by disintegration and balkanization of market relationships (e.g. Ó Gráda, 2015; Shin, 2010), excessive price volatility (Araujo Bonjean & Simonet, 2016; Devereux & Tiba, 2007; Garenne, 2007; Quddus & Becker, 2000; Ravallion, 1987), lack of accurate market information and politically motivated hampering of market access (e.g. Macrae & Zwi, 1992; Marcus, 2003; Ó Gráda, 2015; Olsson, 1993; de Waal, 1993).

This modern experience opposes the available evidence from the pre-industrial markets, which tended to function better during famines than in normal times (e.g. Ó Gráda, 2001, 2005, 2015; Ó Gráda & Chevet, 2002). Several historical studies have concurrently suggested that the further back we go in time, the higher prominence low agricultural productivity and economic backwardness rise in explaining the occurrence and timing of famines (e.g. Alfani & Ó Gráda, 2018; Campbell, 2010; Hoyle, 2010; Mishra et al., 2019). With some notable exceptions (e.g. Slavin, 2014), these findings have poised scepticism that historical markets overwhelmed by supply shocks had the means or volumes needed to alleviate nationwide hardships. The Finnish 1860s famine has provided one of the original pieces of evidence that unlike their modern counterparts, the pre-industrial markets functioned well, even better during famines (Ó Gráda, 2001). To contribute to the literature, we revisit this case of famine.

2.1 | The Finnish 1860s famine and data

Held as the last peace-time population disaster of Western Europe, the Finnish 1860s famine occurred in an overwhelmingly agricultural economy in the hinterland of an industrializing continent and claimed close to 200,000 lives (in excess mortality and in absence of births) out of a population of 1.8 million. With the proximate cause in a substantial drop in harvests (on aggregate losses over 50% depending on location and grain variety) which could not be supplemented with imports, the famine mortality was greatly amplified by the pre-famine increase in

vulnerable population stemming from high population growth, concentrated land ownership and from lack of urban and industrial employment opportunities (Voutilainen, 2016).

In the mid-19th century, although Finland was sparsely settled and often self-reliant regarding grain output, food markets were sufficiently developed. This was true particularly in the historically well-off areas of Southwest Finland as well as in the southeastern parts of the country that laid close to the grain markets of St. Petersburg. Grain sales were an important source of income especially in the South-West Finland, and a substantial amount of grain consumed during harvest failures was purchased beyond local markets all over the country (Soininen, 1974). Due to historical trade connections between rural areas and towns and due to geographical constraints on movement (especially water ways), the regional nature of the inland trade remained in place well into the mid-19th century. Towns and town merchants were responsible for the largest trade volumes, and towns due to their livelihood structure had the most developed markets for food. Rural trade was liberalized in 1859, but it was mostly conducted to meet households' day-to-day needs and not used for large-scale rural food retailing (Voutilainen et al., 2020).

Our analysis of the food markets is based on the price of rye—the most widely traded and most voluminously produced grain of the time (Soininen, 1974). The price data were obtained from local reports sent to the Department of Finance and stored in the Finnish National Archives in Helsinki.[1] Starting in 1857, the officials of 48 administrational districts and in 32 town administrations were required to report monthly price denominations of the local prices of various commodities. The prices were market prices and best ones available from the era (Pitkänen, 1993). The data have been previously analysed by Ó Gráda (2001), who—in contrast to our approach—divided the data in eight provincial aggregates downgrading regional variance and leaving, among other things, the within-province heterogeneity and urban–rural division unhandled. The higher spatial resolution of this study is more in resonance with livelihood and market regions (Maxwell et al., 2020; Voutilainen et al., 2020), thereby providing more accurate measurement of the actual prices faced by the contemporary people.

Figure 1 plots the aggregate price movement between January 1861 and December 1869, which is the time span of the data we analyse. In the beginning of our time span, in 1861,

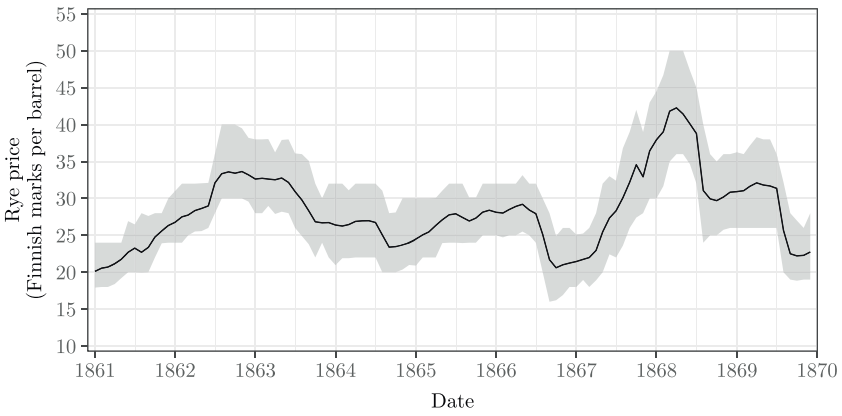

FIGURE 1 Rye price in Finland during 1860s as Finnish marks (FIM) per barrel. Average as black curve and 95% quantile interval as shaded area.

[1]We are grateful to Kari Pitkänen for allowing us to use the data. Additional archival work to fill in the gaps in the data was conducted for the purpose of this study.

the coefficient of variation of regional rye prices was about 0.1, roughly at par with the contemporary Western European economics, and signalling reasonably well-integrated markets in terms of sigma convergence (e.g. Federico et al., 2021). The 1860s famine, which peaked in terms of mortality in the winter and spring of 1867–1868, was foreshadowed by multiple harvest failures earlier in the decade. Harvest failures of 1862, 1865 and especially the famine escalating one in 1867, were tailed by a substantial increase in grain prices over the following winter.

3 | SPATIO-TEMPORAL ERROR CORRECTION MODEL

If markets are perfectly integrated, the prices in any two places should follow the so-called law of one price. Under these circumstances, a price differential will signal opportunities for arbitrage up to the point when prices are either identical or differ to reflect the transportation costs (Ó Gráda, 2015; Studer, 2008). One of the implications of this is that the prices in two markets are characterized by a long-run relationship, deviations from which vanish by ensuing arbitrage (Ó Gráda, 2001; Studer, 2008). This is customarily studied using an error correction approach (e.g. Alogoskoufis & Smith, 1991; Kennedy, 2008). Considering two (log-)price time series $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_T)$, where T is the number of time points, an ECM is traditionally represented as

$$y_t - y_{t-1} = \alpha + \beta(x_t - x_{t-1}) + \gamma(y_{t-1} - x_{t-1}) + \epsilon_t, \tag{1}$$

where α stands for a common trend, β denotes the short-run effect of changes in the reference series x on y, and γ denotes the effect of the difference between y and x and $\epsilon_t \sim \mathrm{N}(0, \sigma_y^2)$.

In normally functioning markets, we expect that $0 < \beta < 1$ and $-1 < \gamma < 0$. The short-term coefficient β captures the price co-movement between the two markets. The error correction term γ is the long-term adjustment parameter: the share of disequilibrium ('error') eradicated in each period. The higher the efficiency of the markets—that is, the closer γ is to -1—the larger the proportion of the error vanishing between time points t and $t + 1$, and therefore the quicker the emerged disequilibrium is arbitraged away.

To provide a flexible method for multiple spatially related series, we generalize the common error correction Equation (1) in the following way. Consider a geographic region that can be partitioned into sub-regions denoted by $i = 1, 2, \ldots, N$. In our case, we regard regions sharing borders as neighbours, even though some other condition for neighbourhood could be used as well. We denote the neighbourhood of a site i with J_i, and an individual neighbour with $j \in J_i$. There is a location specific price $\mu_{i,t}$ of the product for each region and each time point. The prices for all the regions at one time point are denoted by $\boldsymbol{\mu}_t = (\mu_{1,t}, \mu_{2,t}, \ldots, \mu_{N,t})^T$.

Now, if we denote the y_t and x_t in the original ECM by $\mu_{i,t}$ and $\mu_{j,t}$, respectively, the model in the case of only one neighbour looks like

$$\mu_{i,t} = \mu_{i,t-1} + \alpha + \beta_{i,j}(\mu_{j,t} - \mu_{j,t-1}) + \gamma_{i,j}(\mu_{i,t-1} - \mu_{j,t-1}) + \epsilon_{i,t}.$$

The errors $\epsilon_{i,t}$ are assumed to be Gaussian. This model, in turn, can easily be generalized for an arbitrary number of neighbours by adding new terms to all the neighbouring sites in question; thus, we obtain

$$\mu_{i,t} = \mu_{i,t-1} + \alpha + \sum_{j \in J_i} \beta_{i,j} \left(\mu_{j,t} - \mu_{j,t-1}\right) + \sum_{j \in J_i} \gamma_{i,j} \left(\mu_{i,t-1} - \mu_{j,t-1}\right) + \epsilon_{i,t}.$$

We now consider all the regions and their possible connections simultaneously by combining all the site specific models and including the spatial dependency structure in its entirety into a single model frame. The equation above can be written for all regions simultaneously with a matrix notation as

$$\boldsymbol{\mu}_t = \mathbf{1}\alpha + B\boldsymbol{\mu}_t + D\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t.$$

The term $\mathbf{1}$ denotes an all-ones vector of length N, and the entries of the $N \times N$ matrix B are

$$b_{i,j} = \begin{cases} \beta_{i,j}, & \text{if } i \text{ and } j \text{ are neighbours, that is, } j \in J_i \\ 0, & \text{otherwise.} \end{cases}$$

The $N \times N$ matrix D, in turn, has the entries

$$d_{i,j} = \begin{cases} -(\beta_{i,j} + \gamma_{i,j}), & \text{if } i \text{ and } j \text{ are neighbours, that is, } j \in J_i \\ 1 + \sum\limits_{j' \in J_i} \gamma_{i,j'}, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases}$$

With a short notation

$$\boldsymbol{\eta}_t = \mathbf{1}\alpha + D\boldsymbol{\mu}_{t-1}$$

the previous equation can be expressed as

$$\boldsymbol{\mu}_t \sim \mathrm{N}((I - B)^{-1}\boldsymbol{\eta}_t, \sigma_\mu^2 (I - B)^{-1}((I - B)^{-1})^T).$$

Instead of explicitly specifying the cointegration relationships, reference series and other complex interdependency structures between the regions, we started with a generalization of a simple ECM with a fixed neighbourhood structure and, perhaps surprisingly, arrived at a model that bears similarity with simultaneous autoregressive (SAR) models (Anselin, 1988; Ord, 1975), which are widely used in spatial statistics. Compared to a typical SAR model in which the neighbourhood is defined via $B = \rho W$ with a spatial dependency parameter ρ and a fixed spatial weight matrix W, here the non-zero elements of B are unknown parameters with an interpretation corresponding to the original ECM formulation. The exogenous variables commonly used as predictors in a SAR model are replaced here with a lagged dependent variable μ_{t-1}, which follows directly from the above expansion of the original ECM. The advantage of our formulation is that the underlying cointegration relationships are now implicitly included by the spatial spillover of the SAR structure. This enables a situation where a region j at time t may have an effect on a region i at the next point of time $t + 1$, even though the regions i and j are not direct neighbours. An extension of standard SAR model with heterogeneous spatial lag coefficients was developed in (Aquaro et al., 2021), but this model does not decompose the spatial and temporal relationships into short- and long-term effects as our ECM-inspired formulation, and thus lacks some of the interpretative power of our model.

Whether the error correction parameters change during famines has been studied using a predetermined famine period demarked by a dummy variable (e.g. Ó Gráda, 2001, 2005). This, however, risks disregarding the regionally desynchronized evolution of famine conditions and

enforcing a 'cliff-edge' of famine periodization (e.g. Maxwell et al., 2020). Ideally, the model should provide estimates for time-dependency independent of pre-assumptions. Therefore, we allow the parameters α and γ to vary in time. Previous literature has introduced time dependency to the short-run coefficients β (e.g. Li et al., 2006; Ramajo, 2001); however, regarding market integration, the interest lies in the market adjustment parameter γ. Furthermore, we introduce time dependency to the parameter α to account for common nationwide shocks that may surface as an increased price co-movement during the famine (e.g. Ó Gráda, 2001), alongside a common (irregular) seasonal variation due to harvests. Since we do not wish to assume all the sites are identical with respect to α, we introduce also a site specific dependency of the nationwide trend, leading to

$$\boldsymbol{\mu}_t = \lambda\alpha_{t-1} + B\boldsymbol{\mu}_t + D_{t-1}\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t. \tag{2}$$

The term $\boldsymbol{\lambda}$ denotes a vector of the site specific coefficients λ_i that depict how strongly the site i depends on the common α_{t-1}. B is as before. The matrix D_{t-1} depends on time through γ_{t-1}, meaning that instead of having only one matrix compressing the information about the whole period, there is an individual coefficient matrix for each time point. We model the time-varying parameters $\gamma_{i,j,t}$ as random walks and coefficients α_{t-1} as a stationary first-order autoregressive (AR(1)) process.

In practice, our data have $T = 108$ time points and $N = 80$ regions with 298 neighbour pairs. It is plausible that our historical price data contain inaccuracies, for example, due to the temporal coarseness of the measurements or errors in the reporting process itself. The latter is underlined since our data are not complete, but 5% of the observations are missing, probably attributed to the absence of actual price reports or the loss of these reports in the archiving process. Therefore, it is natural to think that the actual price development is a latent process from which we have only noisy observations. The observed log-prices from N sites at the time point $t = 1, 2, \ldots, T$ form a vector $\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{N,t})^T$ for each time point. They are considered to be realizations from a Gaussian distribution

$$\boldsymbol{y}_t \sim \mathrm{N}(\boldsymbol{\mu}_t, \sigma_y^2 I), \tag{3}$$

with the expected value $\boldsymbol{\mu}_t$ as in Equation (2). The variance σ_y^2 is assumed to be constant for all sites and times, and I denotes an identity matrix. The variables $\boldsymbol{\mu}_t$ are the true unobserved prices on the latent level and $\boldsymbol{y}_t$ are our observed data.

4 | RESULTS

We estimate the model based on Equations (2) and (3) using a Bayesian approach.[2] This allows the estimation of complex hierarchical models with prior information and missing data, especially as we wish to take into account the uncertainty due to parameter estimation while interpreting the model parameters (or their arbitrary functions, as in Section 4.2). Bayesian approach also naturally accommodates the possibility that the time-varying components of our model (γ,α) are (nearly) constant in time by averaging over the uncertainty of the respective standard deviation parameters (see below). Following theoretical assumptions underlying price transmission ECMs, we restrict the coefficients $\beta_{i,j}$ to be positive and $-1 \leq \gamma_{i,j,t} \leq 0$ for all i,j and t, and let the

[2] The material to reproduce the analysis is available on https://github.com/tihepasa/bayesecm.

unconstrained $\tilde{\gamma}_{i,j,t} = \text{logit}(-\gamma_{i,j,t})$ follow linear Gaussian random walks (with respect to t). As a prior for region-specific coefficients λ_i, we define the marginal distribution of each λ_i as $\text{N}(1, \sigma_\lambda^2)$, with an additional constraint that the mean of coefficients is exactly 1. This ensures that the product $\lambda\alpha_{t-1}$ is identifiable (Bai & Wang, 2015) while keeping $\boldsymbol{\alpha}$ interpretable.

The prior distributions of our full model are

$$
\begin{aligned}
\alpha_1 &\sim \text{N}\left(\frac{c_\alpha}{1-\phi}, \left(\frac{\sigma_\alpha}{\sqrt{1-\phi^2}}\right)^2\right),\\
\alpha_t &\sim \text{N}(c_\alpha + \phi\alpha_{t-1}, \sigma_\alpha^2), \text{ for } t > 1,\\
c_\alpha &\sim \text{N}(0, 0.1^2),\\
\phi &\sim \text{Beta}(2, 2),\\
\sigma_\alpha &\sim \text{Gamma}(2, 100),\\
\lambda_i &\sim \text{N}(1, \sigma_\lambda^2), \text{ for all } i, \text{ given } \frac{1}{N}\sum_{i=1}^{N}\lambda_i = 1,\\
\sigma_\lambda &\sim \text{N}(0, 0.5^2)[0,],\\
\beta_{i,j} &\sim \text{Gamma}(0.5, 2) \text{ for all } i, j,\\
\tilde{\gamma}_{i,j,1} &\sim \text{N}(-2, 2^2) \text{ for all } i, j,\\
\tilde{\gamma}_{i,j,t} &\sim \text{N}(\tilde{\gamma}_{i,j,t-1}, \sigma_\gamma^2) \text{ for all } i, j, \text{and } t > 1,\\
\sigma_\gamma &\sim \text{Gamma}(2, 10),\\
\mu_{i,1} &\sim \text{N}(3, 0.5^2) \text{ for all } i,\\
\sigma_\mu &\sim \text{Gamma}(2, 20), \text{ and}\\
\sigma_y &\sim \text{Gamma}(2, 20),
\end{aligned}
$$

where $\text{N}(\cdot,\cdot)[0,]$ denotes truncated (at zero) normal distribution.

The priors were chosen based on the approximate prior scale of the variables and then adjusted based on the initial Markov chain Monte Carlo (MCMC) runs for enhanced computational efficiency of the final MCMC sampling. However, the chosen priors did not have strong influence on the posterior estimates compared to more diffuse choices (see supplementary material on GitHub), so they can be regarded as only weakly informative.

The model was estimated with *rstan* (Stan Development Team, 2020), which is an R interface (R Core Team, 2021) for the probabilistic programming language Stan for statistical inference (Carpenter et al., 2017). The samples were drawn using the NUTS sampler (Hoffman & Gelman, 2014) with four chains, each consisting of 8000 iterations, with the 3000 first discarded as a warm-up. The total computation time with parallel chains was about 13 h. The effective sample sizes were approximately between 700 and 45,000, with the most inefficient estimation on the strongly correlating parameters σ_y and σ_μ.

During the estimation of the model, all regions were treated equally without defining whether they were rural districts or urban towns, as we did not have clear a priori knowledge about the potential differences or similarities in say β or γ coefficients between or within these groups. However, there was noticeable posterior evidence that the rural and urban regions behave somewhat differently, and hence we present the results taking this grouping into account.

Instead of describing all region-specific results, in some instances we focus on three example regions, namely the rural districts of Ilmajoki and Masku, and the town of Viipuri. These choices

were based on the estimated model; these regions were found to be important drivers of the prices in their neighbours. Ilmajoki is located on the western coast, whereas Masku lies within South-West Finland, where the vast share of marketed grain surplus was produced (Soininen, 1974). In addition, the choice of Viipuri was motivated by initial observations by Ó Gráda (2001), who stressed the importance of the wider southeastern Viipuri province as a market leader, possibly due to its proximity to the Russian grain markets.

Figure 2 shows examples of estimates for the unobserved log-prices $\mu_{i,t}$ and corresponding 95% posterior intervals for Ilmajoki, Masku and Viipuri. In each instance, the estimates smoothly follow the observed prices. Reasonable estimates were also obtained for missing observations. The estimate of the parameter σ_μ, related to the deviation of the unobserved log-prices, is 0.038 with a 95% posterior interval [0.035, 0.041]. The standard deviation of the observed log-prices, σ_y, is 0.036 with a 95% posterior interval [0.033, 0.038].

Perhaps a natural alternative to our analysis would be to model latent log-prices via a SAR model. To this end, we defined $\mu_t - \mu_{t-1} = \rho W(\mu_t - \mu_{t-1}) + \epsilon$, leading to $\mu_t \sim N(\mu_{t-1}, \sigma_\mu^2(I - \rho W)^{-1}((I - \rho W)^{-1})^T)$, where W is the fixed adjacency matrix after spectral normalization (Kelejian & Prucha, 2010), and $-1 < \rho < 1$ is the unknown spatial dependency parameter. Thus, in addition to latent μ, this model has only three unknown parameters, σ_y, σ_μ, and ρ. These were estimated as 0.034 ([0.032, 0.035]), 0.055 ([0.054, 0.057]) and 0.778 ([0.755, 0.801]), respectively, implying strong spatial dependence and larger unexplained variation in

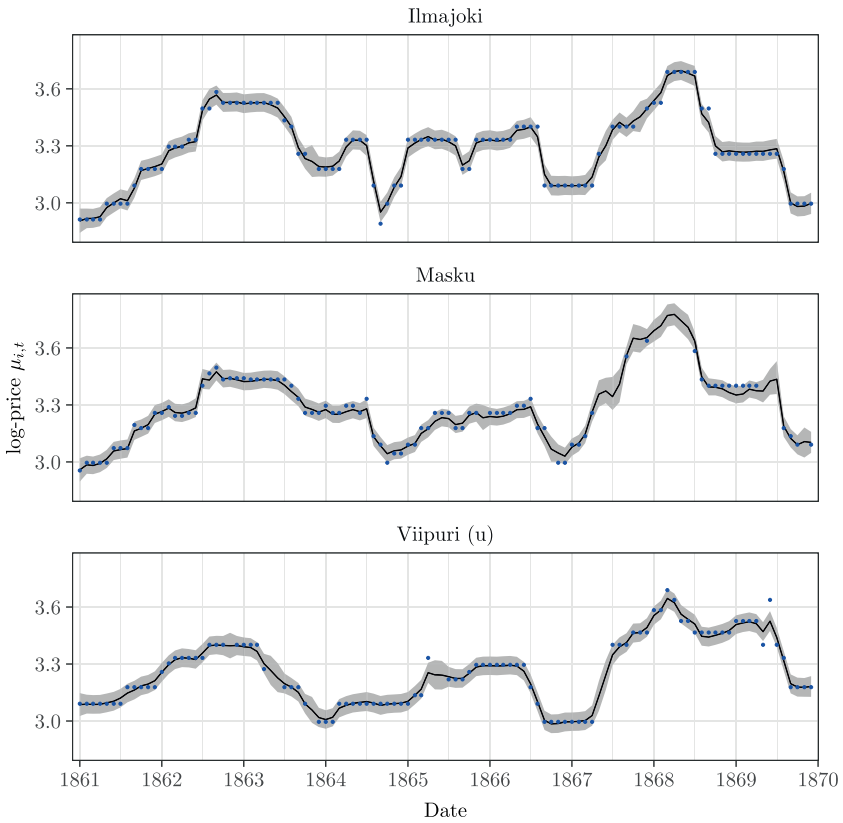

FIGURE 2 Posterior mean and 95% posterior interval for the unobserved log-price $\mu_{i,t}$ in Ilmajoki, Masku and Viipuri (urban) over the time period under study. The dots represent the observed log-prices $y_{i,t}$. [Colour figure can be viewed at wileyonlinelibrary.com]

the latent log-prices than our main model. The estimates of the $\mu_{i,t}$ resembled those in Figure 2 except that the posterior intervals were wider compared to our main model (see supplementary material on GitHub). Overall, such a simplification does not provide the necessary information to identify the spatio-temporal features we are interested in, for example, the short- and long-term decomposition. Thus, we do not consider it more thoroughly here. Note, however, that this alternative is a submodel of our more detailed approach.

4.1 | Coefficient estimates

Figure 3 shows the common time-varying component α_t, which captures not only the inherent seasonal variation but also other unidentified nationwide price variation, for example due to poor harvests. The parameters related to the AR process of α are estimated as $c_\alpha = 0$ with a 95% posterior interval $[-0.003, 0.004]$, $\phi = 0.619$ ($[0.461, 0.766]$) and $\sigma_\alpha = 0.017$ ($[0.013, 0.021]$). In general, the α_t varies around zero, although exhibiting some larger swings. These are mainly price increases due to harvest failures (e.g. in autumn 1862 and 1867) and price drops due to relatively successful new harvests (e.g. in autumn 1868). Figure 3 also shows the general tendency of prices to rise between two harvests (see also Figure 1). During the famine pinnacle of 1867–1868 this tendency was particularly pronounced.

The site-specific coefficients λ_i allowing spatial level differences from the nationwide trend α_t are plotted in Figure 4. The posterior medians vary between 0.224 and 1.650 among the rural sites (upper panel) and 0.654 and 1.927 among the urban sites (lower panel). The posterior mean of the standard deviation σ_λ is 0.463 with a 95% posterior interval [0.316, 0.632]. Overall, the coefficients corresponding to the urban towns are more concentrated to the high end of the range of λs, suggesting that urban towns reacted more strongly to the nationwide shocks captured by α. This stronger dependency could be related to the fact that towns had to purchase grain consumed from surrounding rural regions.

Figure 5 illustrates the short-term coefficients $\beta_{i,j}$ for each of the 298 neighbour pairs. We group also the coefficients $\beta_{i,j}$: rural regions and their rural neighbours, urban locations and their surrounding rural regions. The upper panel reports all rural–rural pairs (both $\beta_{i,j}$ and $\beta_{j,i}$ included), the middle panel reports the rural–urban pairs (i.e., $\beta_{\text{rural,urban}}$) and the lower panel reports

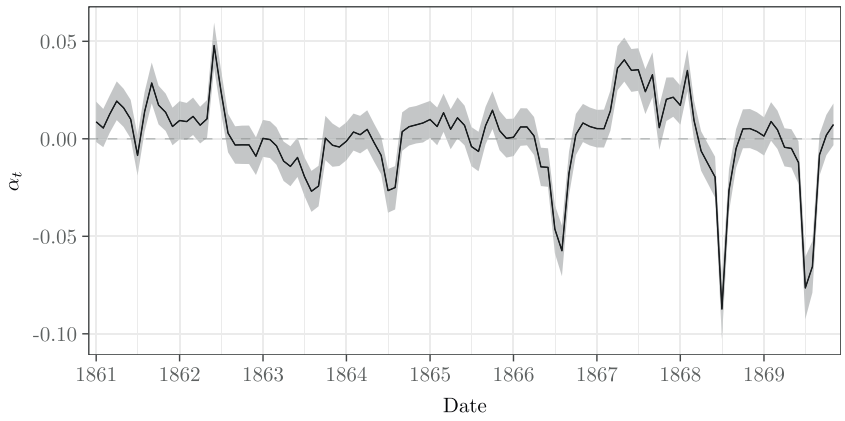

FIGURE 3 Posterior mean and 95% posterior interval for the trend α_t over the period under study.

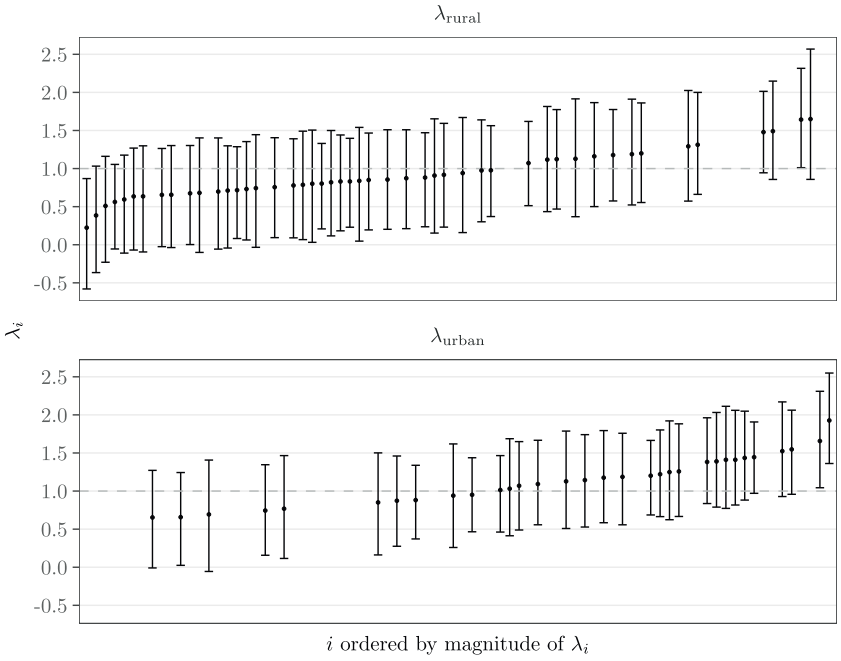

FIGURE 4 Posterior median and 95% posterior interval for the site specific coefficients λ_i by urban and rural groups. The estimates are ordered by their size in order to illustrate the difference between coefficients of urban and rural districts.

the urban–rural pairs ($\beta_{\text{urban,rural}}$). The dots represent the posterior medians of the individual coefficients $\beta_{i,j}$, which are all between 0.006 and 1.020.

As is visible from Figure 5, the short-run coefficients $\beta_{i,j}$ are of similar magnitude both in rural–rural and in rural–urban pairs (upper and middle panel). The vast majority of these coefficients are small and fall between 0 and 0.2, the latter implying a price increase of 0.2% in location i in response to a 1% increase in location j. This suggests that the short-run co-movement of prices (beyond aggregate fluctuations captured in α_t) in rural–rural and rural–urban pairs was weak.

Interestingly, we detected asymmetry in the price co-movement. The urban–rural pairs (lower panel) show that urban prices were generally more sensitive to follow the price development of the surrounding rural region than vice versa. Majority of the coefficients $\beta_{\text{urban,rural}}$ lies between 0.3 and 0.6, with multiple urban locations with coefficients $\beta_{i,j}$ above 0.5. This implies a larger than 0.5% increase in prices in response to 1% increase in the price level in the surrounding rural area.

This means that urban people were susceptible to market-transmitted shocks; conversely, rural prices were merely marginally affected by the urban demand pressure. In all likelihood, this is because urban consumers more frequently resorted to market purchases to obtain the grain consumed, thereby inducing a more developed market system in towns than in the rural regions (Devereux, 1988).

The long-term market adjustment is captured by the parameters $\gamma_{i,j,t}$. There are 298 time series of the error correction terms $\gamma_{i,j,t}$, one for each pair from January 1861 to December 1869, with all the posterior medians between −0.817 and −0.003 for the whole period under our study. The posterior mean of the standard deviation σ_γ is 0.204 with a 95% posterior interval

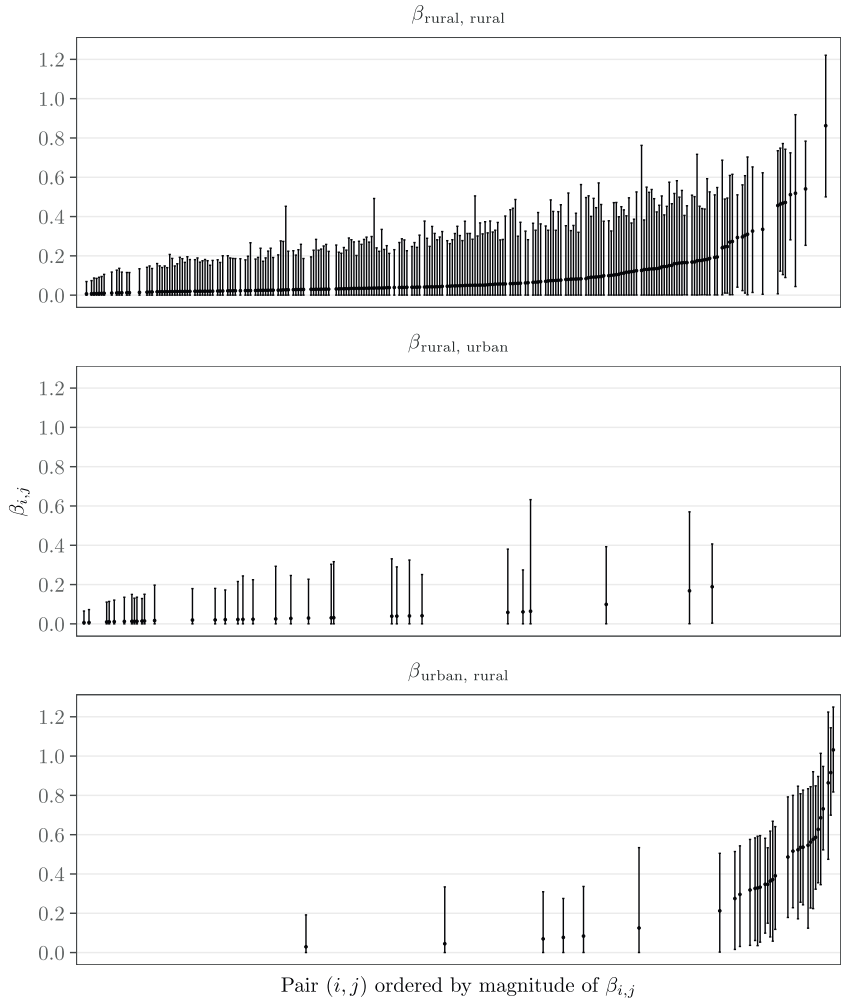

FIGURE 5 Medians and 95% posterior intervals of each coefficient $\beta_{i,j}$ grouped by top: rural districts with rural neighbours, middle: rural districts with urban neighbours and bottom: urban districts with rural neighbours.

[0.143, 0.273], confirming the need for time-varying coefficients (as time-invariant coefficients would correspond to $\sigma_\gamma = 0$).

In Figure 6 we present the coefficients $\gamma_{i,j,t}$ using the same grouping as with the coefficients $\beta_{i,j,t}$ in Figure 5. Most of the coefficients $\gamma_{i,j,t}$ are fairly steady over time, and there are no clear thresholds visible marking the start or end of the famine period. The upper panel shows the error correction coefficients for rural–rural pairs and the middle panel for rural–urban pairs. Both reveal that price correction to emerging disequilibrium was slow among the rural markets and between the rural–urban pairs. Furthermore, both show that while there are some pairs with speedier market adjustments during the famine, in the vast majority of rural–rural and rural–urban pairs, there are no worthwhile changes in the price transmission during the time span in question.

The lower panel in Figure 6 shows the coefficients $\gamma_{i,j,t}$ for urban–rural pairs illustrating how the urban prices adjust to the urban–rural price differentials. In many instances, the urban

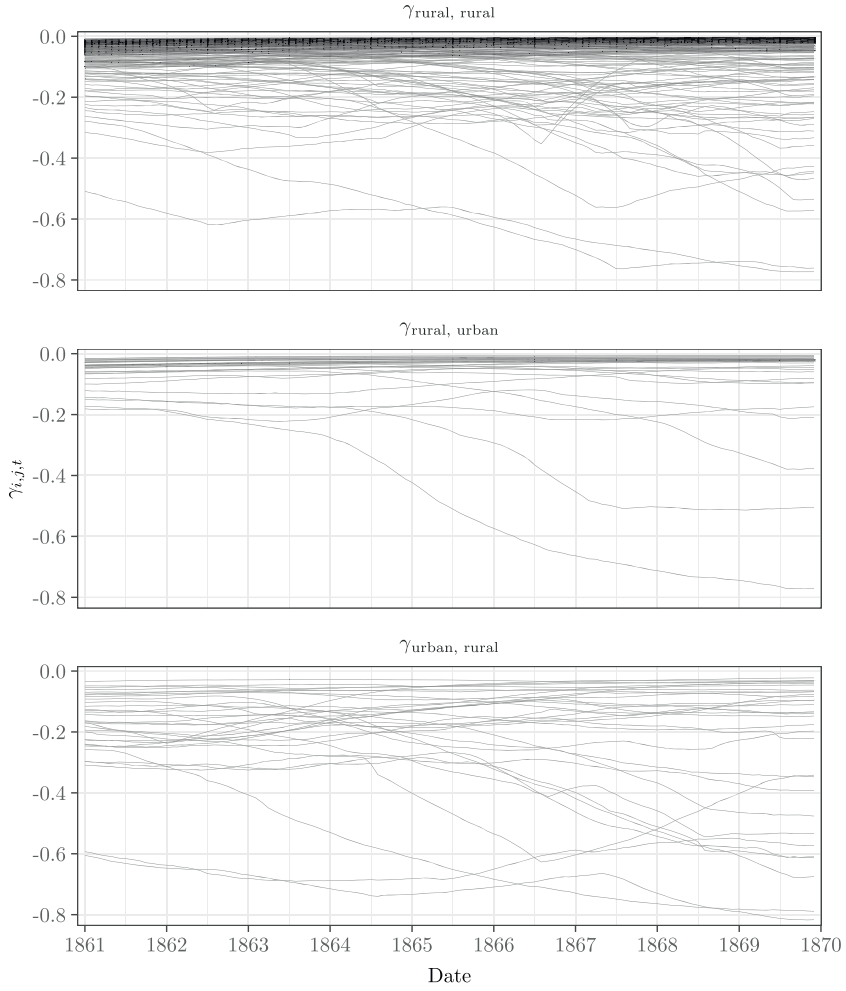

FIGURE 6 Medians of 20,000 posterior draws for all coefficients $\gamma_{i,j,t}$ by groups.

markets adjusted faster to emerging price disequilibrium during the famine than before it. The acceleration of the market adjustment during the famine was most pronounced in northern coastal towns (see supplementary maps on GitHub), but it also occurred in some inland towns. While more marked in the urban markets, the speedier error correction during the famine was not completely confined to towns. There were also some rural regions that witnessed a faster reaction to price differences, many of which were located in south and southwest Finland, with some regions also along the western coast and further inland. Examples of neighbour pairs are displayed in Figure 7 where those $\gamma_{i,j,t}$ series with values smaller than −0.55 at some time point alongside their counterpart series $\gamma_{j,i,t}$ are plotted.

Figures 6 and 7 show that the response of the coefficients $\gamma_{i,j,t}$ to the famine conditions does not happen abruptly, nor do they recover immediately after the famine. This emphasizes the benefits of time-varying coefficients. Many of the coefficients $\gamma_{i,j,t}$ begin to decrease in 1865–1866, some even earlier and do not recover before the end of our time span. This means that price transmissions accelerated much before the famine escalating crop failure in September 1867 and continued at that level after the famine had ceded.

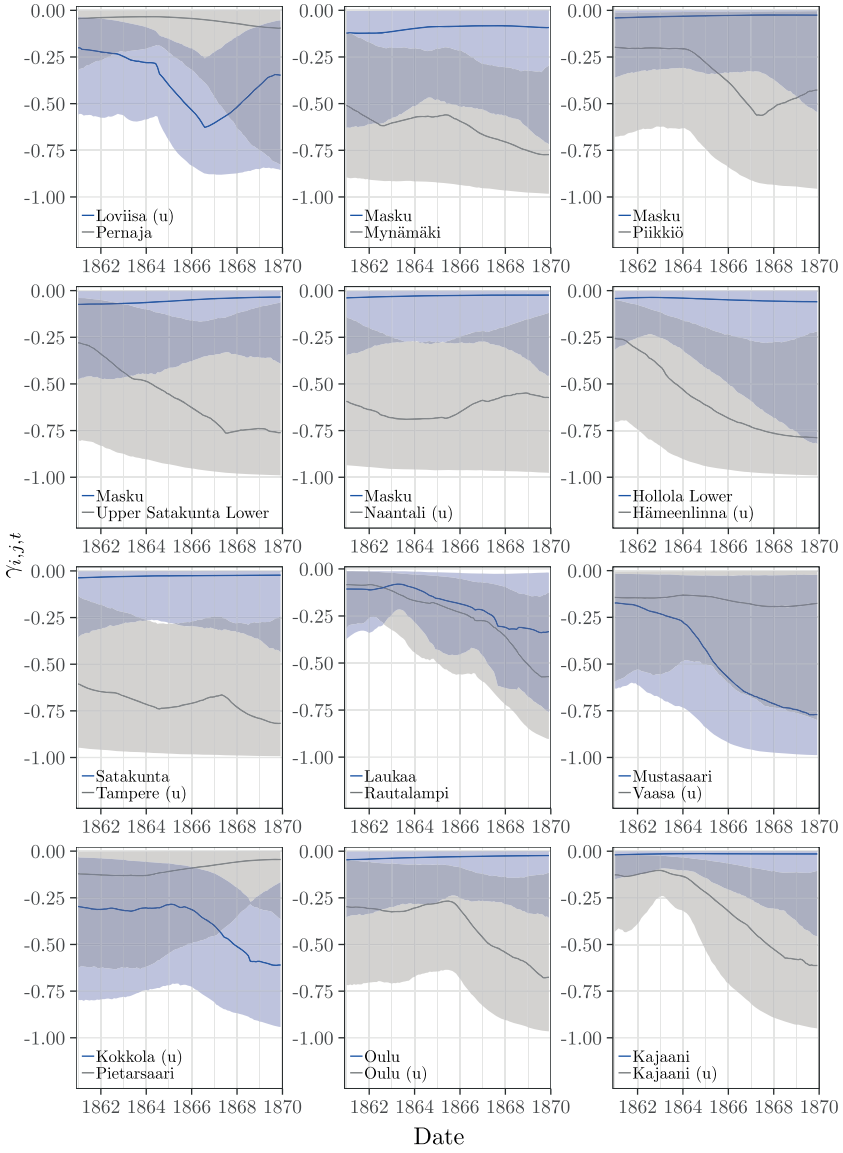

FIGURE 7 Posterior medians and 95% posterior intervals of the coefficients $\gamma_{i,j,t}$ and their counterparts $\gamma_{j,i,t}$ for those pairs where the posterior median is smaller than −0.55 at some time point. The name in the legend corresponds to the site i in the coefficient $\gamma_{i,j,t}$. [Colour figure can be viewed at wileyonlinelibrary.com]

4.2 | Visualizing spatial price propagation

To better understand the overall functioning of the market system and the importance of the close to 600 pairwise coefficients $\beta_{i,j}$ and $\gamma_{i,j,t}$, we examined the expected values of the latent log-prices μ_t in response to a price increase in one region. For this purpose, the trend α_t was set to zero for all t, and the error correction coefficients were selected from July (the last month uninfluenced by a new harvest) for the years 1861 (before famine) and 1868 (the peak year of the famine). The estimated coefficients $\beta_{i,j}$ were used as they were. The initial log-prices were fixed to equal

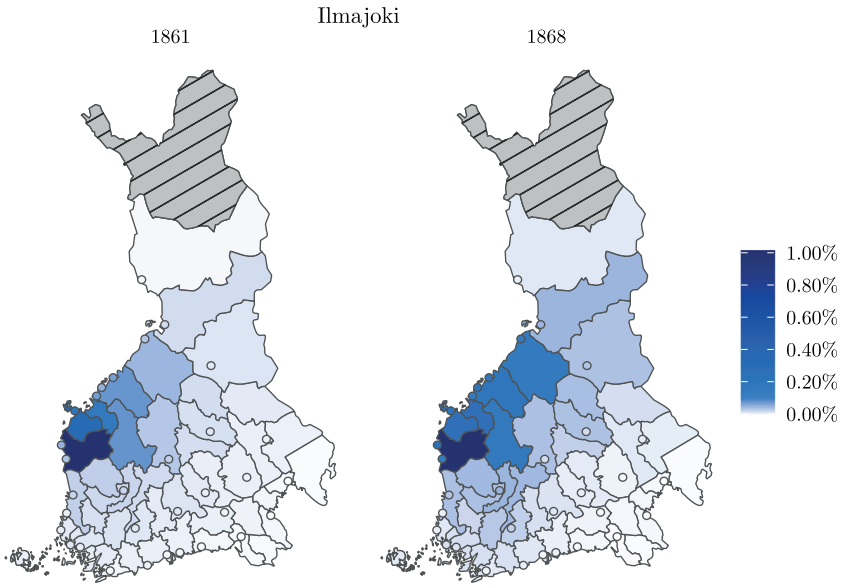

FIGURE 8 Regional maximum percentage increase in prices during a 12-month period with respect to the initial price due to a 1% increase in price in Ilmajoki (dark blue). Values are based on $\boldsymbol{\alpha} = \mathbf{0}$, posterior samples of $\beta_{i,j}$ and $\gamma_{i,j,t}$, with t fixed to July of the year marked in each panel. The striped grey area is not included in our study. [Colour figure can be viewed at wileyonlinelibrary.com]

values (as we were interested in the changes in prices, the actual value could be chosen arbitrarily), except for one specified place where the log-price was increased by 0.01, corresponding to approximately 1% price increase. The expected values were then calculated using the posterior samples of the coefficients $\beta_{i,j}$ and $\gamma_{i,j,t}$ and finally taking an average over all the simulated values of $\mu_{i,t}$. We restricted the simulation to 12 months. This corresponded to the annual harvest cycle. Furthermore, changes in prices after 12 months were generally negligible.

Figures 8–10 present the maximum percentage increase in regional prices due to a 1% increase in price in region j. As in Figure 2 we used Ilmajoki, Masku and Viipuri as examples.

Figure 8 depicts the effects of a 1% change in the rye price in the administrational district of Ilmajoki in July 1861 and 1868. The resulting maximum price increase is modest, and only in one region it surpasses 0.3% with respect to its initial level over the 12-month window. The price increase is larger north of Ilmajoki, and this tendency strengthens during the famine: in 1868, the price shock travels further northeastward and reaches further inland. The pattern agrees with the increased price transmission that happened in many Ostrobothnian coastal towns during the famine (see Figures 6 and 7 and supplementary maps on GitHub). The asymmetry of the spread is distinct: the price increase in the southern neighbouring district of Upper Satakunta Upper is at maximum 0.05% (in 1868) in response to a 1% shock whereas in the northern neighbouring district of Mustasaari the corresponding value is 0.28%.

Figure 9 depicts a price increase in the southwestern district of Masku. Unlike in the case of Ilmajoki, the spatial reach of the shock does not change much during the famine, but the price transmission intensifies close to the shock origin. For example, in the northern neighbouring region of Upper Satakunta Lower, a 1% price increase in Masku results in price increases of c. 0.23% and 0.64% in 1861 and 1868 respectively. The fact that the spatial reach remains constant probably reflects the fact that the southwestern Finnish markets were fairly well integrated to

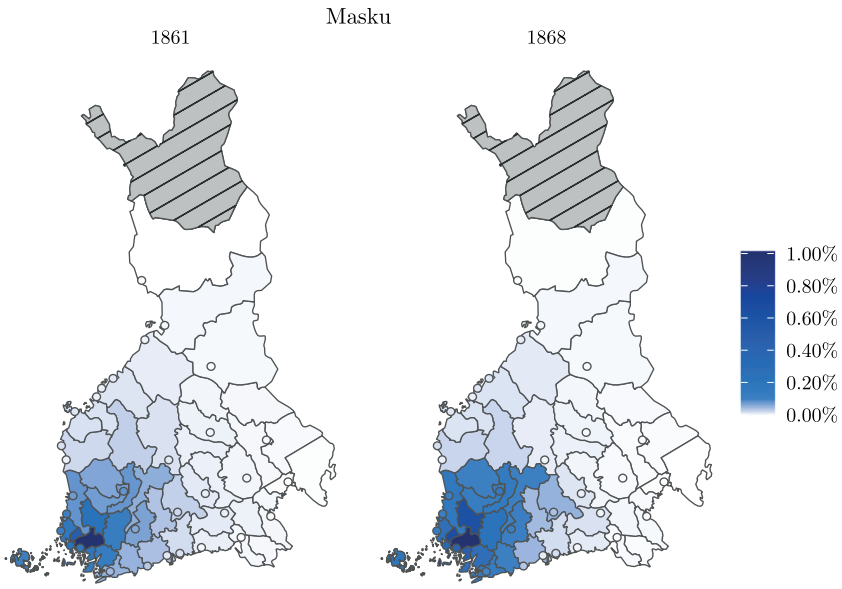

FIGURE 9 Regional maximum percentage increase in prices during a 12-month period with respect to the initial price due to a 1% increase in price in Masku (dark blue). Values are based on $\boldsymbol{\alpha} = \mathbf{0}$, posterior samples of $\beta_{i,j}$ and $\gamma_{i,j,t}$, with t fixed to July of the year marked in each panel. The striped grey area is not included in our study. [Colour figure can be viewed at wileyonlinelibrary.com]

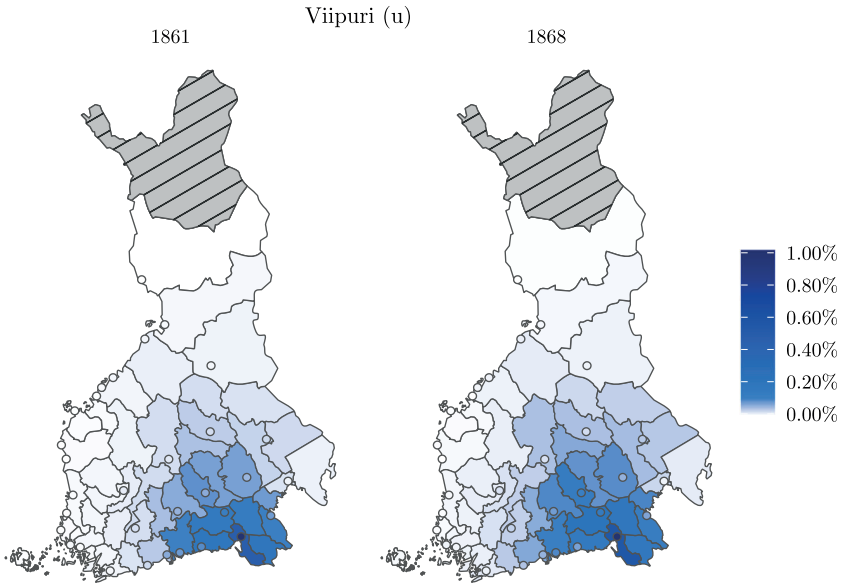

FIGURE 10 Regional maximum percentage increase in prices during a 12-month period with respect to the initial price due to a 1% increase in price in town of Viipuri (dark blue). Values are based on $\boldsymbol{\alpha} = \mathbf{0}$, posterior samples of $\beta_{i,j}$ and $\gamma_{i,j,t}$, with t fixed to July of the year marked in each panel. The striped grey area is not included in our study. [Colour figure can be viewed at wileyonlinelibrary.com]

begin with and suggests that well-integrated markets saw little change in their operation during the famine.

Figure 10 depicts a shock to the southeastern town of Viipuri. Here, the spread patterns are like those in Ilmajoki. The spatial spread of the shock is asymmetric, with more pronounced westward travel from the shock origin, without apparent difference between the years 1861 and 1868.

These maps suggest that the price propagation retained their pre-famine routes: the famine typically did not carve out new trade paths. Furthermore, the markets were fairly thin, and price transmission was strong only to places close-by to the shock origin. Further research is needed to tell whether observed route stabilities, spatial asymmetries and lack of long-distance price transmission stemmed from liabilities of existing trade relationships, lack of information or, for example insurmountable transportation costs.

5 | DISCUSSION

To analyse the grain market integration during the Finnish 1860s famine, we introduced a spatial context to the well-known error correction framework and modelled all the regional price time series simultaneously. This allowed us to omit the complex procedures to predetermine the market leader and the (time-varying) long-term relationships between the series. Also, allowing the common trend and error correction parameters to vary smoothly in time enabled us to estimate the temporal changes in the market integration without relying on artificial, sharp time period demarking the famine.

Depending on the available data and prior information, our model could be further modified and extended, for example, through more detailed modelling of the trend or error correction terms. For example, the common trend term α could be specified with some functional form of seasonal variation related to, for instance, harvests. We treated the error correction terms as independent random walks, but these could be allowed to depend on each other according to some (spatial) correlation structure or grouping. Time-varying components could also be assumed to vary more smoothly using integrated random walks, or piecewise constant given prior knowledge of potential the change points. It is also possible to let the short-run coefficients vary in time, although this induces a significant increase in the computational burden due to the subsequent time-varying covariance matrix in likelihood computations. Additional data on the regions' characteristics could be incorporated in the model to explain the differences in the short-run and long-run coefficients of the regions for a further insight of the market dynamics. Finally, we have defined the neighbourhood based on border sharing, but it could be interesting to study whether addition of second-order neighbours (neighbours of neighbours) would affect the results, especially for urban towns surrounded by a single rural region. Further extensions to enable simultaneous modelling of multiple products could also be an interesting future venue (an alternative agent-based modelling approach to model multivariate spatio-temporal panel data was used in Cun and Pesaran (2021)).

Regarding the patterns of market integration, we detected certain differences from those in the original contribution of Ó Gráda (2001). We estimated generally weaker price co-movement and, importantly, documented increased price transmission principally in the urban markets. In addition to modelling differences, it is likely that some of the differences can be attributed to the use of provincial aggregates in Ó Gráda (2001), which may overemphasize the role of small urban locations.

A possible reason for the increased price transmission during the famine may be the connection between an ill-developed pre-famine market system, the pre-famine prevalence of subsistence farming, and the introduction of deficit producers to the food markets after crop failures (Devereux, 1988). The fact that we observed little change in the behaviour of the reasonably well-developed southwestern and southeastern Finnish food markets during the famine aligns with this interpretation.

Our results show that the regional aspects of the early 19th century Finnish grain markets (e.g. Voutilainen et al., 2020) were still in place during the 1860s famine. The overall weak price transmission not only provides an explanation for the emergence of persistent east–west price gaps observed by Ó Gráda (2001), but also yields an important wider implication for the study of market integration. The results also suggest that moderate price dispersion (displayed in Figure 1) and reasonably high level of sigma convergence (measured in coefficient of variation of regional prices) are not universally coincided with efficient price transmission between the regional markets. Further research is needed to understand how low levels of price dispersion were achieved in this kind of setting and whether it was driven by certain key markets or reasonable symmetry of harvest outcomes. Our results show that the trade routes appeared robust to a large-scale harvest shock of 1867. Markets rarely changed their spatial orientation in response to this, even though the harvest failures came with substantial geographic variation. Transportation costs, lack of information and preexisting trade connections probably explain this.

Where markets behaved better during the famine, they facilitated the spread of the price shocks. The thin markets had the advantage of limiting the shocks to locations close by. The increased market transmission during the famine proliferated the spread of the shocks, especially along the western coast.

ACKNOWLEDGEMENTS

Tiia-Maria Pasanen was supported by the Finnish Cultural Foundation. Jouni Helske was supported by the Academy of Finland grants 331817 and 311877. Miikka Voutilainen was supported by the Academy of Finland grant 308975. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

ORCID

Tiia-Maria Pasanen https://orcid.org/0000-0002-3195-0928
Jouni Helske https://orcid.org/0000-0001-7130-793X

REFERENCES

Alfani, G. & Ó Gráda, C. (2018) The timing and causes of famines in Europe. *Nature Sustainability*, 1, 283–288.

Alogoskoufis, G. & Smith, R. (1991) On error correction models: specification, interpretation, estimation. *Journal of Economic Surveys*, 5, 97–128.

Andrée, B.P.J., Chamorro, A., Kraay, A., Spencer, P. & Wang, D. (2020) Predicting food crises. *World Bank Working Paper 9412*.

Anselin, L. (1988) *Spatial econometrics: methods and models*. Dordrecht: Kluwer.

Aquaro, M., Bailey, N. & Pesaran, M.H. (2021) Estimation and inference for spatial models with heterogeneous coefficients: an application to US house prices. *Journal of Applied Econometrics*, 36, 18–44. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2792

Araujo Bonjean, C. & Simonet, C. (2016) Are grain markets in Niger driven by speculation? *Oxford Economic Papers*, 68, 714–735.

Bai, J. & Wang, P. (2015) Identification and Bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics*, 33, 221–240. Available from: https://doi.org/10.1080/07350015.2014.941467

Campbell, B.M. (2010) Nature as historical protagonist: environment and society in preindustrial England. *The Economic History Review*, 63, 281–314.
van Campenhout, B. (2007) Modelling trends in food market integration: method and an application to Tanzanian maize markets. *Food Policy*, 32, 112–127.
Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M. et al. (2017) Stan: a probabilistic programming language. *Journal of Statistical Software, Articles*, 76, 1–32. Available from: https://www.jstatsoft.org/v076/i01
Chilosi, D., Murphy, T.E., Studer, R. & Tunçer, A.C. (2013) Europe's many integrations: geography and grain markets, 1620–1913. *Explorations in Economic History*, 50, 46–68.
Cun, W. & Pesaran, M.H. (2021) A spatiotemporal equilibrium model of migration and housing interlinkages. *CESifo Working Paper Series 9343*, CESifo. Available from: https://ideas.repec.org/p/ces/ceswps/_9343.html
Deng, L.B. (2007) Increased rural vulnerability in the era of globalization: conflict and famine in Sudan during the 1990s. In: Devereux, S. (Ed.) *The new famines: why famines persist in an era of globalization*. Abingdon: Routledge, pp. 245–268.
Devereux, S. (1988) Entitlements, availability and famine: a revisionist view of Wollo, 1972–1974. *Food Policy*, 13, 270–282.
Devereux, S. (2009) Why does famine persist in Africa? *Food Security*, 1, 25–35.
Devereux, S. & Tiba, Z. (2007) Malawi's first famine, 2001–2002. In: Devereux, S. (Ed.) *New famines: why famines persist in an era of globalization*. Abingdon: Routledge, pp. 143–177.
Engle, R.F. & Granger, C.W.J. (1987) Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society*, 251–276.
Fackler, P.L. & Tastan, H. (2008) Estimating the degree of market integration. *American Journal of Agricultural Economics*, 90, 69–85. Available from: http://www.jstor.org/stable/30139492
Federico, G., Schulze, M. & Volckart, O. (2021) European goods market integration in the very long run: from the black death to the first world war. *Journal of Economic History*, 81, 276–308.
Garenne, M. (2007) An atypical urban famine: Antananarivo, Madagascar 1985–1986. In: Devereux, S. (Ed.) *The new famines: why famines persist in an era of globalization*. Abingdon: Routledge, pp. 178–196.
Giles, J.A. & Giles, D.E.A. (1993) Pre-test estimation and testing in econometrics: recent developments. *Journal of Economic Surveys*, 7, 145–197. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6419.1993.tb00163.x
Giles, D.E. & Godwin, R.T. (2012) Testing for multivariate cointegration in the presence of structural breaks: *p*-values and critical values. *Applied Economics Letters*, 19, 1561–1565. Available from: https://doi.org/10.1080/13504851.2011.639727. Publisher: Routledge.
Gonzalo, J. & Lee, T.-H. (1998) Pitfalls in testing for long run relationships. *Journal of Econometrics*, 86, 129–154. Available from: https://www.sciencedirect.com/science/article/pii/S0304407697001115
Hjalmarsson, E. & Österholm, P. (2010) Testing for cointegration using the Johansen methodology when variables are near-integrated: size distortions and partial remedies. *Empirical Economics*, 39, 51–76.
Hoffman, M.D. & Gelman, A. (2014) The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15, 1593–1623.
Howe, P. & Devereux, S. (2004) Famine intensity and magnitude scales: a proposal for an instrumental definition of famine. *Disasters*, 28, 353–372.
Hoyle, R.W. (2010) Famine as agricultural catastrophe: the crisis of 1622–4 in east Lancashire. *The Economic History Review*, 63, 974–1002.
Hunter, J. & Ogasawara, K. (2019) Price shocks in regional markets: Japan's Great Kantō Earthquake of 1923. *The Economic History Review*, 72, 1335–1362.
Juselius, K. (2006) *The cointegrated VAR model: methodology and applications*. Oxford: Oxford University Press.
Kelejian, H.H. & Prucha, I.R. (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157, 53–67.
Kennedy, P. (2008) *A guide to econometrics*. Malden: Blackwell.
Li, G., Wong, K.K., Song, H. & Witt, S.F. (2006) Tourism demand forecasting: a time varying parameter error correction model. *Journal of Travel Research*, 45, 175–185.

Li, F., Wang, Z. & Liu, G. (2013) Towards an error correction model for dam monitoring data analysis based on cointegration theory. *Structural Safety*, 43, 12–20.
Møller, N.F. & Sharp, P. (2014) Malthus in cointegration space: evidence of a post-Malthusian pre-industrial England. *Journal of Economic Growth*, 19, 105–140.
Macrae, J. & Zwi, A.B. (1992) Food as an instrument of war in contemporary African famines: a review of the evidence. *Disasters*, 16, 299–321.
Marcus, D. (2003) Famine crimes in international law. *American Journal of International Law*, 97, 245–281.
Matz, J.A., Kalkuhl, M. & Abegaz, G.A. (2015) The short-term impact of price shocks on food security—evidence from urban and rural Ethiopia. *Food Security*, 7, 657–679.
Maxwell, D., Khalif, A., Hailey, P. & Checchi, F. (2020) Determining famine: multidimensional analysis for the twenty-first century. *Food Policy*, 92, 101832.
Mishra, V., Tiwari, A.D., Aadhar, S., Shah, R., Xiao, M., Pai, D.S. et al. (2019) Drought and famine in India, 1870–2016. *Geophysical Research Letters*, 46, 2075–2083.
Nkang, N.M., Abang, S.O., Akpan, O.E. & Edet, E.O. (2006) Rice production, imports, and food security in Nigeria: an application of cointegration and error correction model. *Journal of Food Agriculture and Environment*, 4, 86.
Ó Gráda, C. (2001) Markets and famines: evidence from nineteenth-century Finland. *Economic Development and Cultural Change*, 49, 575–590.
Ó Gráda, C. (2005) Markets and famines in pre-industrial Europe. *Journal of Interdisciplinary History*, 36, 143–166.
Ó Gráda, C. (2015) *Eating people is wrong, and other essays on famine, its past, and its future*. Princeton, NJ: Princeton University Press.
Ó Gráda, C. & Chevet, J.-M. (2002) Famine and market in Ancien Régime France. *Journal of Economic History*, 62, 706–733.
Olsson, L. (1993) On the causes of famine: drought, desertification and market failure in the Sudan. *Ambio*, 22, 395–403.
Ord, K. (1975) Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70, 120–126.
Persson, K.G. (1999) *Grain markets in Europe, 1500–1900: integration and deregulation*. Cambridge: Cambridge University Press.
Pitkänen, K.J. (1993) *Deprivation and disease: mortality during the great Finnish famine of the 1860s*. Helsinki: Suomen Väestötieteen yhdistys.
Quddus, M. & Becker, C. (2000) Speculative price bubbles in the rice market and the 1974 Bangladesh famine. *Journal of Economic Development*, 25, 155–175.
R Core Team. (2021) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/
Ramajo, J. (2001) Time-varying parameter error correction models: the demand for money in Venezuela, 1983. I-1994. IV. *Applied Economics*, 33, 771–782.
Ravallion, M. (1987) *Markets and famines*. Oxford: Clarendon Press.
Seaman, J. & Holt, J. (1980) Markets and famines in the Third World. *Disasters*, 4, 283–297.
Shin, M. (2010) A geospatial analysis of market integration: the case of the 2004/5 food crisis in Niger. *Food Security*, 2, 261–269.
Slavin, P. (2014) Market failure during the Great Famine in England and Wales (1315–1317). *Past and Present*, 222, 9–49.
Soininen, A.M. (1974) *Vanha maataloutemme: maatalous ja maatalousväestö Suomessa perinnäisen maatalouden loppukaudella 1720-luvulta 1870-luvulle*. Helsinki: Suomen historiallinen seura.
Stan Development Team. (2020) RStan: the R interface to Stan. Available from: http://mc-stan.org/. R package version 2.21.2.
Studer, R. (2008) India and the great divergence: assessing the efficiency of grain markets in eighteenth and nineteenth-century India. *The Journal of Economic History*, 68, 393–437.
Svanidze, M., Götz, L., Djuric, I. & Glauben, T. (2019) Food security and the functioning of wheat markets in Eurasia: a comparative price transmission analysis for the countries of Central Asia and the South Caucasus. *Food Security*, 11, 733–752.
Voutilainen, M. (2016) *Poverty, inequality and the Finnish 1860s famine*. Jyväskylä: University of Jyväskylä.

Voutilainen, M., Turunen, R. & Ojala, J. (2020) Multi-currency regime and markets in early nineteenth-century Finland. *Financial History Review*, 27, 115–138.
de Waal, A. (1993) War and famine in Africa. *Ids Bulletin*, 24, 33–40.

How to cite this article: Pasanen, T.-M., Voutilainen, M., Helske, J. & Högmander, H. (2022) A Bayesian spatio-temporal analysis of markets during the Finnish 1860s famine. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 71(5), 1282–1302. Available from: https://doi.org/10.1111/rssc.12577

II

SPATIO-TEMPORAL MODELING OF CO-DYNAMICS OF SMALLPOX, MEASLES, AND PERTUSSIS IN PRE-HEALTHCARE FINLAND

by

Pasanen, T.-M., Helske, J., Högmander, H., and Ketola, T. 2024

PeerJ 12:e18155

DOI: https://doi.org/10.7717/peerj.18155

Spatio-temporal modeling of co-dynamics of smallpox, measles, and pertussis in pre-healthcare Finland

Tiia-Maria Pasanen[1], Jouni Helske[1,2], Harri Högmander[1] and Tarmo Ketola[3,4]

[1] Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland
[2] INVEST Research Flagship Centre, University of Turku, Turku, Finland
[3] Department of Forestry, University of Helsinki, Helsinki, Finland
[4] Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

ABSTRACT

Infections are known to interact as previous infections may have an effect on risk of succumbing to a new infection. The co-dynamics can be mediated by immunosuppression or modulation, shared environmental or climatic drivers, or competition for susceptible hosts. Research and statistical methods in epidemiology often concentrate on large pooled datasets, or high quality data from cities, leaving rural areas underrepresented in literature. Data considering rural populations are typically sparse and scarce, especially in the case of historical data sources, which may introduce considerable methodological challenges. In order to overcome many obstacles due to such data, we present a general Bayesian spatio-temporal model for disease co-dynamics. Applying the proposed model on historical (1820–1850) Finnish parish register data, we study the spread of infectious diseases in pre-healthcare Finland. We observe that measles, pertussis, and smallpox exhibit positively correlated dynamics, which could be attributed to immunosuppressive effects or, for example, the general weakening of the population due to recurring infections or poor nutritional conditions.

Subjects Ecology, Microbiology, Epidemiology, Statistics, Data Science
Keywords Spatio-temporal, Infection co-dynamics, Pertussis, Measles, Smallpox, Bayesian analysis

Submitted 6 February 2024
Accepted 1 September 2024
Published 26 September 2024

Corresponding author
Tiia-Maria Pasanen, tiia-maria.h.pasanen@jyu.fi

Academic editor
Nicholas Geard

Additional Information and Declarations can be found on page 20

DOI **10.7717/peerj.18155**

INTRODUCTION

Infections exist rarely in isolation, and their effects on hosts are known to interact and to have both positive and negative relationships between each other (*Gupta, Ferguson & Anderson, 1998*; *Rohani et al., 2003*; *Shrestha et al., 2013*; *Mina et al., 2015*; *Nickbakhsh et al., 2019*). For example, cross immunity may prevent others from infecting the host, competition for the same resources or susceptible host can have strong effects on epidemics, and sometimes one infection paves a way for another (*Gupta, Ferguson & Anderson, 1998*; *Rohani et al., 2003*; *Graham, 2008*). Perhaps historically the best-known relationship between infections is the immunosuppressive effect of measles on the following pertussis epidemic by increasing the severity of the epidemic (see *Coleman, 2015*; *Mina et al., 2015*; *Noori & Rohani, 2019*). Coinfections of parasites (*Graham, 2008*) and viruses and respiratory bacterial infections are well known (e.g., *Bakaletz, 2017*; *Wong et al., 2023*), whereas understanding coinfections and cotransmissions of, for example, zika, dengue,

and chikungunya viruses presents a current serious challenge for public health (*Vogels et al., 2019*).

Demographic consequences of epidemics are most dramatically seen in large cities and in densely populated areas, which is reflected in the epidemiological research in general (*Mueller et al., 2020*). However, as rural areas constitute a large part of most countries, the spatio-temporal dynamics of epidemics in populations with low densities deserve more attention (*Mueller et al., 2020*). In rural areas, populations often consist of loosely connected metapopulations rather than large and epidemiologically more autonomous populations in cities. This has most likely strong repercussions for the drivers of epidemics (*Ball et al., 2015*) and also for the co-occurrence of infections. However, these issues are rarely addressed in the literature, possibly due to the statistical challenges encountered with sparse and scarce data, as well as the difficulty of modeling the dynamics of several infections simultaneously both in space and time.

The discrepancy between studying dense and sparse populations is evident and can be seen, for example, by comparing our case of rural Finland in 1820–1850 to the seminal research of *Rohani et al. (2003)*. Their study is based on five large European cities, where the weekly number of fatalities frequently exceeds 30 and even 80. In our data, the recorded incidents in most of the towns rarely exceed one person per month, as we study a small and mainly agrarian population in the southern part of Finland with circa 1.2 to 1.6 million individuals (*Voutilainen, Helske & Högmander, 2020*). The population, without proper healthcare (*Saarivirta, Consoli & Dhondt, 2012*), was spread over a vast area in geographically separated, but socially connected, small towns and villages. Based on the data from 1882, population sizes of towns varied between 300 and 25,000 (*Statistical Office of Finland, 1882*; *Ketola et al., 2021*). Despite the obvious uncertainty of population censuses during that era (*Voutilainen, Helske & Högmander, 2020*), the contrast between our data and most of the published datasets is striking. Statistical modeling of such data is problematic due to incomplete information from some locations and the rare occurrence of events, hampering the ability of generally used models to consider several infectious diseases at the same time and on both temporal and spatial scales.

To estimate the spatio-temporal co-dynamics of deaths due to pertussis, measles, and smallpox, we build a model that can overcome the limitations inherent in our data. The model jointly estimates the spread of multiple infections, enabling the exploration of the temporal and spatial dependence structures both within and between the infections. Our general Bayesian model consists of a multivariate latent incidence process, a seasonal component, and multiple predictors whose effects may vary between the towns. This allows us to study the dynamics of the diseases simultaneously despite having only incomplete information about the deaths. The results we get from modeling the mere presence-absence data are compared with those of modeling the corresponding death counts, and the simplification is deemed to be a reasonable option in our case. Given the limitations of our data, we do not aim to make causal claims on a biological level, and rather than focusing on the magnitude of infections and the intensity of deaths, our primary interest lies in understanding how the prevalence of these three diseases varied both spatially

and temporally in pre-industrial Finland, and if there were possible associations in these dynamics across the diseases.[1]

[1]This text was originally published as a preprint (https://export.arxiv.org/abs/2310.06538).

MATERIALS AND METHODS

Data

During the study period, 1820–1850, the parishes in Finland kept track, among others, of baptisms, burials and causes of deaths, according to common and long held principles (*Pitkänen, 1977*). Even though the death diagnostics were based on symptoms, some infections can be considered to be diagnosed rather accurately due to their characteristic features. These diseases include pertussis (whooping cough), measles, and smallpox, which we consider. These infections were the main reason for child mortality, and, overall, they were responsible for approximately 5, 3, and 3 percent of total deaths, respectively, according to our data. Based also on the available records in our data, the median ages of deaths in complete years were 0 (sd = 3.6) for pertussis, 2 (sd = 3.9) for measles, and 2 (sd = 7.5) years for smallpox. Smallpox vaccinations were started in Finland in 1802 and were slowly progressing during the study period (*Briga, Ketola & Lummaa, 2022*). However, general healthcare was almost non-existent as in 1820 there were only 373 hospital sickbeds for 1.2 million inhabitants (*Saarivirta, Consoli & Dhondt, 2012*).

Our data consist of the daily numbers of deaths, classified by the cause of death, between January 1820 and December 1850 from $N = 387$ different regions (towns) in mainland Finland with the exclusion of northern areas. The time window is chosen such that there were no major famines, wars, border changes, or other potentially confounding events, which could have altered the geographical partition or the dynamics of the epidemics. The general stability achieved is advantageous in the modeling.

Although using a daily time scale would, theoretically, be ideal for modeling disease dynamics, the infrequency of deaths implies that our data lack sufficient information to study temporal dependencies at such a detailed level. This issue is likely further pronounced by the spatial heterogeneity of the data and the potentially complex lagged auto- and cross-dependencies between the infection dynamics of the diseases. Therefore, the daily counts of deaths are aggregated over time into a monthly level, decreasing the number of zero observations yet maintaining a reasonable time resolution for observing the spread of the diseases on our geographical scale. This yields a total of $T = 372$ time points. The counts of the observed numbers of deaths by disease, considering the aggregated data, are visualized in Fig. 1. In each case of the three infections, about 93–96% of the death counts are zero or one, and less than 2% of the counts are more than three, despite the aggregation. The reliability of the actual death counts varies considerably both temporally and spatially owing to the heterogeneous quality of the parish records and the cause of death classifications. Moreover, these deficiencies are not necessarily independent of the true number of deaths. It is also noteworthy that despite the records of baptisms and burials, there are no reliable estimates of the population sizes at the town level (*Voutilainen, Helske & Högmander, 2020*). Hence the intuitive idea of using local relative mortality is unfortunately beyond reach. Because of this and the aforementioned reliability issues—and since most of the

counts are still either zero or one even after the aggregation to a monthly level—only the dichotomous knowledge of the death occurrence is used in the main analysis. Even so, the count data are considered in the model comparison section.

About 24% of the data concerning death occurrences in a town and a month are missing, and from 57 out of the overall 387 regions there are no observations at all. The missingness pattern is common to all the diseases, as for a particular town and month, we have observed the death counts for all the diseases or for none of them. This is because the missing data can be attributed to the absence of parish records that document all deaths. Such unavailability of data may be a result of incomplete digitization of the parish records, loss of the documents for example due to a fire in the local rectory or church, or simply because there were no deaths in a given month. Therefore, the missing data could potentially depend on the unknown regional population size, but not on the specific cause of death, given that the proportions of deaths attributable to the particular cause of death are relatively small. Nevertheless, our proposed model provides estimates of the monthly probabilities of observing at least one death also for the towns with missing observations.

Model

We construct a general model to describe the spatial and temporal dependencies both within and between the infections under study. We also want to enable exploiting other relevant information as explanatory variables. In epidemiological context, there typically occur spatial or temporal trends or seasonal effects, which can be included as separate components in the model. Due to the nature of our data, we model the probability of observing at least one death caused by a disease in a certain town in a certain month. The model consists of a trend, a seasonal effect, and a regression part reflecting the local effects of the previous state of infection in the focal town and its neighboring towns.

Formally, let $y_{i,t}^d$ denote a dichotomous variable of an event where at least one death occurs due to a disease d in a region $i = 1, \ldots, N$ at a time point $t = 1, \ldots, T$, where N is the number of regions and T the number of time points. Let K_x^d indicate the number of explanatory variables x, based on the features of the region i. Accordingly, z denotes the explanatory variables, and K_z^d their number, related to the neighborhood of the region i. Thus the model for observing at least one death caused by the disease d in the region i at the time point t can be written as follows:

$$y_{i,t}^d \sim \text{Bernoulli}\left(\text{logit}^{-1}\left(\eta_{i,t}^d\right)\right), \tag{1}$$

where

$$\eta_{i,t}^d = \lambda_i^d \tau_t^d + s_t^d + a_i^d + b_i^d \sum_{k=1}^{K_x^d} \beta_k^d x_{i,k}^d + c_i^d \sum_{k=1}^{K_z^d} \gamma_k^d z_{i,k}^d. \tag{2}$$

Here Bernoulli distribution with a logit link is chosen due to our dichotomous consideration of the occurrences of death, but other distributions with appropriate link functions can be applied for different types of response variables.

The first three terms being summed in Eq. (2) form a base level, in our case, for the probability of observing at least one death caused by the disease d at each town and month.

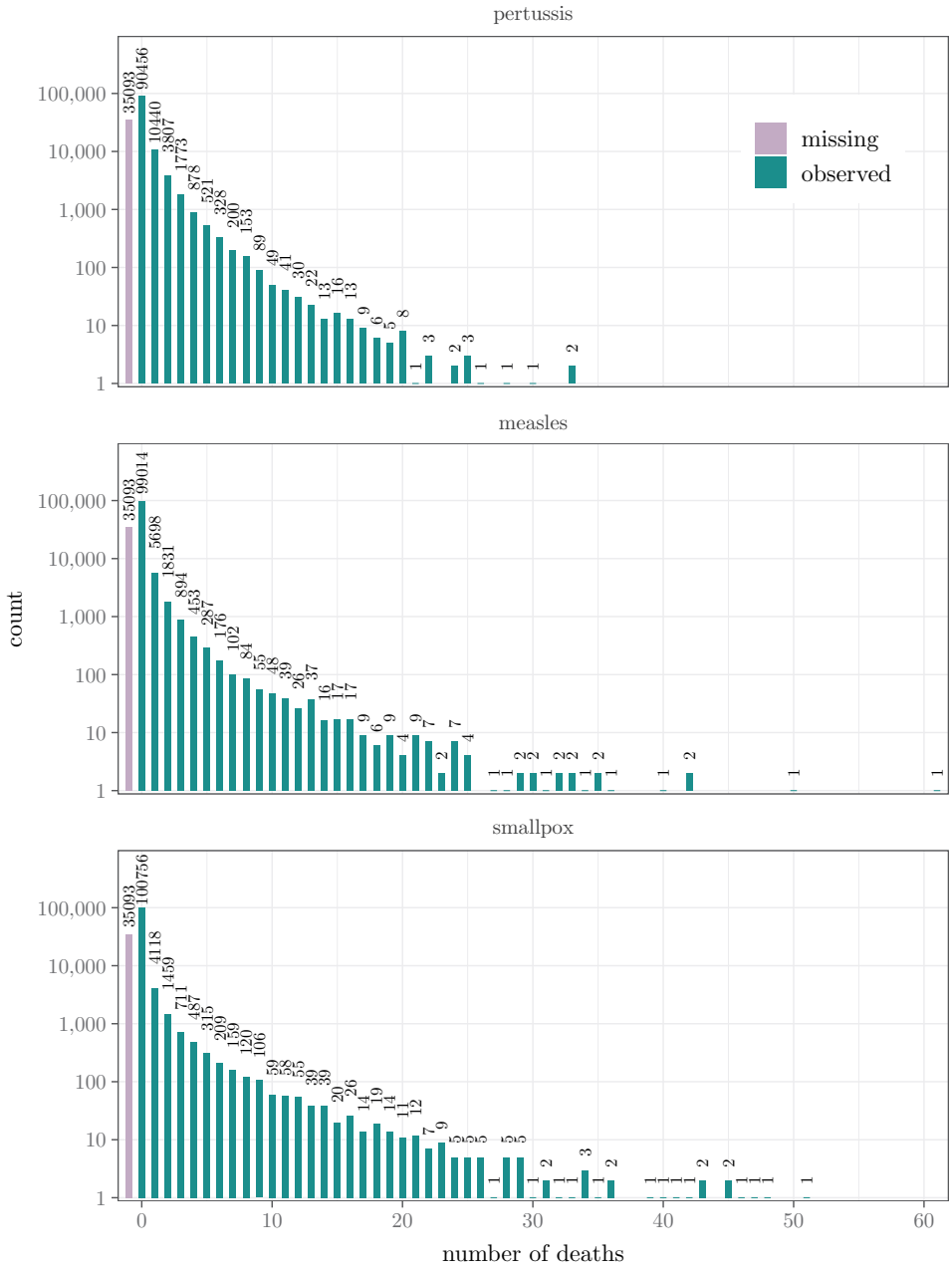

Figure 1 The counts of the observed numbers of deaths over all towns and months plotted by disease. The first bar indicates the number of missing observations. The numbers above the bars are the counts. Note that the vertical axis is on a logarithmic scale.

Full-size DOI: 10.7717/peerj.18155/fig-1

More specifically, the first term consists of the time dependent latent factor τ_t, describing the nationwide incidence (on log-odds scale), or trend, of the disease d, and the regional adjustments or loadings λ_i, with respect to the mean level. As in general dynamic factor models, the products $\lambda_i\tau_t$ are not identifiable without constraints (*Bai & Wang, 2015*). Instead of the common approach of fixing one of the loadings λ_i to 1, we constrain the mean of the loadings to 1, enabling the interpretation of the factor τ_t as the nationwide incidence level. Due to the nature of the other terms in Eq. (2), this incidence level gives the national average log-odds of observing at least one death in an "average" town in a given month if no deaths were observed in the previous month in the focal town or in its neighboring towns. The second term s_t is a monthly seasonal effect, which is the average deviation from the nationwide incidence level, summing up to 0 over the months. The third term a_i is a regional, zero-mean constant reflecting local deviations from the nationwide incidence level τ_t due to unobserved local demographic, geographic, social or other characteristics associated with mortality.

The last two sum terms in Eq. (2) form the regression part of the model. The first sum includes the covariates $x_{i,k}$ related to the focal region i, and the second sum the covariates $z_{i,k}$ related to the neighboring regions. These variables have both nationwide coefficients β and γ, and their local adjustments b_i and c_i amplifying or diminishing the nationwide level. As the regional constants a_i, also the multiplicative local adjustments may account for any unobserved heterogeneity between the regions, such as the local population sizes or densities. The adjustment parameter b_i reflects the features of the focal region i, and c_i those of the neighborhood of the region i (possibly relative to i). We assume that b_i is the same for all covariates x_k, and, accordingly, c_i for all z_k, since the underlying regional characteristics modifying the nationwide mortality effects β and γ do not depend on the covariates.

In our study, there are three covariates assigned to the town i and another three to its neighbors for all diseases $d = p, m, s$, where p stands for pertussis, m measles, and s smallpox. The local explanatory variables $x_{i,k}$ are the presences of deaths caused by the three different diseases in the previous month in the focal town, whereas the regional neighborhood predictors $z_{i,k}$ are the averages of the same presences of deaths over the local neighborhood. We define two regions being neighbors when they share a border. By this definition, all the towns have at least one neighbor. Other definitions of neighborhood could be used as well, for example based on the transportation network or distance, leading to weighted averages of death occurrences.

As noted earlier, our data contain a large number of missing observations. We assume that the probability of having a missing response or predictor variable is independent of the value of the response variable. This presumption may be considered valid, given that the lack of parish records on deaths is unlikely to depend on the causes of deaths in a particular month. Under this assumption, we can model the observed data analogously to a complete-case analysis in an unbiased manner, assuming our model is correctly specified (*van Buuren, 2018*). This eliminates the need for multiple imputation or sampling missing observations using MCMC algorithms which do not use gradient information (*i.e.,*

algorithms capable of sampling discrete variables), both of which would be computationally unfeasible in our Bayesian spatio-temporal context.

In practice, the complete-case analysis in our context means that to use an observation as a response, we require that both the current response variable as well as all of the related covariates are observed. If any of them is missing, we omit the particular combination of town and month as a response. On the other hand, when calculating the neighborhood covariates $z_{i,k}$, which in our case are the averages over the observations within the neighborhood, we omit the neighbors with missing observations so that they are not included even as a denominator in the evaluation of the mean. If all neighbor observations are missing, the corresponding covariate is defined as missing.

We model the latent factor $\tau_t = (\tau_t^p, \tau_t^m, \tau_t^s)$, the temporal process describing the baseline of the nationwide incidence rates, as an intercorrelated random walk, $\tau_{t+1} \sim \mathrm{N}(\tau_t, \Sigma)$. Here Σ is an unconstrained 3×3 covariance matrix parametrized using the standard deviations σ_τ^d and the correlation matrix R. In our application this latent process, together with the regional constant and the monthly effect, can be interpreted as the probability to observe at least one new death in a particular town when no deaths caused by any of the three diseases were observed in the previous month in the focal town or in its neighborhood.

In the Bayesian modeling framework, we need prior distributions for all the parameters to be estimated. The incidence factors τ^d follow a $\mathrm{N}(-2, 2^2)$ prior at the first time point and form a random walk at the later time points. For the correlation matrix R we use an LKJ(1) prior with a Cholesky parametrization (*Lewandowski, Kurowicka & Joe, 2009*), *i.e.*, a uniform prior over valid 3×3 correlation matrices. For the regional parameters we apply Gaussian priors: $\lambda_i^d \sim \mathrm{N}(1, \sigma^2_{\lambda^d})$, $a_i^d \sim \mathrm{N}(0, \sigma^2_{a^d})$, $b_i^d \sim \mathrm{N}(1, \sigma^2_{b^d})$, and $c_i^d \sim \mathrm{N}(1, \sigma^2_{c^d})$. The prior means of the local adjustments b_i^d and c_i^d are set to 1, which enables us to interpret β_k^d and γ_k^d as nationwide effects as is the case with additive multilevel models having population-level and group-level effects. While we use a hard equality constraint for the mean of the λ_i^ds to ensure the identifiability and more efficient estimation of the model, the hierarchical priors for b_i^d and c_i^d are sufficient for their identifiability. For the unknown deviations σ_{τ^d}, σ_{λ^d}, σ_{a^d}, σ_{b^d}, and σ_{c^d} we assign Gamma priors with shape parameter 2 and rate parameter 1. The nationwide coefficients β_k^d and γ_k^d have $\mathrm{N}(0, 2^2)$ priors. The seasonal effects s_t^d follow a standard normal prior with the aforementioned sum-to-zero constraint. The priors can be seen as weakly informative, and they are chosen primarily to enhance the computational efficiency (*Banner, Irvine & Rodhouse, 2020*).

Naturally, any of the components in the model could be excluded by setting the corresponding coefficients or standard deviations to zero. Our Bayesian model encompasses all such simplified alternatives, with the corresponding model and parameter uncertainty reflected by the estimated posterior distributions, leading to more truthful uncertainty estimates compared to merely imposing prior constraints on certain effects to be zero.

RESULTS

The model is estimated using Markov chain Monte Carlo (MCMC) with *cmdstanr* (*Gabry & Češnovar, 2022*), which is an R interface (*R Core Team, 2022*) for the probabilistic

programming language Stan for statistical inference (*Stan Development Team, 2022*). To draw the posterior samples we use NUTS sampler (*Hoffman & Gelman, 2014*; *Betancourt, 2018*) with four chains, each consisting of 7,500 iterations, the first 2,500 of which discarded as a warm-up. With parallel chains the computation takes about ten hours. The model is estimated on a supercomputer node with four cores of Xeon Gold 6230 2.1 GHz processors and 40 GB of RAM. According to the MCMC diagnostics of *cmdstanr* (*Vehtari et al., 2021*), the model converges without divergences, the $\widehat{R}$ statistics are always below 1.005, and the effective sample sizes are approximately between 700 and 43,000. The lowest effective sample size is the one of the deviation parameter of the constants considering measles, σ_{α_m}. The R and Stan codes, the data used for the analysis, and Supplementary Figures and Tables are available on GitHub (https://github.com/tihepasa/InfectionDynamics). All the figures were created using the R packages *ggplot2* (*Wickham, 2016*) and *ggpubr* (*Kassambara, 2023*).

To visualize the temporal and spatial patterns of the death occurrences and to see how the model estimates the corresponding probabilities to observe at least one death, the data and the predictions based on the model are plotted as time series and as maps in Figs. 2 and 3. The estimates are computed as $\int p(y'_{t+k}|y_1,\ldots,y_t,\theta)p(\theta|y_1,\ldots,y_T)\,d\theta$. In other words, they are k-step ahead in-sample predictions (often called fitted values in the time series literature, see, *e.g.*, *Hyndman & Athanasopoulos (2021)*), where $k-1$ is the number of preceding missing months, and they are calculated conditional on the posterior distribution of the model parameters, including the latent incidence process τ. For the first time point in this computation we also assume that the missing observations are zeros in order to have covariates for all the sites.

The temporal behavior of the data is similar to the corresponding estimates. The slight differences may be due to the fact that the proportions are based only on the data available, whereas the model predictions cover all the month-town combinations. The spatial patterns of the modeled probabilities reflect the infection distributions visible in the data. Pertussis, measles, and smallpox all have emphasis on the eastern half of Finland, with especially measles extending its prevalence to the southern parts of the country as well. When it comes to the completely missing sites, the medians of the estimated average probabilities over time are 3.7 percentage points higher for them than for the sites with observed data in the case of pertussis, 0.3 percentage points higher in the case of measles, and 0.2 percentage points lower in the case of smallpox. The differences are quite small, and by the prediction account, the model seems to work well.

In what follows, we present the results in detail. They confirm that all the components in the model are relevant, capturing different aspects of the spatio-temporal dynamics of the epidemiological phenomena.

The nationwide incidence time series of the diseases are depicted by the factors τ_t^d. The corresponding estimates are shown in Fig. 4 on a probability scale ($\text{logit}^{-1}(\tau_t^d)$). In general, they seem to have the same shapes as the observed nationwide monthly proportions of towns where at least one death caused by pertussis, measles, or smallpox was recorded (see Fig. 2). There is one major disease outbreak regarding smallpox, whereas the other diseases have several peaks, pertussis varying the most. No clear periodicity can be seen in any of the

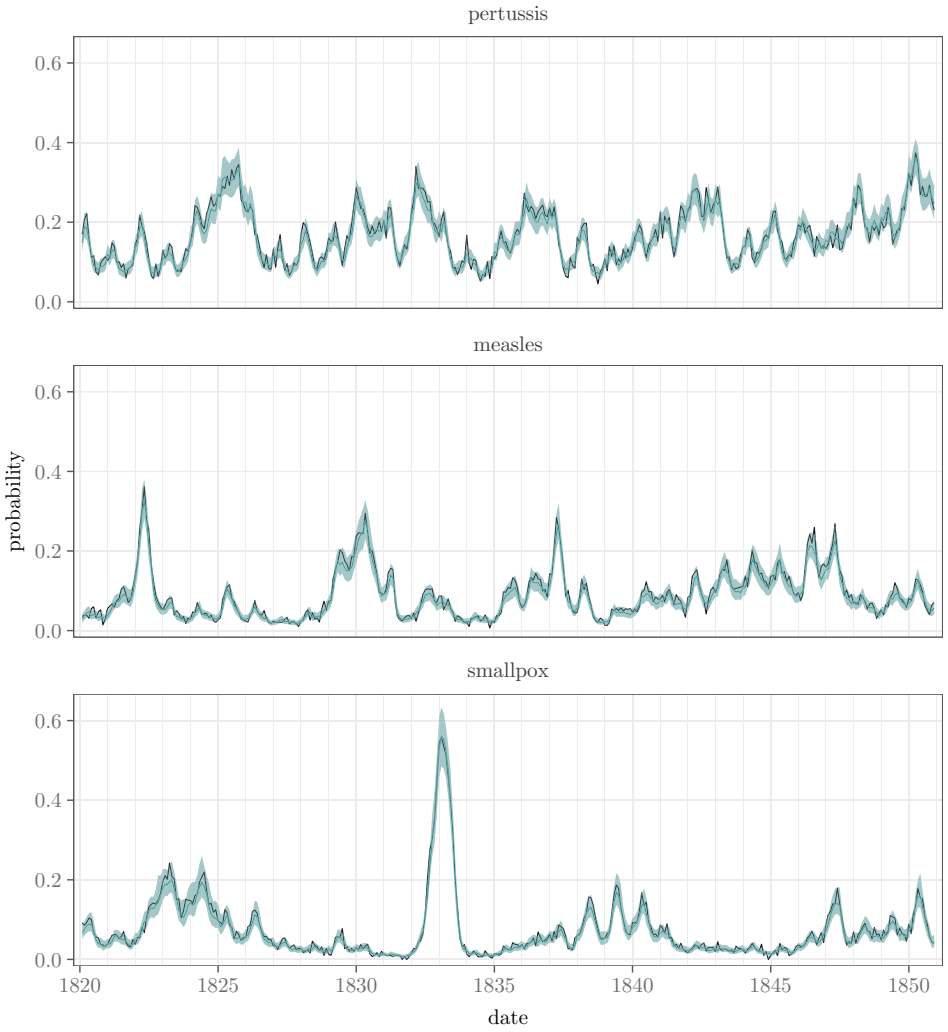

Figure 2 The prevalence of the diseases is illustrated with the dark lines depicting the proportions of the towns where at least one death was observed. The lighter turquoise lines show the posterior means of the corresponding estimates and the shaded areas their 95% posterior intervals. Note that the data line is calculated over the towns with observations, whereas the estimate line averages all the towns.

Full-size DOI: 10.7717/peerj.18155/fig-2

series, which was also confirmed by estimating dominant frequencies *via* spectral analysis using the R package *forecast* (*Hyndman & Khandakar, 2008*).

The seasonal effects s_t^d, or the average monthly deviations from the nationwide incidence level, are shown in Fig. 5. According to the estimates, the seasonal effect of pertussis peaks at the beginning of the calendar year, while the effect decreases during the summer and increases again towards the end of the year. In contrast, the only distinctive seasonal effects

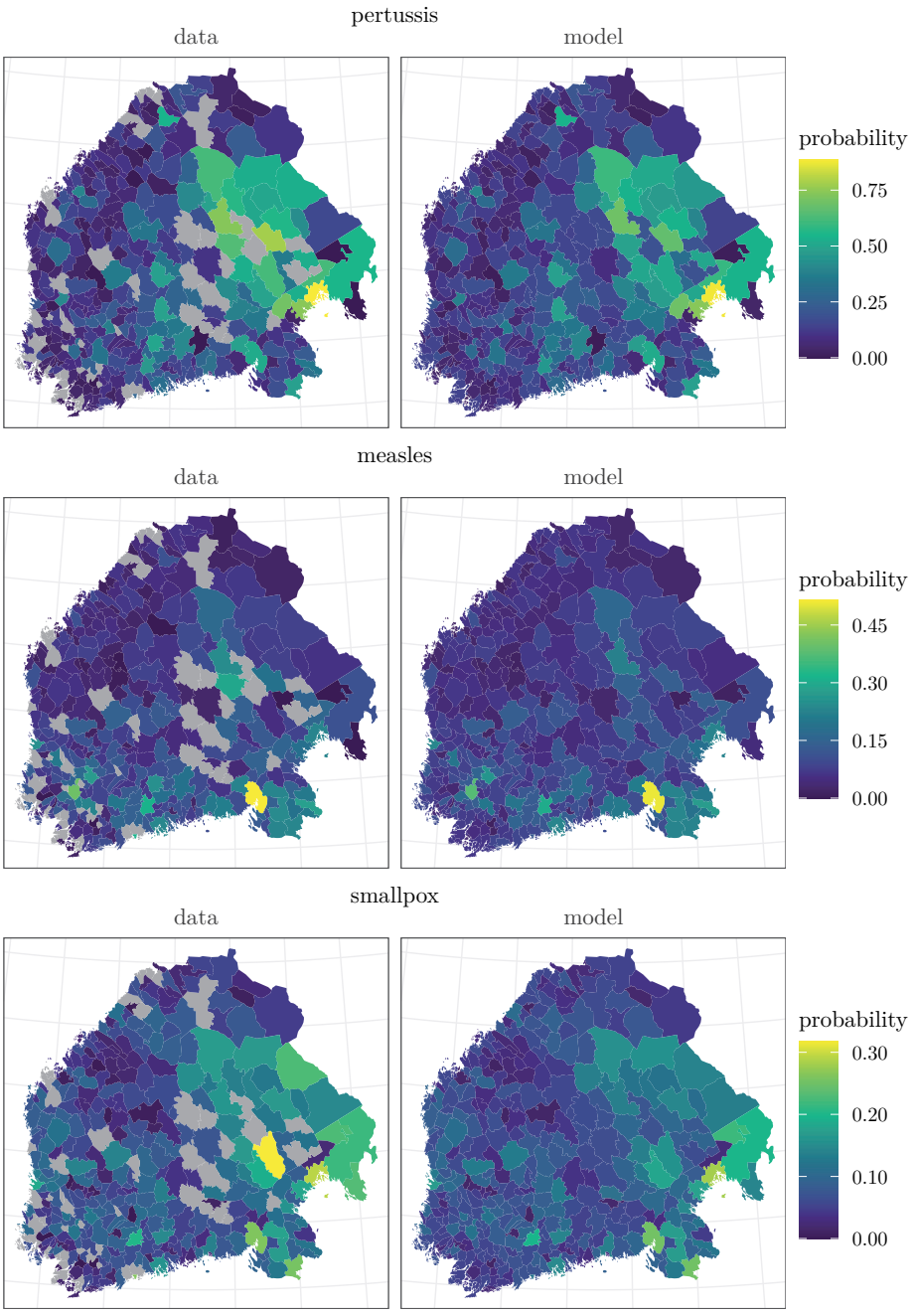

Figure 3 The left panels show the proportions of the months when deaths were recorded over the study period of 31 years. The gray areas indicate the towns where all data are missing. The right panels present the regional averages of the predicted conditional probabilities to observe at least one death caused by each disease in each month given the actual observations from the previous month. Note that the data are averaged over the observed towns and months, whereas the model covers all the towns and months.
Full-size DOI: 10.7717/peerj.18155/fig-3

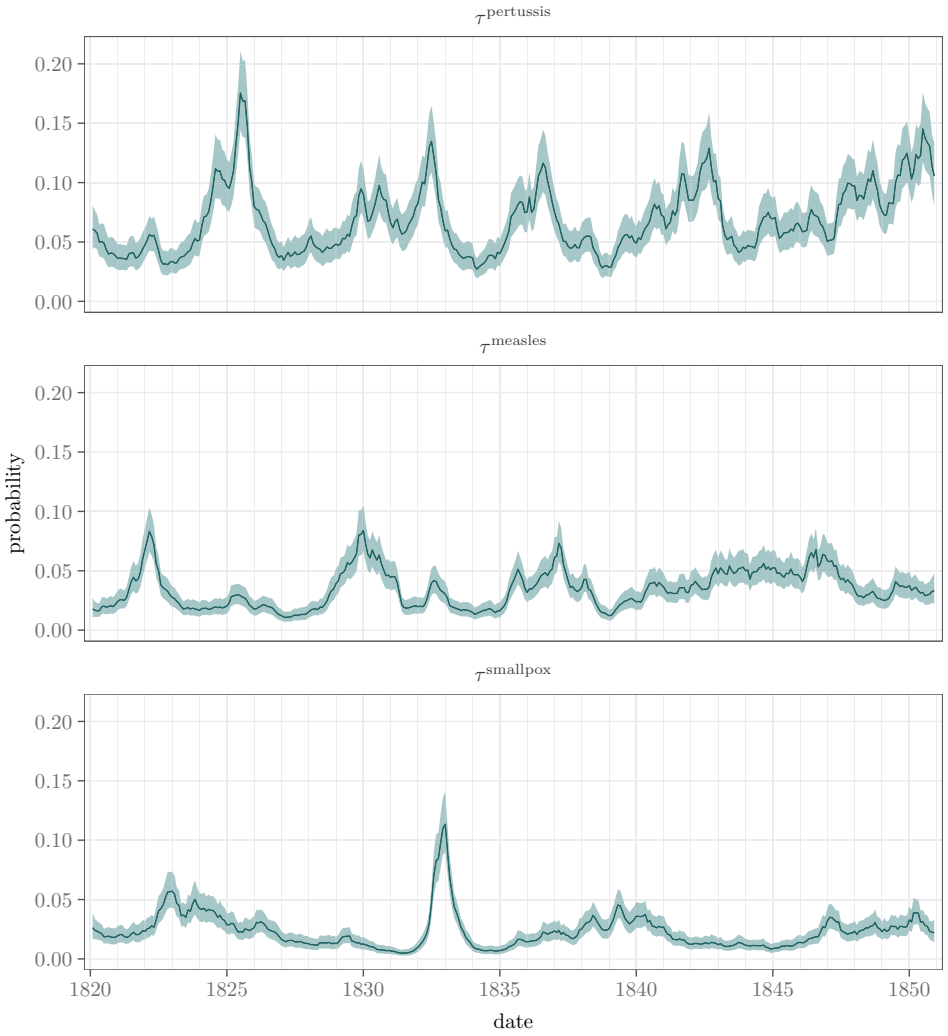

Figure 4 Posterior means and 95% posterior intervals for the unobserved incidence factors τ_t^d for pertussis, measles, and smallpox over the time period under study. The curves are on a probability scale.
Full-size DOI: 10.7717/peerj.18155/fig-4

related to measles and smallpox are the peaks in the spring and the minor decreases at the end of the year.

Measured by the τ factors, we found a distinctive correlation between the infections of measles and pertussis, 0.33 with a 95% posterior interval [0.08, 0.55]. Omitting the specific seasonal term s in the model yields almost the same correlation 0.31 [0.10, 0.50]. The correlation between smallpox and measles is ambiguous, being 0.24 [−0.01, 0.46], though it increases to 0.46 [0.26, 0.63] in the model without the seasonal component. This implies that monthly effects explain partly but not exhaustively the connection between these diseases. Smallpox and pertussis seem to be mutually independent, 0.06 [−0.19, 0.30], which is also the case with the model without the seasonal terms, 0.15 [−0.07, 0.37].

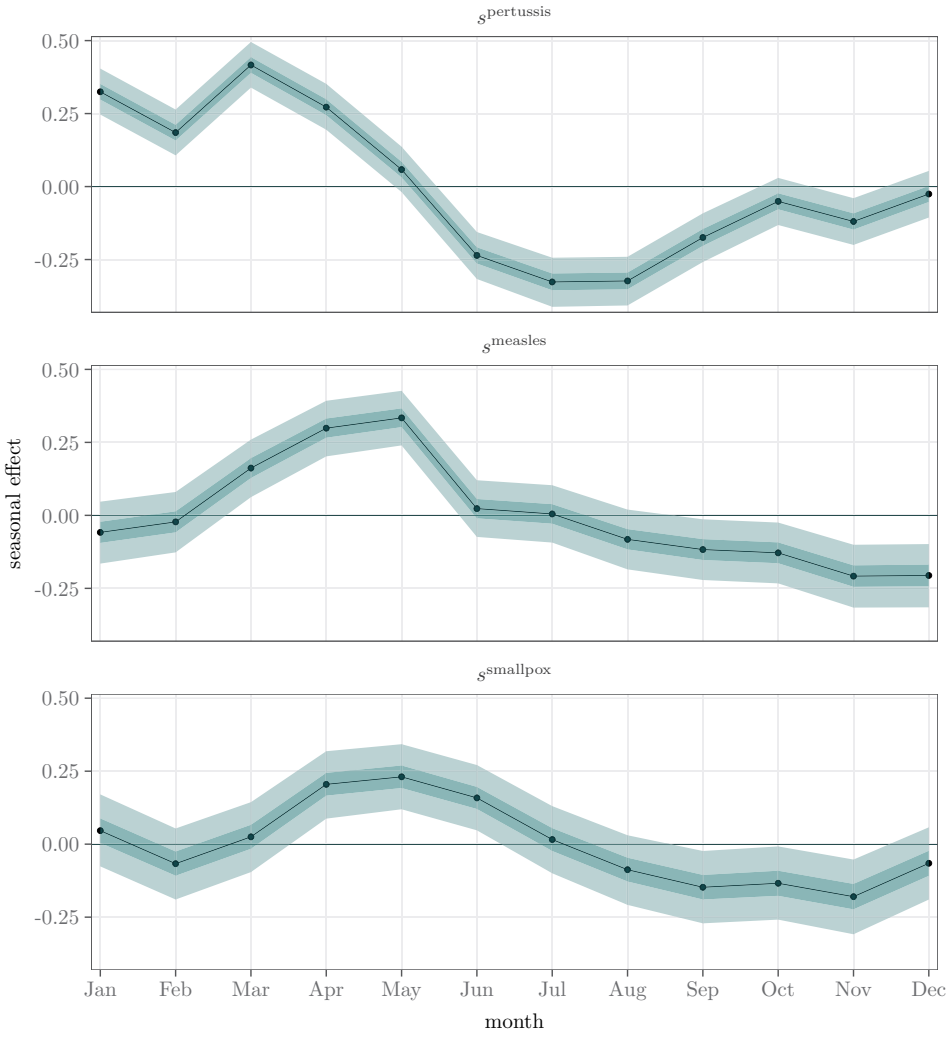

Figure 5 **Posterior means (black), 50% (dark turquoise) and 95% (light turquoise) posterior intervals of monthly seasonal effects s_t^d for pertussis, measles, and smallpox over a year.**
Full-size DOI: 10.7717/peerj.18155/fig-5

According to the regional loadings λ_i^d, adjusting the nationwide factors τ_t^d, it was more likely to die of any of these diseases in eastern and southeastern Finland than in other parts of the study area. This is also in accordance with the maps of the data in Fig. 3. The posterior means of the loadings λ are plotted in Fig. 6. Considering the loadings, there is most local variation in pertussis, $\sigma_{\lambda^p} = 0.35$ with a 95% posterior interval $[0.31, 0.40]$. With regard to measles and smallpox, the loadings vary less, $\sigma_{\lambda^m} = 0.17$ $[0.14, 0.20]$ and $\sigma_{\lambda^s} = 0.15$ $[0.13, 0.17]$.

The final term affecting the base level of the probability to observe at least one death caused by pertussis, measles, or smallpox consists of the regional constants a_i, shown in Fig.

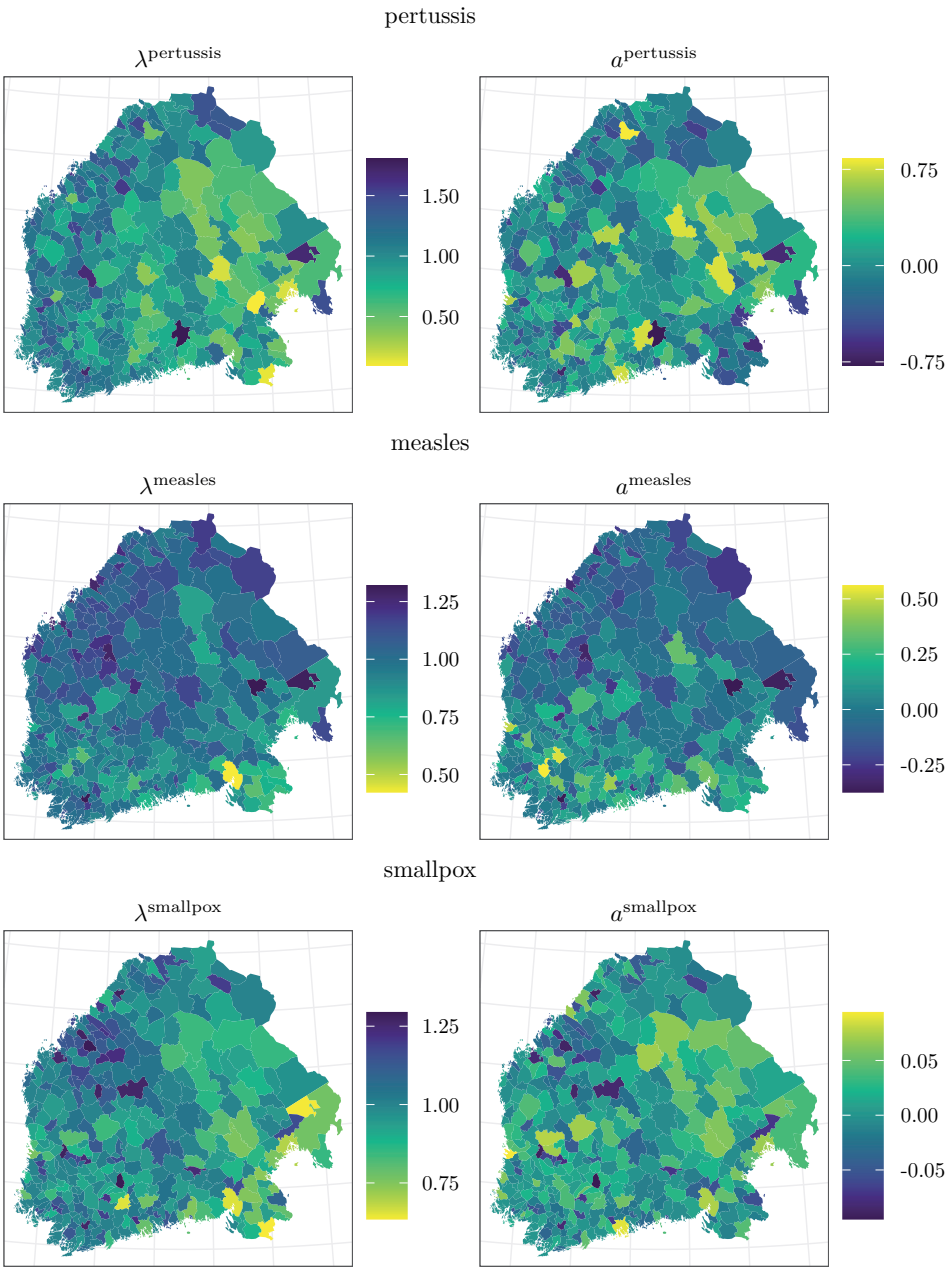

Figure 6 The left panels show the posterior means of the local loadings λ_i, adjusting the national incidence factors τ_t. Since the factors are negative, the smaller the loading is, the greater the probability of at least one death is. The right panels illustrate the posterior means of the regional constants a_i.

Full-size DOI: 10.7717/peerj.18155/fig-6

6. Those related to pertussis and smallpox seem to be larger in eastern and southwestern inland areas, whereas those considering measles are the largest in southern Finland.

The estimates of the nationwide regression coefficients β and γ are represented in Table 1. All effects differing from zero are positive, meaning they increase the probability to detect at least one death. The probability to observe one or more deaths induced by one of these diseases is increased most prominently if there are recorded deaths caused by the same disease in the same town, or in its neighbors, in the previous month. However, there is more uncertainty in the effects of neighbors than in those of the towns themselves. The risk that there is at least one death caused by pertussis is increased by the occurrence of measles, whereas the corresponding effect of smallpox is not distinctive. Measles is probably affected more by smallpox than by pertussis. In turn, measles seems to affect smallpox more than pertussis does.

When it comes to the local adjustments b_i and c_i, their standard deviations are clearly above zero, varying between 0.22 (σ_{b^m}) and 0.55 (σ_{c^p}), which indicates that the local adjustments differ geographically. There are no obvious interpretations of their spatial patterns (see maps of b and c in Supplementary Figures 1 and 2 on GitHub). This is credible since the coefficients represent the combined effects of multiple unobserved features that are not necessarily spatially organized.

For full results of all time and town invariant parameter estimates with their prior and posterior intervals, see Supplementary Table 1 on GitHub.

Model comparison

While our main interest was studying the past spatio-temporal dynamics of infections and disease associations within and between the diseases, we also examined the necessity and reasonableness of modeling the disease interdependencies and the response aggregation. We compared our model with a corresponding one without the dependencies between the infections by excluding the other diseases as explanatory variables and omitting the correlation between the incidence factors τ_t in the simpler model. Since the original data contained the numbers of deaths instead of the dichotomous aggregates we used as a response, we also estimated corresponding models with the difference of using the counts as a response and a negative binomial distribution to model them. Additionally, the briefly aforementioned model without a seasonal component was included in the comparison in the case of both types of responses. This resulted in six different model versions for comparison: dependent diseases, independent diseases, and dependent diseases without a seasonal effect, each for both Bernoulli and negative binomial distributions.

The negative binomial model can be formally written as

$$y_{i,t}^d \sim \mathrm{NB}\left(\exp\left(\eta_{i,t}^d\right), \exp\left(\alpha_{\phi^d} + \phi_i^d\right)\right), \tag{3}$$

where the mean parameter $\eta_{i,t}^d$ is defined as in Eq. (2), and the nationwide dispersion parameters α_{ϕ^d} and the local dispersion parameters ϕ_i^d depend on the response disease. The priors are the same as with the Bernoulli model, with the addition of $\alpha_{\phi^d} \sim \mathrm{N}(0, 1^2)$, $\phi^d \sim \mathrm{N}(0, \sigma_{\phi^d}^2)$, and $\sigma_{\phi^d} \sim \mathrm{Gamma}(2, 1)$. From the negative binomial model, we could then compute our quantity of interest, the probability of observing at least one death in a

Table 1 Posterior means and 95% posterior intervals of the nationwide regression parameters grouped by the response disease. The superscript indicates the response disease and the subscript the explanatory disease.

	Within towns			Between towns		
		Mean	(2.5, 97.5%)		Mean	(2.5, 97.5%)
pertussis → pertussis	β_p^p	1.56	(1.47, 1.64)	γ_p^p	1.23	(1.10, 1.35)
measles → measles	β_m^m	1.90	(1.82, 1.98)	γ_m^m	2.21	(2.06, 2.38)
smallpox → smallpox	β_s^s	2.43	(2.34, 2.53)	γ_s^s	2.57	(2.40, 2.74)
measles → pertussis	β_m^p	0.11	(0.04, 0.18)	γ_m^p	0.12	(0.00, 0.24)
smallpox → pertussis	β_s^p	0.04	(−0.04, 0.12)	γ_s^p	0.11	(−0.02, 0.23)
pertussis → measles	β_p^m	0.13	(0.06, 0.20)	γ_p^m	0.11	(−0.01, 0.23)
smallpox → measles	β_s^m	0.15	(0.05, 0.24)	γ_s^m	0.24	(0.08, 0.39)
pertussis → smallpox	β_p^s	0.05	(−0.03, 0.13)	γ_p^s	0.19	(0.05, 0.33)
measles → smallpox	β_m^s	0.21	(0.12, 0.31)	γ_m^s	0.38	(0.21, 0.54)

specific town and month, which could be compared with the corresponding estimates of the Bernoulli model.

As a scoring rule for the model comparison, we used the expected log predictive density (ELPD), which measures the goodness of the entire predictive distribution (*Vehtari, Gelman & Gabry, 2017*). The ELPD was estimated *via* an approximate leave-one-out cross-validation using the R package *loo* (*Vehtari et al., 2023*). We left out one month and town from all the diseases at a time to estimate the ELPD. Models with higher values of ELPD correspond to greater posterior predictive accuracy for predicting new data points compared to models with lower ELPD values.

According to the differences in the ELPDs in Table 2, the best performing model is the one utilizing Bernoulli distribution and considering the diseases dependent. Omitting the dependencies results in the second-best model, with the difference in ELPD over three times the standard error. As could be expected, omitting the seasonal effect further impairs the model. When it comes to the negative binomial models with counts as responses, the order of the dependent, independent, and seasonless models is the same. The Bernoulli models outperform the negative binomial ones in all cases. Overall, we see that directly using the dichotomized data *versus* modeling the count data has a greater impact than considering the infection dependencies or seasonality in our model. However, even though in terms of predictive performance the differences between different Bernoulli models are relatively small, we used the most complex model in our main analysis. This is in line with the common Bayesian paradigm of incorporating the uncertainty of the model structure in the model (*Vehtari & Ojanen, 2012*).

For the dependent and independent models, we performed additional prediction checks by discarding the last two years of the data and estimating the probabilities for those years. We also calculated the ELPDs considering the removed years, see Table 2. The modifications without the seasonal effect were not included in this comparison due to their already evident poor performance and the fact that they were originally fitted merely to investigate the importance of the obvious seasonal variation. The posterior means were

Table 2 Differences of the ELPDs and the standard errors of the ELPD differences for the leave-one-out cross-validation. The values in the first two columns are computed over all the years for the models estimated with the full data, and the last two columns are the values calculated over the last two years for the models discarding those years while estimating the models.

	Full data		Last two years	
	$\text{ELPD}_{\text{diff}}$	SE_{diff}	$\text{ELPD}_{\text{diff}}$	SE_{diff}
Bernoulli dependent	0.00	0.00	0.00	0.00
Bernoulli independent	−36.71	11.75	−46.84	4.21
Bernoulli dependent, without season	−80.50	13.08	–	–
Negative binomial dependent	−353.88	43.75	−381.38	16.70
Negative binomial independent	−400.59	45.01	−490.50	18.90
Negative binomial dependent, without season	−437.72	45.98	–	–

quite similar in all cases, but the posterior intervals were wider for measles and smallpox in the case of the negative binomial model, as can be seen from Fig. 7.

Overall, the results of the Bernoulli, as well as the negative binomial, models seem to indicate similar interdependencies between the diseases. In the case of the negative binomial model, the estimates of all the time and town invariant parameters with their prior and posterior intervals are shown in Supplementary Table 2 on GitHub. Also, figures corresponding to the ones representing the results of the Bernoulli model are available in GitHub (Supplementary Figures 3-9).

Furthermore, to inspect the effect of the earlier disease history, a Bernoulli model according to Eqs. (1) and (2) was fitted with additional lags of two months for the three focal and three neighbor covariates. The approach resulted in having six scalar regression coefficients (three βs and three γs) more for each response disease than in our main model. The results are well aligned with the one-month lag model. The amount of available observations decreases by about three percentage points when introducing the two-month lag since we must know not only the previous observation but also the one preceding that. Thus this model is not completely comparable to our main model. Using the same data for both one-month and two-month lag models, the model with two-month lag performs better, measured with the ELPD: one-month lag model results in $\text{ELPD}_{\text{diff}} = -989.87$ and $\text{SE}_{\text{diff}} = 53.89$, compared with the two-month lag model. Nevertheless, there are some convergence and efficacy issues. Out of the 20,000 iterations, there are 28 diverging ones which can potentially bias the results, thus they are not completely reliable (*Betancourt, 2018*). The outperformance of the two-month lag model in the sense of the ELPD might be also related to smaller variance regardless of the possible bias, which is consistent with the seemingly better fitting predictions gained from our main model than the two-month lag model (see Supplementary Figure 10). The full results are in GitHub in Supplementary Table 3 and Supplementary Figures 10-16.

Additionally, we fitted a model using our data aggregated on a weekly level, which increases the amount of missing data from 24% in the monthly data to 48% in the weekly data. We used lags from 1 to 4 weeks to cover as much delayed effect as with our main model. Due to the unequal number of weeks per calendar month, incorporating the monthly effect

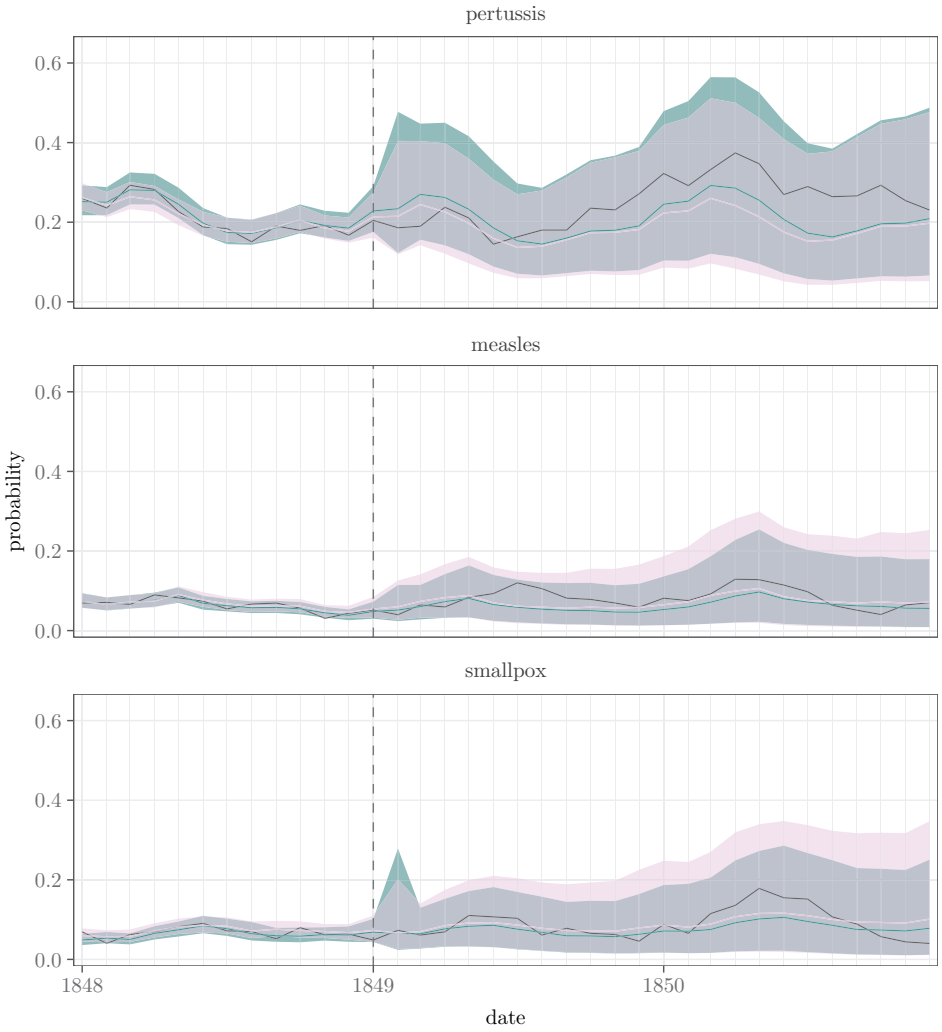

Figure 7 The dark gray lines depict the proportions of the towns where at least one death was observed. The turquoise lines show the posterior means of the corresponding estimates and the shaded areas their 95% posterior intervals in the case of the Bernoulli model, whereas the pink lines and areas represent the same values for the negative binomial model. The dotted vertical line indicates the time point after which the model estimates are predicted by the models estimated without the data of the last two years.
Full-size DOI: 10.7717/peerj.18155/fig-7

is not straightforward, so we omitted the seasonal effect. Unfortunately, this model did not converge, potentially because of the increased amount of missing data, or the complex dependency structures due to varying sub-monthly incubation and time-to-fatality times. The weekly data and model code are available in the supplementary material.

DISCUSSION

We developed a Bayesian model to explore the spatio-temporal dynamics and co-dynamics of three fatal childhood infections—measles, smallpox, and pertussis—in pre-healthcare Finland (1820–1850). The main novelties of the approach are, firstly, the consideration of both the spatial and temporal aspects simultaneously, and, secondly, considering the connections not only within but also between the three diseases. Furthermore, our dataset is substantially different in comparison to the corresponding previous epidemiological literature. Instead of data regarding large cities or being pooled over countries, we exploited records from a sparsely populated nation, where 1.2–1.6 million inhabitants were spread over vast areas in hundreds of small towns without modern healthcare. Our model allows the inclusion of several explanatory elements which all capture different features. According to our results, all the components are meaningful and statistically distinctive, and the incorporation of the possibility of dependencies between the diseases leads to a model describing the data better than one merely assuming independent diseases. The data and the model framework are available on GitHub, providing a template for other researchers.

Based on our results, the main components explaining the temporal and geographical variation in the probabilities of observing at least one death caused by pertussis, measles, or smallpox are the nationwide incidence factors with their local adjustments. The estimated incidence factors follow the temporal behavior of the observed data, and the regional adjustments resemble the spatial patterns of the data (Figs. 2 and 4, and 3 and 6).

Measured by pairwise correlations of the incidence factors, a distinctive positive co-occurrence of measles and pertussis was discovered. Previous research has found positive, negative, and inconsistent co-occurrences of these infections, see, *e.g.*, *Rohani et al. (2003)*, *Coleman (2015)*, and *Noori & Rohani (2019)*. We also found a notable connection between measles and smallpox with a model without the seasonal component, but this correlation is not present in the full model including the seasonality. This indicates that their dynamics follow a similar, seasonal pace. Overall, the seasonal effect is visible among all the diseases. In addition to the nationwide incidence level, the seasonality increases the mortality during the first half of the year, depending on the disease, see Fig. 5. The seasonalities may reflect increased transmission during social gatherings, or they can be due to some environmental and climatic drivers (*Metcalf et al., 2009*; *Metcalf et al., 2017*). The work of *Briga et al. (2021)*, based on selected data covering longer periods, indicates that of the infections of pertussis, measles, and smallpox only pertussis was linked with new year and Easter in Finland in the 18th and 19th centuries.

Furthermore, lagged dependencies within and between the infections were discovered as positive temporal and spatial effects of the explanatory variables. Recorded deaths in the focal town and in its adjacent towns in the previous month increased the risk of dying of the same disease. Between the infections, these effects were notably smaller (Table 1). It should be noticed that the coefficients reflecting the effect of the history of the focal town and its neighbors are not directly comparable, as the value of the focal covariate is either 0 or 1, but the neighborhood covariate is a proportion between 0 and 1.

According to the results, the risk of succumbing to pertussis, measles, and smallpox was mediated by occurrences of the other infections in the area. All these three diseases tended to increase the mortality related to the two other diseases, as all the pairwise interaction parameter estimates are positive. This might be due to general immunosuppression or to decreased condition following the previous infection. The strongest associations were found between measles and pertussis, and measles and smallpox. The possibility that pertussis is driven by the immunosuppressive effects of measles as suggested by *Coleman (2015)* and *Noori & Rohani (2019)* implies that the risk of dying of pertussis is increased by a recent measles infection. This is also supported by findings of *Mina et al. (2015)* showing that measles vaccination, by preventing measles-associated immune memory loss, decreases the risk of other infections. Our observations (see Table 1) are aligned with these results. However, also a reverse connection was recovered: the recorded deaths caused by pertussis in the same town during the previous month increased the risk of observing one or more measles induced deaths almost equally. A stronger lagged effect was discovered between measles and smallpox. Also these interactions were found to act in both directions.

To gain further insights into the specific effects of immunosuppression and impaired health conditions, longer than the one-month (or two-month) lags that we used here, should likely be employed. Unfortunately, our data do not suffice for identifying such effects as accounting longer histories or using finer timescale is challenging due to the missing data and the relative rarity of the deaths. Also herd immunity would be an important aspect to consider, but until proper population size estimates are available, it remains a topic for future work. Concerning our results, the lack of controlling for the longer term immunity might obscure some of the findings when compared to more contemporary datasets. Although immunosuppressive mechanisms of measles are well known in the literature, for many other diseases those are less known, for example, the effect of pertussis on measles (however, see, *e.g.*, *Macina & Evans (2021)*). Thus, we suggest carefulness in interpreting our results as they might reflect the shorter term effects caused by the overall condition of the patients rather than true immunosuppression.

The observed spatially varying local risks of at least one death due to pertussis, measles, or smallpox may arise from the closeness of potential sources of infection, differences in cultural, housing, or nutritional circumstances, or even genetics (*Honkola et al., 2018*; *Voutilainen, 2017*; *Kerminen et al., 2017*). As can be seen from Fig. 3, the probabilities of detecting one or more deaths caused by pertussis and smallpox were greater in the eastern parts of Finland, whereas measles was clearly an infection emphasized in the southern parts, being in concordance what was suggested by *Pitkänen, Mielke & Jorde (1989)* and *Ketola et al. (2021)*.

When it comes to the long term temporal behavior of the infections, it seems that epidemics in small populations, consisting of sparse metapopulations of tiny towns, might be dominated by reintroductions and fade-outs rather than by endemic dynamics more typical in densely populated cities and countries (*Keeling & Grenfell, 1997*; *Grenfell & Harwood, 1997*; *Rohani et al., 2003*; *Ketola et al., 2021*). In *Briga, Ketola & Lummaa (2022)* epidemics were found to reoccur in cycles of roughly four years in the 18th and 19th centuries in chosen Finnish towns with the highest quality data. The length and phase

of such patterns are likely to vary due to annual and geographical differences in seasons, making them challenging to estimate from our scarce data. Our study covering 31 years did not reveal any long-term nationwide periodicities.

We modeled the deaths caused by measles, smallpox, and pertussis *via* a binary Bernoulli distribution, where value 1 denotes that there was at least one reported death given the disease, town, and month, and 0 for no reported deaths. This approach, while sacrificing some detail, allowed us to capture the broad trends and patterns in the data, and to make meaningful inferences about the spatio-temporal co-dynamics of these diseases. In contrast to the generally held view that dichotomizing data should be avoided, in our case directly modeling binary presence-absence data seemed to be beneficial compared to modeling observed death counts, potentially due to accuracy issues in the actual counts. However, both approaches led to practically identical main conclusions. The model comparisons also exemplified how our approach is applicable to other kinds of responses than Bernoulli variables.

We accounted for spatial dependencies using explanatory variables based on a neighborhood structure defined by a shared border between two towns. To model and quantify the evident epidemiological transmission dynamics, we included neighbor effects enabling the situation in the adjacent towns in the previous month to affect the probability to observe one or more deaths in the focal town. Our choice of neighborhood is straightforward, omitting the actual intensity of communication between the neighboring towns, hence possibly shrinking or magnifying the true dynamics of the infections. If there were more detailed data or complementary information about the social connections, other definitions for neighborhood, even with an appropriate weighing mechanism, could be employed. We tried to consider each pair of neighbors individually, but the information in the data was not sufficient for model identifiability, owing to the rarity of cases in neighboring towns. Naturally, including alternative appropriate and available covariates as explanatory variables is possible as well. The general spatio-temporal model developed for the purpose of exploring the dynamics and co-dynamics particularly in the case of sparse and scarce data is applicable to other corresponding datasets, for example, based on the historical parish records from other Nordic countries, or data on modern day rural areas.

ACKNOWLEDGEMENTS

The authors wish to acknowledge CSC –IT Center for Science, Finland, for computational resources. The authors also thank Virpi Lummaa.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Tiia-Maria Pasanen was supported by the Finnish Cultural Foundation, the Emil Aaltonen Foundation and the Research Council of Finland grant 331817. Jouni Helske was supported by the Research Council of Finland grants 331817 and 355153. Tarmo Ketola was supported by the Research Council of Finland grant 278751. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
The Finnish Cultural Foundation, the Emil Aaltonen Foundation and the Research Council of Finland: 331817.
The Research Council of Finland: 331817, 355153.
The Research Council of Finland: 278751.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Tiia-Maria Pasanen conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Jouni Helske conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Harri Högmander conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Tarmo Ketola conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data, code files for analyses, and additional figures and tables are available at GitHub and Zenodo:

- https://github.com/tihepasa/infectionDynamics/releases/tag/v2.0.2

- tihepasa. (2024). tihepasa/infectionDynamics: infectionDynamics supplements (v2.0.1). Zenodo. https://doi.org/10.5281/zenodo.11183838
10.5281/zenodo.10591760.

REFERENCES

Bai J, Wang P. 2015. Identification and Bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics* **33(2)**:221–240 DOI 10.1080/07350015.2014.941467.

Bakaletz LO. 2017. Viral—bacterial co-infections in the respiratory tract. *Current Opinion in Microbiology* **35**:30–35 DOI 10.1016/j.mib.2016.11.003.

Ball F, Britton T, House T, Isham V, Mollison D, Pellis L, Tomba GS. 2015. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics* **10**:63–67 DOI 10.1016/j.epidem.2014.08.001.

Banner KM, Irvine KM, Rodhouse TJ. 2020. The use of Bayesian priors in ecology: the good, the bad and the not great. *Methods in Ecology and Evolution* **11(8)**:882–889 DOI 10.1111/2041-210X.13407.

Betancourt M. 2018. A conceptual introduction to hamiltonian Monte Carlo. ArXiv arXiv:1701.02434.

Briga M, Ketola T, Lummaa V. 2022. The epidemic dynamics of three childhood infections and the impact of first vaccination in 18th and 19th century Finland. *MedRxiv* DOI 10.1101/2022.10.30.22281707.

Briga M, Ukonaho S, Pettay JE, Taylor RJ, Ketola T, Lummaa V. 2021. The seasonality of three childhood infections in a pre-industrial society without schools. DOI 10.1101/2021.10.08.21264734.

Coleman S. 2015. The historical association between measles and pertussis: a case of immune suppression? *SAGE Open Medicine* **3**:1–6 DOI 10.1177/2050312115621315.

Gabry J, Češnovar R. 2022. cmdstanr: R Interface to 'CmdStan'. *Available at https://mc-stan.org/cmdstanr/*.

Graham AL. 2008. Ecological rules governing helminth—microparasite coinfection. *Proceedings of the National Academy of Sciences of the United States of America* **105(2)**:566–570.

Grenfell B, Harwood J. 1997. (Meta)population dynamics of infectious diseases. *Trends in Ecology & Evolution* **12(10)**:395–399 DOI 10.1016/S0169-5347(97)01174-9.

Gupta S, Ferguson N, Anderson R. 1998. Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* **280(5365)**:912–915 DOI 10.1126/science.280.5365.912.

Hoffman MD, Gelman A. 2014. The No-U-Turn sampler: adaptively setting path lengths in hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15(1)**:1593–1623.

Honkola T, Ruokolainen K, Syrjänen KJJ, Leino U-P, Tammi I, Wahlberg N, Vesakoski O. 2018. Evolution within a language: environmental differences contribute to divergence of dialect groups. *BMC Evolutionary Biology* **18**:1 DOI 10.1186/s12862-017-1072-2.

Hyndman RJ, Athanasopoulos G. 2021. *Forecasting: principles and practice*. 3rd edition. Melbourne: OTexts.

Hyndman RJ, Khandakar Y. 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* **26(3)**:1–22 DOI 10.18637/jss.v027.i03.

Kassambara A. 2023. ggpubr: 'ggplot2' based publication ready plots. R package version 0.6.0. *Available at https://CRAN.R-project.org/package=ggpubr*.

Keeling MJ, Grenfell BT. 1997. Disease extinction and community size: modeling the persistence of measles. *Science* **275(5296)**:65–67 DOI 10.1126/science.275.5296.65.

Kerminen S, Havulinna AS, Hellenthal G, Martin AR, Sarin A-P, Perola M, Palotie A, Salomaa V, Daly MJ, Ripatti S, Pirinen M. 2017. Fine-scale genetic structure in Finland. *G3: Genes, Genomes, Genetics* **7(10)**:3459–3468 DOI 10.1534/g3.117.300217.

Ketola T, Briga M, Honkola T, Lummaa V. 2021. Town population size and structuring into villages and households drive infectious disease risks in pre-healthcare Finland. *Proceedings of the Royal Society B: Biological Sciences* **288(1949)**:20210356 DOI 10.1098/rspb.2021.0356.

Lewandowski D, Kurowicka D, Joe H. 2009. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* **100(9)**:1989–2001 DOI 10.1016/j.jmva.2009.04.008.

Macina D, Evans KE. 2021. Pertussis in individuals with co-morbidities: a systematic review. *Infectious Diseases and Therapy* **10(3)**:1141–1170 DOI 10.1007/s40121-021-00465-z.

Metcalf CJE, Bjørnstad ON, Grenfell BT, Andreasen V. 2009. Seasonality and comparative dynamics of six childhood infections in pre-vaccination Copenhagen. *Proceedings of the Royal Society B: Biological Sciences* **276(1676)**:4111–4118 DOI 10.1098/rspb.2009.1058.

Metcalf CJE, Walter KS, Wesolowski A, Buckee CO, Shevliakova E, Tatem AJ, Boos WR, Weinberger DM, Pitzer VE. 2017. Identifying climate drivers of infectious disease dynamics: recent advances and challenges ahead. *Proceedings of the Royal Society B: Biological Sciences* **284(1860)**:20170901 DOI 10.1098/rspb.2017.0901.

Mina MJ, Metcalf CJE, De Swart RL, Osterhaus A, Grenfell BT. 2015. Long-term measles-induced immunomodulation increases overall childhood infectious disease mortality. *Science* **348(6235)**:694–699 DOI 10.1126/science.aaa3662.

Mueller JT, McConnell K, Burow PB, Pofahl K, Merdjanoff AA, Farrell J. 2020. Impacts of the COVID-19 pandemic on rural America. *Proceedings of the National Academy of Sciences of the United States of America* **118(1)**:2019378118 DOI 10.1073/pnas.2019378118.

Nickbakhsh S, Mair C, Matthews L, Reeve R, Johnson PC, Thorburn F, Von Wissmann B, Reynolds A, McMenamin J, Gunson RN, Murcia PR. 2019. Virus—virus interactions impact the population dynamics of influenza and the common cold. *Proceedings of the National Academy of Sciences of the United States of America* **116(52)**:27142–27150.

Noori N, Rohani P. 2019. Quantifying the consequences of measles-induced immune modulation for whooping cough epidemiology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374(1775)**:20180270 DOI 10.1098/rstb.2018.0270.

Pitkänen K. 1977. The reliability of the registration of births and deaths in Finland in the eighteenth and nineteenth centuries: some examples. *Scandinavian Economic History Review* **25(2)**:138–159 DOI 10.1080/03585522.1977.10407878.

Pitkänen KJ, Mielke JH, Jorde LB. 1989. Smallpox and its eradication in Finland: implications for disease control. *Population Studies* **43(1)**:95–111 DOI 10.1080/0032472031000143866.

R Core Team. 2022. R: a language and environment for statistical computing. Vienna, Austria: *Available at https://www.R-project.org/*.

Rohani P, Green C, Mantilla-Beniers N, Grenfell B. 2003. Ecological interference between fatal diseases. *Nature* **422(6934)**:885–888 DOI 10.1038/nature01542.

Saarivirta T, Consoli D, Dhondt P. 2012. The evolution of the Finnish health-care system early 19th century and onwards. *International Journal of Business and Social Sciences* **3(6)**:243–257.

Shrestha S, Foxman B, Weinberger DM, Steiner C, Viboud C, Rohani P. 2013. Identifying the interaction between influenza and pneumococcal Pneumonia using incidence data. *Science Translational Medicine* **5**:191ra84.

Stan Development Team. 2022. The Stan C++ library. Version 2.30. *Available at https://mc-stan.org/*.

Statistical Office of Finland. 1882. Statistical yearbook of Finland 1882. Helsinki, Finland: Statistical Office of Finland *Available at https://urn.fi/URN:NBN:fi-fe2023013124832*.

Van Buuren S. 2018. *Flexible imputation of missing data*. Boca Raton: CRC Press DOI 10.1201/9780429492259.

Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner P-C, Paananen T, Gelman A. 2023. loo: efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.6.0. *Available at https://mc-stan.org/loo/*.

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 DOI 10.1007/s11222-016-9696-4.

Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. 2021. Rank-normalization, folding, and localization: an improved $\widehat{R}$ for assessing convergence of MCMC (with Discussion). *Bayesian Analysis* **16(2)**:667–718 DOI 10.1214/20-BA1221.

Vehtari A, Ojanen J. 2012. A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys* **6(none)**:142–228 DOI 10.1214/12-SS102.

Vogels CB, Rückert C, Cavany SM, Perkins TA, Ebel GD, Grubaugh ND. 2019. Arbovirus coinfection and co-transmission: a neglected public health concern? *PLOS Biology* **17(1)**:e3000130 DOI 10.1371/journal.pbio.3000130.

Voutilainen M. 2017. Marriage and household structure in rural pre-famine Finland, 1845–65. *COLLeGIUM: Studies across Disciplines in the Humanities and Social Sciences* **22**:67–82.

Voutilainen M, Helske J, Högmander H. 2020. A Bayesian reconstruction of a historical population in Finland, 1647–1850. *Demography* **57(3)**:1171–1192 DOI 10.1007/s13524-020-00889-1.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag New York. *Available at https://ggplot2.tidyverse.org*.

Wong A, Barrero Guevara LA, Goult E, Briga M, Kramer SC, Kovacevic A, Opatowski L, Domenech de Cellès M. 2023. The interactions of SARS-CoV-2 with cocirculating pathogens: epidemiological implications and current knowledge gaps. *PLOS Pathogens* **19(3)**:e1011167 DOI 10.1371/journal.ppat.1011167.

III

HIDDEN MARKOV MODELLING OF SPATIO-TEMPORAL DYNAMICS OF MEASLES IN 1750–1850 FINLAND

by

Pasanen, T.-M., Helske, J., and Ketola, T. 2024

Preprint

DOI: https://doi.org/10.48550/arXiv.2405.16885

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

Tiia-Maria Pasanen[1], Jouni Helske[1,2], and Tarmo Ketola[3,4]

[1]Department of Mathematics and Statistics, University of Jyväskylä
[2]INVEST Research Flagship Centre, University of Turku
[3]Department of Forestry, University of Helsinki
[4]Department of Biological and Environmental Science, University of Jyväskylä

December 2, 2024

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 into 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 routes after major administrative transformations in Finland.

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. 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 where spatial and temporal effects are not separable [23]. Approximate methods like INLA [35] may be 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 [27, 36]. Furthermore, 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.

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 [18, 17]. 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.

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. Employing more states may aid in discovering latent phenomena related to the epidemics. 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 time-invariant.

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 illustrate our approach by modelling the spatio-temporal occurrence of measles in Finland during the 18$^{\text{th}}$ and 19$^{\text{th}}$ centuries, offering important insights into the historical measles epidemics in Finland. Overall, such descriptive analysis in epidemiology may be of great value [14].

Extremely low population size coupled with hundreds of sparsely spaced small towns makes it challenging for epidemics to remain fully endemic [21]. Thus, it is probable that in such conditions of Finland, fadeouts and reintroductions dominated the infection dynamics, implying that changes in the transmission sources could have strongly affected these dynamics. Finland and Sweden are separated by a large geographical barrier, the Gulf of Bothnia, and this, as well as the Russian border on the east, must have influenced the trade connections and the movement of people. The annexation of Finland to Russia, after a war in 1808–1809, resulted in the transfer of the capital from the western position of Turku to a more eastern location of Helsinki, increased dramatically trade and social communication, and altered drastically Finland's connectivity to other countries [28]. We exploit our method to describe the reoccurring and presumably significantly foreign connection driven patterns of the epidemics in a compact yet informative way.[1]

2 Data and methods

2.1 Data

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. Also earlier, albeit incomplete, records exist [30]. Even though death diagnostics were symptom-based, the distinctive features of certain diseases allowed for relatively accurate diagnoses. One of those infections is measles, which we focus on. According to our

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

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 [40, 43]. However, the healthcare system was primitive and virtually non-existent before the 19$^{\text{th}}$ century, lagging behind in development compared to the other Nordic countries and Europe until the 20$^{\text{th}}$ century [37, 38].

Communication within Finland and between neighbouring countries, combined with limited knowledge of infections and a poor healthcare system, were factors that enabled recurring epidemics of various diseases, including measles. While armed conflicts between countries were not uncommon, wars and border changes affected, among many other things, the common welfare of people. Finland had been part of Sweden for centuries, but in 1721, Russia annexed the southeastern parts of Finland, or Old Finland, including the significant trading center of Vyborg. 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, impacting markets and trade connections [28]. 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.

We analyse information obtained from parish records, using monthly data from January 1750 to December 1850 from 387 different towns located in the southern half of Finland. The original records offer count information on the deaths at the town level, but since the counts are quite small (99% of all measles counts in our data are three or less) and there are no areal population sizes available—merely nationwide approximations [43]—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. [29], 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. [11, 10]. Pasanen et al. [29] also implies that the dichotomisation of the observations results in conclusions corresponding to those of undichotomised data.

The data come with a significant amount of missing observations for several reasons. These include misdiagnoses, varying personal practices in maintaining the registers [30], and loss of records due to fires and other conditions over time. There are a total of 1212 time points, each having observations from 75–272 sites simultaneously, so no month has information from all the sites at once. From another perspective, only one town has no missing observations at all, whereas all observations are missing from 44 towns. Overall, at the beginning of the time series, the missingness is approximately 70%, decreasing steadily to about 20% at the end of the time series, so that 40% of all data are missing.

2.2 Hidden Markov Models

Discrete time hidden Markov models (HMM) [33] 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, two states could be the presence and absence of a disease. Next, we introduce the method in more detail.

The first component of the model is the hidden stochastic process x_t, $t = 1, \ldots, T$, with the state space $\{1, \ldots, S\}$, i.e., having values from 1 to S. Usually the temporal dependency of the process x_t encompasses only one previous time point so that it satisfies the first-order Markov property

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

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 $s \in \{1, \ldots, S\}$ 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 $s, s' \in \{1, \ldots, S\}$. Additionally, initial state probabilities $\rho_s = P(x_1 = s)$, $s \in \{1, \ldots, S\}$, 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, $t = 1, \ldots, T$, where y_t depends only on the current state of the hidden process. In consequence, the observations are conditionally independent:

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

Each state x_t is associated with parameters θ_{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 θ_{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 Ω, where $\omega_{s,t} = p(y_t|\theta_{x_t=s})$. Now the HMM can be defined by the set $\{\Omega, A, \boldsymbol{\rho}\}$.

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 1.

While various methods have been suggested for estimating the number of states, S, for example, Pohle et al. [31], 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 $\{\Omega, A, \boldsymbol{\rho}\}$, and on the hidden state trajectory $x_1, \ldots, x_T$, given the observations y_t. As an example, consider HMM consisting of binary observations and two states, with $\boldsymbol{\rho} = (1, 0)^T$, $a_{1,1} = 0.9$, $a_{2,2} = 0.5$, $\omega_{1,t} = \text{Bernoulli}(0.1)$, 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 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,

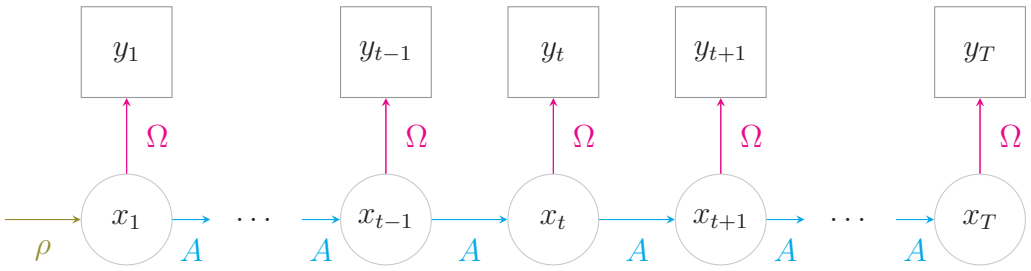

Figure 1: Directed graph of a hidden Markov model. The state x_1 in the first time point depends on the initial probabilities ρ, and the following transitions are determined by the transition probabilities in A. The observations y_t depend on the current state x_t via the emission probabilities in Ω,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

$p(\Omega, A, \rho|y)$, where $y = (y_1, \ldots, y_T)$. 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, 33, and the references therein]. The latent states $x = (x_1, \ldots, x_T)$ 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. 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 [44].

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 $i = 1, \ldots, N$, and correspondingly its neighbour with $j = 1, \ldots, N$. 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

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

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

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

where $\tau \in \mathbb{R}^+$ is a precision parameter of the variables φ_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

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

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 φ_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 significant amount of spatial association, the parameter α 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 [25]. Therefore, in some cases, it might be convenient to choose ICAR in the first place.

2.4 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 historical Finnish parish register data to study the dynamics of the measles epidemics. More precisely, our response variable $y_{i,t}$ is dichotomous, indicating the presence of at least one measles induced death observation in town i in time point t.

As before, we denote each site with $i = 1, \ldots, N$, and each time point with $t = 1, \ldots, T$. In this setting, the probability of observing at least one death $y_{i,t}$ caused by measles in site i and time point t can be modeled as

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

where the states

$$x_t | x_{t-1} \sim \text{Categorical}(A_{x_{t-1},\cdot}). \tag{7}$$

The notation $A_{x_{t-1},\cdot}$ means the row x_{t-1} including all columns of the 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 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 formula 7 defines the connection between the underlying states x_t and the transition matrix A.

The state specific constant $\mu_{x_t=s}$, $s \in \{1, \ldots, S\}$, 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, μ_1 and μ_S, respectively. We use these two to scale the ones for the other states. We do this by defining a simplex $\boldsymbol{m}$ 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

$$\mu_s = \mu_1 + (\mu_S - \mu_1) \sum_{i=1}^{s} m_i \text{ for all } s = 1, \ldots, S. \tag{8}$$

As we are using a Bayesian approach, this allows us to employ informative priors for the constants of the first and last state, μ_1 and μ_S.

To allow local deviations from the nationwide base level, we add a site dependent term λ_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 μ_s as a nationwide average level.

The actual spatial dependence between the sites is introduced in the third term, $\sigma_\varphi \varphi_{i,x_t}$. The dependency is achieved via the ICAR structure outlined above, i.e., adapting the formulas 3 and 4, and letting $\boldsymbol{\varphi}_{x_t=s} \sim \text{N}(0, (D - W)^{-1})$, where $s \in \{1, \ldots, S\}$, and $\sigma_\varphi = \tau^{-\frac{1}{2}}$. To employ the ICAR component, we need a definition for the neighbourhood. Here two regions are considered as neighbours if they share a border. Other definitions, for example, based on the distance from the centers 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. The state specific spatial term φ_{i,x_t} is zero on average over the sites to guarantee identifiability using soft constraint as in Morris et al. [25].

The final component γ_t in the model is a monthly average deviation additional to all the other components. It is constrained to sum to zero over a year.

As mentioned in subsection 2.1, our data are missing a large number of observations. However, as long as we assume 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 [44]. Thus, the absent information can be omitted while estimating the model, yet the model provides estimates for all sites and time points. Even though the missing information is omitted, its effect and the uncertainty is reflected in the width of the posterior intervals of the estimates.

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

$$
\begin{aligned}
\mu_1 &\sim \mathrm{N}(-4.5, 0.25^2),\\
\mu_S &\sim \mathrm{N}(-1.75, 0.5^2)[\mu_1,],\\
\boldsymbol{m}_{2:S} &\sim \mathrm{Dirichlet}(1_1, \ldots, 1_{S-1}),\\
\lambda_i &\sim \mathrm{N}(0, \sigma_\lambda^2), \text{ for all } i = 1, \ldots, N, \text{ given } \frac{1}{N}\sum_{i=1}^{N} \lambda_i = 0,\\
\sigma_\lambda &\sim \mathrm{N}(0, 1^2)[0,],\\
\sigma_\varphi &\sim \mathrm{N}(0, 1^2)[0,],\\
\boldsymbol{\varphi}_s &\sim \mathrm{N}(0, (D - W)^{-1}), \text{ for all } s = 1, \ldots, S,\\
\gamma_t &\sim \mathrm{N}(0, 1^2), \text{ for all } t = 1, \ldots, 12, \text{ given } \frac{1}{12}\sum_{i=1}^{12} \gamma_t = 0,\\
\boldsymbol{\rho} &\sim \mathrm{Dirichlet}(1_1, \ldots, 1_S), \text{ and}\\
A_{s,\cdot} &\sim \mathrm{Dirichlet}(p_1, \ldots, p_S), \text{ where } p_{s'} = \begin{cases} 0.5, \text{ when } s' \neq s \\ 2 \cdot S, \text{ when } s' = s, \end{cases}
\end{aligned}
$$

where $\mathrm{N}(.,.)[z,]$ denotes a normal distribution truncated at z, and $\boldsymbol{m}_{2:S}$ denotes the components from 2 to S of the vector $\boldsymbol{m}$, omitting the first component since it is zero. The priors for the state specific constants μ_1 and μ_S are chosen based on the data, to roughly match with the 10th and 90th percentiles of the monthly averages of the observed death occurrences. Such utilisation of the data is a common practice, for example, in *rstanarm* [26]. The row sums of the transition matrix A are one, as well as the sums of the parameters ρ and a, which is achieved with the Dirichlet priors. These priors are mostly weakly informative and chosen to enhance the computational efficiency [26, 5]. In our case, the number of time points is $T = 1212$ and the number of sites $N = 387$. For our final model, we set the number of states to be $S = 5$.

3 Results

The model is estimated with MCMC using *cmdstanr* [15], which is an R interface [32] for the probabilistic programming language Stan for statistical inference [39]. Posterior samples are drawn using NUTS sampler [19, 9] with four chains. Each chain consists of 7,500 iterations, the first 5,000 being discarded as warm-up. The computation with parallel chains takes approximately 30 hours. The estimation is done on a supercomputer node with four cores of Xeon Gold 6230 2.1 GHz processors, each core allocated 15 GB of RAM. According to the MCMC diagnostics of the *cmdstanr* [41], the model converges without any divergences. The $\widehat{R}$ statistics are all below 1.004, and the bulk effective sample sizes are roughly between 640 and 28000, the most inefficient one being $\varphi_{29,5}$, which corresponds to a town and month that, while not associated with a missing observation itself, is missing five out of six of the neighbouring observations. The R and Stan codes and the data used for the analysis are available on GitHub (`https://github.com/tihepasa/spatialHMM`).

The data, aggregated temporally, the predicted probabilities of observing at least one death caused by measles, and the state trajectories that depict the path of the most probable states are illustrated in Figure 2. The predictions are computed according to the model as $\text{logit}^{-1}(\mu_{x_t} + \lambda_i + \sigma_\varphi \varphi_{i,x_t} + \gamma_t)$, followed by taking an average over all the sites. 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 12% 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 approximately 90% of all of the probabilities being over 89%, and only five, randomly distributed in terms of time, falling below 50%.

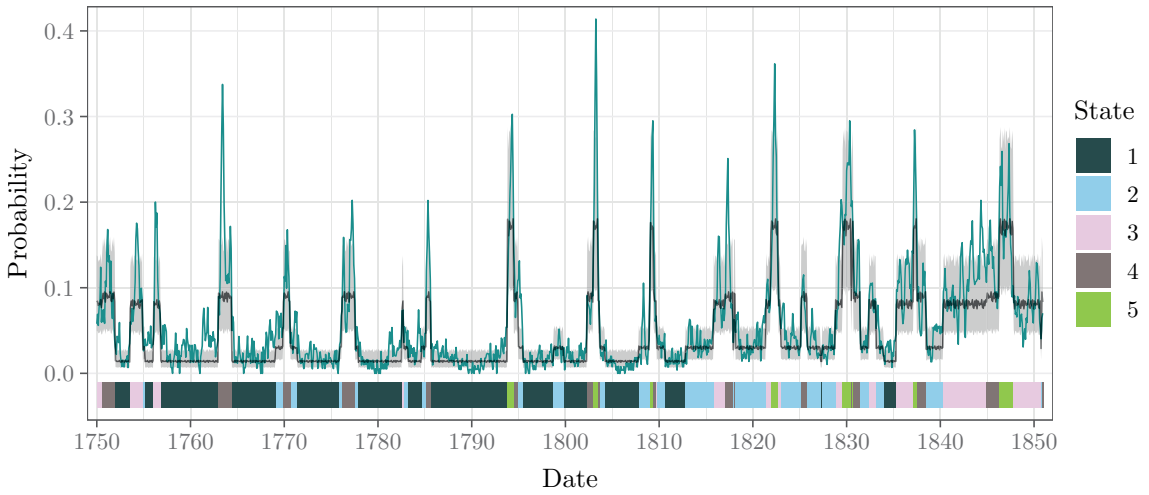

Figure 2: The turquoise line denotes the proportion of towns where at least one death caused by measles was observed in that month according to the data. The black line represents the nationwide average over local probabilities to observe at least one death caused by measles given by the model: $\text{logit}^{-1}(\mu_{x_t}+\lambda_i+\sigma_\varphi\varphi_{i,x_t}+\gamma_t)$. The grey area around the black line shows the 95% posterior interval of the predicted average. The colors below the curve indicate in which state it was most likely to be at that time.

The spatial dimension of the states can be seen in Figure 3, 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 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 2. 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.

According to the transition probabilities, shown in Table 1, 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 States 3 and 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 ρ_s, along with other scalar parameters, can be found from Table 2. 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.

Table 1: 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)	0.02 (0.01, 0.03)	0.01 (0.00, 0.02)	0.00 (0.00, 0.01)	0.00 (0.00, 0.01)
State 2	0.04 (0.02, 0.07)	0.91 (0.88, 0.95)	0.02 (0.01, 0.05)	0.02 (0.01, 0.04)	0.00 (0.00, 0.01)
State 3	0.01 (0.00, 0.03)	0.03 (0.01, 0.07)	0.91 (0.87, 0.95)	0.02 (0.01, 0.05)	0.02 (0.01, 0.05)
State 4	0.02 (0.00, 0.06)	0.05 (0.02, 0.10)	0.02 (0.00, 0.05)	0.89 (0.83, 0.93)	0.02 (0.00, 0.05)
State 5	0.01 (0.00, 0.03)	0.01 (0.00, 0.03)	0.04 (0.01, 0.010)	0.08 (0.03, 0.16)	0.87 (0.78, 0.94)

The scalar constants μ_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 2. When these are transformed into a probability scale, the base probabilities range from 1% in State 1 to 16% in State 5.

The local constants λ_i, independent of the state, are shown in Figure 4. As expected, there are no visible spatial patterns in these parameters aimed to capture spatially independent differences of the sites. The deviation σ_λ of the constants λ_i is 0.59 with posterior interval [0.53, 0.65], see also Table 2, 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 $\sigma_\varphi \varphi_{i,s}$. The deviation parameter σ_φ is constant for all the sites and states and is estimated to be 0.89 [0.82, 0.96], see also Table 2. The state-specific spatial terms $\sigma_\varphi \varphi_{i,s}$ are shown in Figure 5. The spatial patterns are similar to those of the estimated probabilities to observe

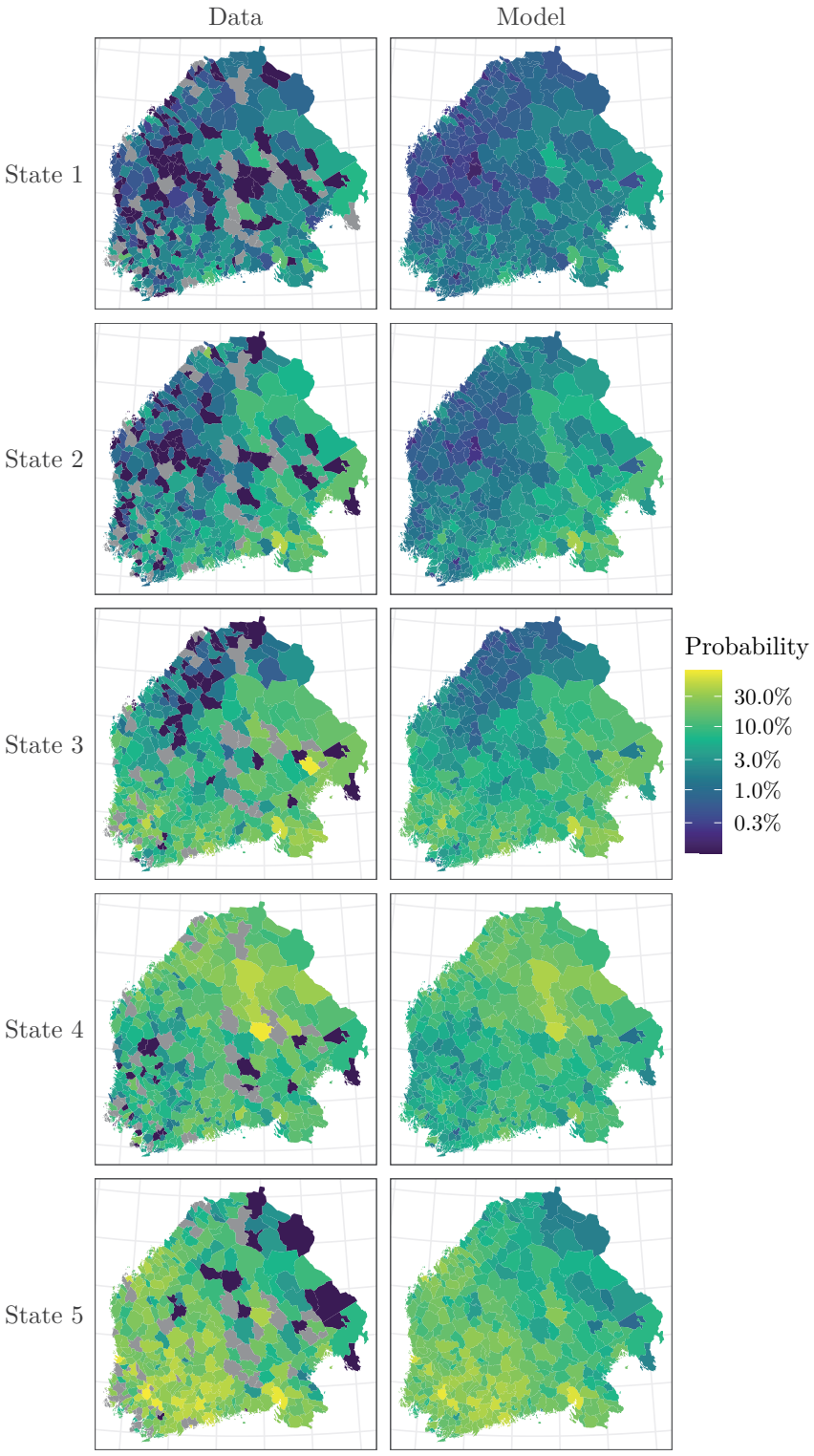

Figure 3: On the left column are the probabilities to observe at least one death caused by measles in each town as means over the months when each state was most likely, and on the right column are the corresponding estimated probabilities $\text{logit}^{-1}(\mu_{x_t} + \lambda_i + \sigma_\varphi \varphi_{i,x_t} + \gamma_t)$. The rows represent different states. Gray area indicates a site with only missing observations in months corresponding to that state. Note that the color scale is logarithmic.

at least one death caused by measles in Figure 3.

To capture the yearly temporal variation, the model includes a seasonal component γ_t, which is shown in Figure 6. These monthly terms sum to zero over a year, i.e., they depict

Table 2: Posterior means and standard deviations, 95% posterior intervals, bulk and tail effective sample sizes (ess) and $\widehat{R}$ statistics of the initial probabilities ρ_s, the state specific constants μ_s, the deviation of local constants σ_λ, and the deviation parameter σ_φ of the ICAR component.

	mean	sd	2.5%	97.5%	ess bulk	ess tail	$\widehat{R}$
ρ_1	0.17	0.14	0.01	0.52	17657	5620	1.00
ρ_2	0.17	0.14	0.01	0.53	18953	5898	1.00
ρ_3	0.33	0.18	0.05	0.73	20350	6151	1.00
ρ_4	0.17	0.14	0.00	0.53	15146	6368	1.00
ρ_5	0.17	0.14	0.01	0.53	16702	7113	1.00
μ_1	-4.55	0.04	-4.64	-4.47	3595	5045	1.00
μ_2	-3.87	0.04	-3.96	-3.79	4056	7210	1.00
μ_3	-2.76	0.03	-2.82	-2.69	4229	6986	1.00
μ_4	-2.43	0.03	-2.49	-2.36	1011	4169	1.00
μ_5	-1.68	0.03	-1.74	-1.61	3109	5485	1.00
σ_λ	0.59	0.03	0.53	0.65	2941	5410	1.00
σ_φ	0.89	0.03	0.82	0.96	1368	2622	1.00

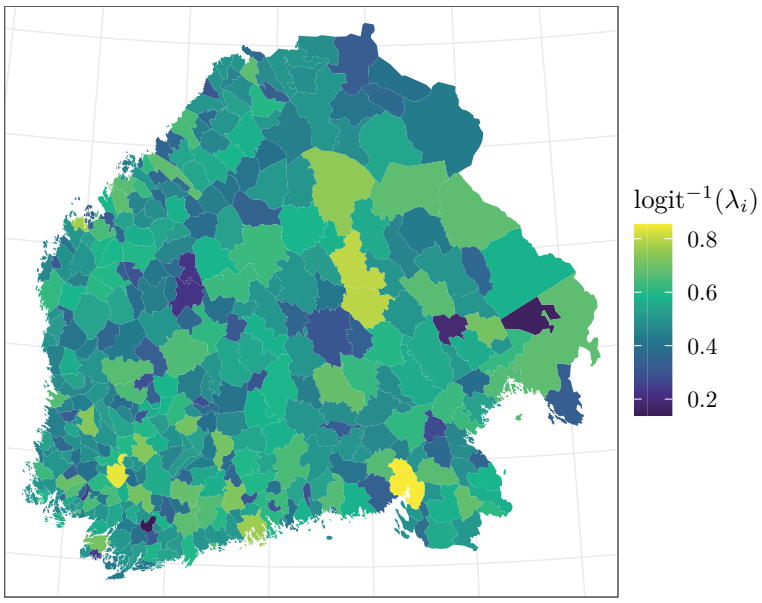

Figure 4: Inverse logit transformations of the posterior means of the local constants λ_i.

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. 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 [29] for measles during 1820–1850 in Finland.

While our HMM was estimated as homogeneous in terms of the transition matrix A, Figure 2 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

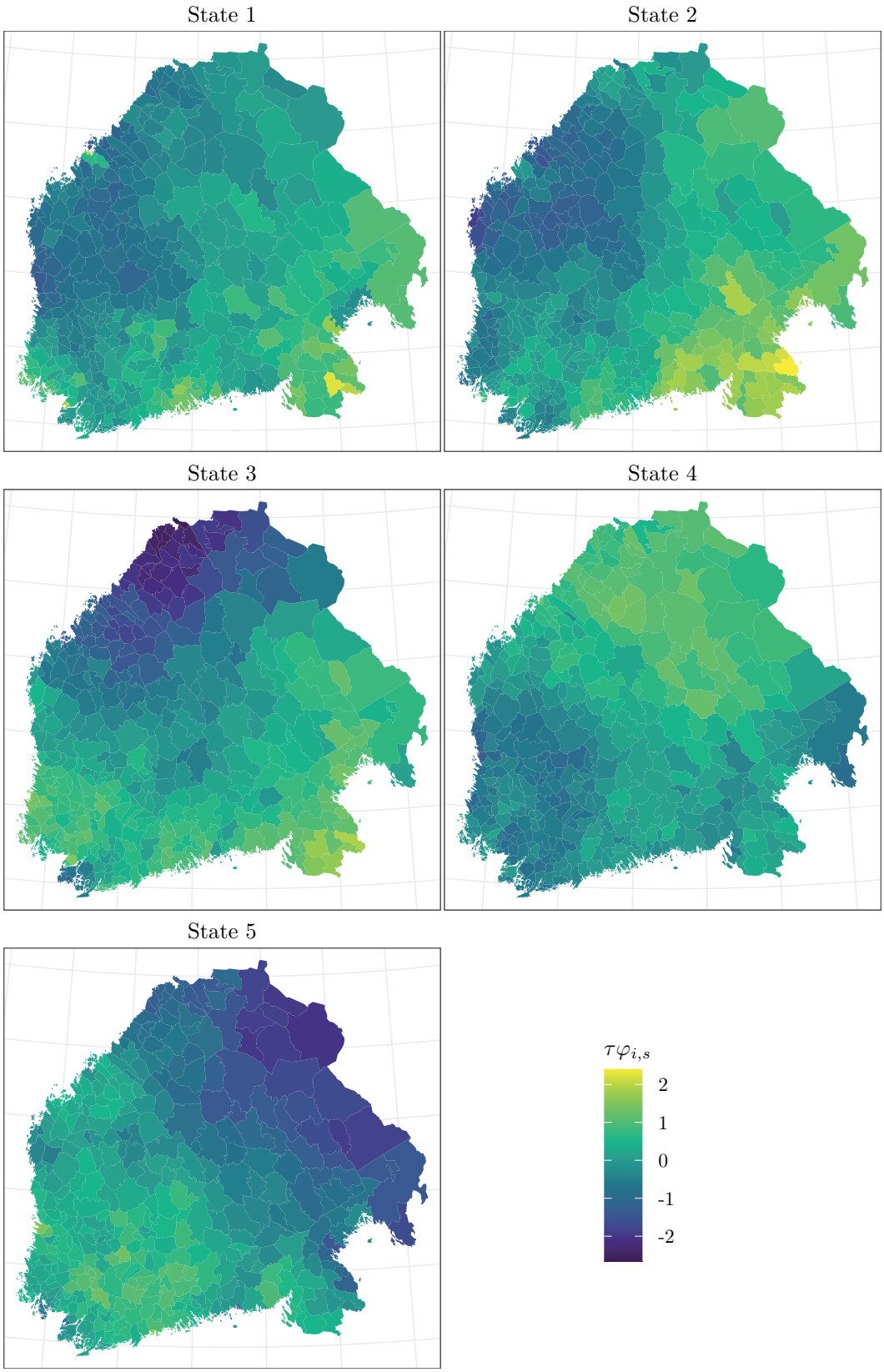

Figure 5: Posterior means of the spatial terms $\sigma_\varphi \varphi_{i,s}$ by state.

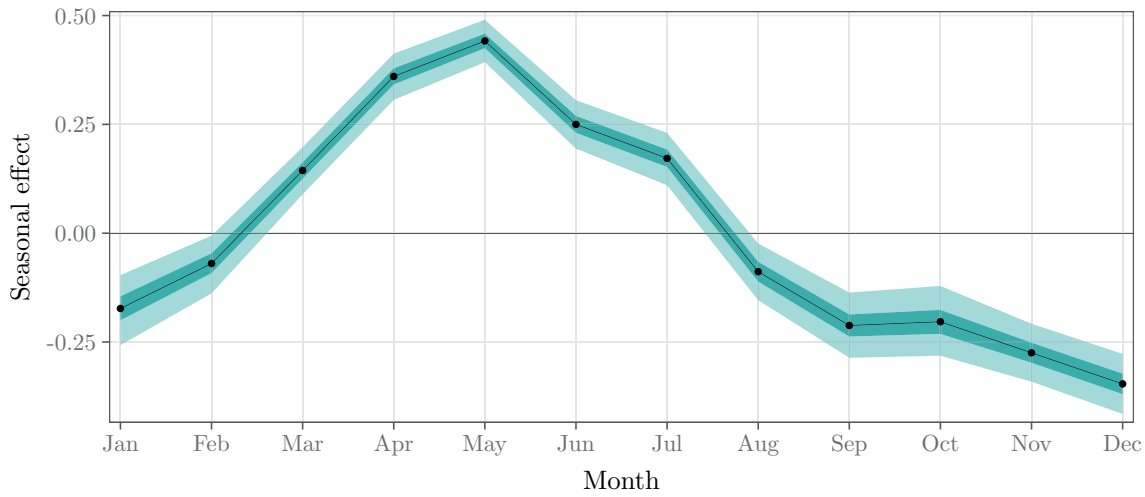

Figure 6: The black dots and line show the posterior mean of the monthly seasonal term γ_t. The dark and pale turquoise areas represent the 50% and 95% posterior intervals, respectively.

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 modeled 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 September 1812 to February 1813). Table 3 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 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 [28].

Table 3: 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, 0.13)	0.05 (0.05, 0.05)	0.11 (0.11, 0.12)	0.02 (0.02, 0.02)
After	0.04 (0.04, 0.04)	0.39 (0.39, 0.39)	0.34 (0.33, 0.34)	0.14 (0.14, 0.14)	0.09 (0.09, 0.10)

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 [e.g., 31]. 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.

For comparison, the supplementary material contains figures (Supplementary Figures 1–5) and tables (Supplementary Tables 1–2) for the four-state model, corresponding the ones represented here for the five-state model in Figures 2–6 and Tables 1–2. Based on Figures 2 and 3 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).

4 Discussion

In this study, we illustrated 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 Finns 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 3).

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 the eastern influence and trade connections [28]. 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 3), possibly indicating persistent infectious pressure from the closest city in the area, Saint Petersburg. 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 [43] 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 [21, 16, 34, 22]. Epidemics are affected by transport networks, geographical obstacles and borders [42]. Finland and Sweden are separated by a large geographical barrier: The Gulf of Bothnia is 80–150 kilometers wide and 700 kilometers 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 [28]. 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 $y_{i,t} \sim \text{Bernoulli}(\mu_t + \lambda_i + \sigma_\varphi \varphi_{i,t} + \gamma_t)$, where $\mu_t \sim \text{N}(\mu_{t-1}, \sigma_\mu^2)$, and $\boldsymbol{\varphi}_t \sim \text{N}(\boldsymbol{\varphi}_{t-1}, (D - W)^{-1})$. 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]. However, 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. Additional downside would be 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, yet hopefully easy to understand, 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 μ_s and spatial terms $\varphi_{i,t}$, also the local constant λ_i, deviation parameter σ_φ and the seasonal term γ_t could depend on state if seen appropriate. Our attempts to include the state dependent local constants λ_i and deviations σ_φ lead to multimodality and convergence issues. Thus, the inclusion of those further complexities in our model remains as a future work.

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 [29]. The model may also be extended to a case with exogenous covariates for the

observational distributions or for the transition probabilities.

Acknowledgements

Tiia-Maria Pasanen was supported by the Finnish Cultural Foundation, the Emil Aaltonen Foundation and the Research Council of Finland grant 331817. Jouni Helske was supported by the Research Council of Finland grants 331817 and 355153. Tarmo Ketola was supported by the Research Council of Finland grant 278751. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. The authors also thank Virpi Lummaa and Harri Högmander.

References

[1] Amorós, R., Conesa, D., López-Quílez, A., Martinez-Beneito, M.A., 2020. A spatio-temporal hierarchical Markov switching model for the early detection of influenza outbreaks. Stochastic environmental research and risk assessment 34, 275–292. doi:https://doi.org/10.1007/s00477-020-01773-5.

[2] Bakka, H., Rue, H., Fuglstad, G.A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., Lindgren, F., 2018. Spatial modeling with R-INLA: A review. WIREs Computational Statistics 10, e1443. doi:https://doi.org/10.1002/wics.1443.

[3] Balocchi, C., Jensen, S.T., 2019. Spatial modeling of trends in crime over time in Philadelphia. The Annals of Applied Statistics 13, 2235–2259. doi:https://doi.org/10.1214/19-aoas1280.

[4] Banerjee, S., Carlin, B., Gelfand, A., 2015. Hierarchical Modeling and Analysis for Spatial Data, Second Edition. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis. doi:https://doi.org/10.1201/b17115-10.

[5] Banner, K.M., Irvine, K.M., Rodhouse, T.J., 2020. The use of Bayesian priors in Ecology: The good, the bad and the not great. Methods in Ecology and Evolution 11, 882–889. doi:https://doi.org/10.1111/2041-210X.13407.

[6] Baum, L.E., Petrie, T., 1966. Statistical inference for probabilistic functions of finite state Markov chains. The Annals of Mathematical Statistics 37, 1554–1563. doi:https://doi.org/10.1214/aoms/1177699147.

[7] Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological) 36, 192–225. doi:https://doi.org/10.1111/j.2517-6161.1974.tb00999.x.

[8] Besag, J., York, J., Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics 43, 1–20. doi:https://doi.org/10.1007/BF00116466.

[9] Betancourt, M., 2018. A conceptual introduction to Hamiltonian Monte Carlo. arXiv doi:https://doi.org/10.48550/arXiv.1701.02434. [Preprint]. Posted July 16, 2018 [accessed May 3, 2024].

[10] Briga, M., Ketola, T., Lummaa, V., 2022. The epidemic dynamics of three childhood infections and the impact of first vaccination in 18th and 19th century Finland. medRxiv doi:https://doi.org/10.1101/2022.10.30.22281707. [Preprint]. Posted October 31, 2022 [accessed May 25, 2023].

[11] Briga, M., Ukonaho, S., Pettay, J.E., Taylor, R.J., Ketola, T., Lummaa, V., 2021. The seasonality of three childhood infections in a pre-industrial society without schools. medRxiv doi:https://doi.org/10.1101/2021.10.08.21264734. [Preprint]. Posted October 18, 2021 [accessed September 29, 2023].

[12] Bürkner, P.C., Charpentier, E., 2020. Modelling monotonic effects of ordinal predictors in Bayesian regression models. British Journal of Mathematical and Statistical Psychology 73, 420–451. doi:https://doi.org/10.1111/bmsp.12195.

[13] Clayton, D., Bernardinelli, L., 1996. Bayesian methods for mapping disease risk, in: Geographical and Environmental Epidemiology: Methods for Small Area Studies. Oxford University Press. doi:https://doi.org/10.1093/acprof:oso/9780192622358.003.0018.

[14] Fox, M.P., Murray, E.J., Lesko, C.R., Sealy-Jefferson, S., 2022. On the need to revitalize descriptive epidemiology. American Journal of Epidemiology 191, 1174–1179. doi:https://doi.org/10.1093/aje/kwac056.

[15] Gabry, J., Češnovar, R., 2022. cmdstanr: R Interface to 'CmdStan'. Https://mc-stan.org/cmdstanr/, https://discourse.mc-stan.org.

[16] Grenfell, B., Harwood, J., 1997. (Meta)population dynamics of infectious diseases. Trends in Ecology & Evolution 12, 395–399. doi:https://doi.org/10.1016/S0169-5347(97)01174-9.

[17] Helske, S., Helske, J., 2019. Mixture hidden Markov models for sequence data: The seqHMM package in R. Journal of Statistical Software 88, 1–32. doi:https://doi.org/10.18637/jss.v088.i03.

[18] Helske, S., Helske, J., Eerola, M., 2018. Combining sequence analysis and hidden Markov models in the analysis of complex life sequence data, in: Ritschard, G., Studer, M. (Eds.), Sequence Analysis and Related Approaches. Life Course Research and Social Policies. Springer International Publishing. volume 10, pp. 185–200. doi:https://doi.org/10.1007/978-3-319-95420-2_11.

[19] Hoffman, M.D., Gelman, A., 2014. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 15, 1593–1623.

[20] Jasra, A., Holmes, C.C., Stephens, D.A., 2005. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. Statistical Science 20, 50 – 67. doi:https://doi.org/10.1214/088342305000000016.

[21] Keeling, M.J., Grenfell, B.T., 1997. Disease extinction and community size: Modeling the persistence of measles. Science 275, 65–67. doi:https://doi.org/10.1126/science.275.5296.65.

[22] Ketola, T., Briga, M., Honkola, T., Lummaa, V., 2021. Town population size and structuring into villages and households drive infectious disease risks in pre-healthcare Finland. Proceedings of the Royal Society B: Biological Sciences 288, 20210356. doi:https://doi.org/10.1098/rspb.2021.0356.

[23] Knorr-Held, L., 2000. Bayesian modelling of inseparable space-time variation in disease risk. Statistics in Medicine 19, 2555–2567. URL: https://doi.org/10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-%23.

[24] Knorr-Held, L., Richardson, S., 2003. A hierarchical model for space–time surveillance data on meningococcal disease incidence. Journal of the Royal Statistical Society Series C: Applied Statistics 52, 169–183. doi:https://doi.org/10.1111/1467-9876.00396.

[25] Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S.J., Gelman, A., DiMaggio, C., 2019. Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan. Spatial and Spatio-temporal Epidemiology 31, 100301. doi:https://doi.org/10.1016/j.sste.2019.100301.

[26] Muth, C., Oravecz, Z., Gabry, J., 2018. User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. The Quantitative Methods for Psychology 14, 99–119. doi:10.20982/tqmp.14.2.p099.

[27] Ogden, H., 2017. On asymptotic validity of naive inference with an approximate likelihood. Biometrika 104, 153–164. doi:https://doi.org/10.1093/biomet/asx002.

[28] Ojala, J., Räihä, A., 2017. Navigation Acts and the integration of North Baltic shipping in the early nineteenth century. International Journal of Maritime History 29, 26–43. doi:https://doi.org/10.1177/0843871416678166.

[29] Pasanen, T.M., Helske, J., Högmander, H., Ketola, T., 2024. Spatio-temporal modeling of co-dynamics of smallpox, measles, and pertussis in pre-healthcare Finland. PeerJ 12, e18155. doi:https://doi.org/10.7717/peerj.18155.

[30] Pitkänen, K., 1977. The reliability of the registration of births and deaths in Finland in the eighteenth and nineteenth centuries: Some examples. Scandinavian Economic History Review 25, 138–159. doi:https://doi.org/10.1080/03585522.1977.10407878.

[31] Pohle, J., Langrock, R., Van Beest, F.M., Schmidt, N.M., 2017. Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement. Journal of Agricultural, Biological and Environmental Statistics 22, 270–293. doi:https://doi.org/10.1007/s13253-017-0283-8.

[32] R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

[33] Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77, 257–286. doi:https://doi.org/10.1109/5.18626.

[34] Rohani, P., Green, C., Mantilla-Beniers, N., Grenfell, B., 2003. Ecological interference between fatal diseases. Nature 422, 885–888. doi:https://doi.org/10.1038/nature01542.

[35] Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71, 319–392. doi:https://doi.org/10.1111/j.1467-9868.2008.00700.x.

[36] Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P., Lindgren, F.K., 2017. Bayesian computing with INLA: A review. Annual Review of Statistics and Its Application 4, 395–421. doi:https://doi.org/10.1146/annurev-statistics-060116-054045.

[37] Saarivirta, T., Consoli, D., Dhondt, P., 2009. Suomen terveydenhuoltojärjestelmän ja sairaaloiden kehittyminen. Vaatimattomista oloista modernin terveydenhuollon eturintamaan. Kasvatus & Aika 3. URL: https://journal.fi/kasvatusjaaika/article/view/68129.

[38] Saarivirta, T., Consoli, D., Dhondt, P., 2012. The evolution of the Finnish health-care system early 19th century and onwards. International Journal of Business and Social Sciences 3, 243–257.

[39] Stan Development Team, 2024. Stan modeling language users guide and reference manual, version 2.34. URL: https://mc-stan.org.

[40] Tilastollinen päätoimisto, 1899. Pääpiirteet Suomen väestötilastosta vuosina 1750-1890: 1: Väestön tila. SVT: Suomen virallinen tilasto VI. Väkiluvun-tilastoa 29 Table 1 A.

[41] Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., Bürkner, P.C., 2021. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC (with discussion). Bayesian Analysis 16, 667–718. doi:https://doi.org/10.1214/20-BA1221.

[42] Vora, A., Burke, D.S., Cummings, D.A.T., 2008. The impact of a physical geographic barrier on the dynamics of measles. Epidemiology and Infection 136, 713–720. doi:https://doi.org/10.1017/S0950268807009193.

[43] Voutilainen, M., Helske, J., Högmander, H., 2020. A Bayesian reconstruction of a historical population in Finland, 1647–1850. Demography 57, 1171–1192. doi:https://doi.org/10.1007/s13524-020-00889-1.

[44] Zucchini, W., MacDonald, I., 2009. Hidden Markov Models for Time Series: An Introduction Using R. Chapman and Hall/CRC. doi:https://doi.org/10.1201/9781420010893.

IV

ADAPTATION TO PATERNAL LEAVE POLICIES IN FINNISH MUNICIPALITIES: CHANGING GENDER NORMS AND CROSS-BORDER POLICY LEGACIES

by

Pasanen T.-M., Helske, S., Giuliani, G. A., Chapman, S. N., and Helske, J. 2024

Preprint

DOI: https://doi.org/10.31235/osf.io/k27yw

Adaptation to paternal leave policies in Finnish municipalities: changing gender norms and cross-border policy legacies

Tiia-Maria Pasanen[1*], Satu Helske[2], Giovanni Amerigo Giuliani[3], Simon N. Chapman[2], and Jouni Helske[2]

[1]Department of Mathematic and Statistics, University of Jyväskylä, Finland

[2]INVEST Research Flagship Centre, University of Turku, Finland

[3]Department of Political and Social Sciences, University of Bologna, Italy

[*]Corresponding author: Tiia-Maria Pasanen; tiia-maria.h.pasanen@jyu.fi

Abstract

Countries with stronger support to gender equality tend to invest in policies supporting the dual-earner or earner–carer family models, while countries with stronger support to conservative gender norms tend to support the male breadwinner model. However, whilst gender equality norms can be endorsed by the majority of the population, conservative norms could still be largely supported at the sub-national level, potentially leading to lower uptake of parental leave among fathers in more conservative areas.

This study seeks to examine shifting norms in fathers' parental leave uptake in Finnish municipalities in the 2010s, around the first reform that gave fathers an independent right to a 6-week quota of "solo" parental leave. We applied a Bayesian spatio-temporal model on administrative data from Finnish municipalities and estimated local norms based on voting data. Furthermore, we used the proportion of Swedish residents as a proxy for cross-border policy influences from the neighbouring country Sweden, where paternal leave-taking has been a longer phenomenon.

Local support to de-familialising policies was found to predict higher leave-taking, but only under less supportive policy configuration. The proportion of Swedish-speaking residents was found to be increasingly important for predicting paternal leave-taking. We interpret this result as a sign of cross-border influences from Sweden. Interestingly, uptake increased the fastest in a more conservative region, probably due to its strong linguistic and cultural links to Sweden. Furthermore, we observed spatial dependencies between neighbouring municipalities, which supports our spillover hypothesis; that interactions among families nearby lead them to adopt new practices.

These results trigger new theoretical considerations concerning the role of gender norms in affecting citizens' behaviour in the family policy field and how these norms interact with policy reforms, the role of national institutions in affecting welfare state preferences at the local level, and the importance of community socialisation.

Keywords: Fathers, parental leave, gender norms, geographical context, spatio-temporal

modelling, Bayesian statistics

Introduction

In recent years, family policies have shifted towards more gender-neutral and father-specific parental leaves to encourage fathers to take more parental leave (Daly, 2020). This shift is part of broader efforts to promote gender equality: increasing fathers' involvement in care work is seen as crucial for reducing the gender wage gap and improving women's economic well-being (Ponthieux and Meurs, 2015). Despite the policy developments, many fathers still forgo leave or use less than their entitlement (Koslowski et al., 2019). While fathers have increased leave-taking and the time spent with their children, parenting remains gendered even in the most supportive policy contexts (Eydal and Rostgaard, 2023; Blum et al., 2023; Duvander et al., 2019).

The design of parental leave policies affect uptake and can support gender equality or reinforce inequalities (Koslowski et al., 2019; O'Brien and Wall, 2017). Long periods of low-paid or unpaid sharable leave tend to be predominately taken by mothers, while ear-marking well-paid, non-transferrable solo leave for the father has been shown to be effective in boosting active paternal participation beyond infancy to the long term and shifting gender norms towards more equal parenting (Duvander et al., 2019; Fernández-Cornejo et al., 2016; Koslowski et al., 2019; O'Brien and Wall, 2017; Omidakhsh et al., 2020; Ottosen, 2014; Tamm, 2019).

On the other hand, individual gender role attitudes influence the configuration of family policies: countries with stronger support to gender equality tend to invest in policies supporting the dual-earner family model or the earner–carer family model, while countries with stronger support to conservative gender norms tend to have policies supporting the male breadwinner family model (Duvander and Ellingsæter, 2016; Häusermann, 2018). However, gender equality norms can be endorsed by the majority of the population at the national level, whilst conservative norms could still be largely supported at the sub-national level (Lomazzi, 2017; Pfau-Effinger, 2023). As a consequence, though paternal leave entitlements may been implemented and available throughout the society, their uptake can substantially vary: fathers in conservative areas are assumed to be slower to adopt new leave practices, while their counterparts in more liberal areas may be quicker to adapt to the expansion of paternal entitlements.

This study seeks to examine shifting norms in fathers' parental leave uptake in Finnish municipalities in the 2010s, around the first reform that gave fathers an independent right to a six-week quota of "solo" parental leave, intended to be taken after the mother's return to work or studies. While Finland represents the Nordic family policy model, where both parents are considered responsible for both earning and caring (Korpi, 2000; Lewis, 2009), and has a long tradition of supporting maternal labour force participation, Finland used to lag behind its Nordic counterparts when it comes to developing leave policies that support gender equality and fathers' independent leave entitlements. Using administrative data from Finnish municipalities and an innovative method for determining local gender norms as well as cross-border policy influences from Sweden—the neighbouring country that has been the forerunner in gender-equal parental leave—this study provides new insights into regional analysis and how local contexts and cultural influences shape responses to national policies.

Gender Norms and the Shape of Family Policy: Theoretical Considerations

The evolution of gender norms

There is a very well-informed literature that has investigated how social norms contribute to shaping family policy at the national level (e.g., (Sjöberg, 2004; Pfau-Effinger, 2014; Kangas and Rostgaard, 2007; Rossier et al., 2011; Emmenegger and Manow, 2014; Pavolini et al., 2021)). With social norms, we refer to those shared modes of regulations connected to forms of social recognition (Dubois, 2002). Social norms rest more or less explicitly on underlying values, pointing to what is socially acceptable and what is not (Bicchieri et al., 2023). Consequently, social norms are prescriptive since they imply sanctions for those who decide not to comply with them (Rossier et al., 2011). Furthermore, social norms are plural since a diverse and even conflicting array of norms may coexist in the same sphere—including the family one (Rossier et al., 2011). Norms supported and promoted by the highest source of authority—such as the State or the most influential group—are defined as dominant and end up being shared by the majority of the society, though to a different extent.

When focusing on the family sphere, individual attitudes concerning gender norms are assumed to influence the specific configuration of family policy at the national level. These attitudes can be interpreted as the cognitive representation of what is believed appropriate for males and females in a specific context and are thus associated with preferences for a specific family model and gender roles inside and outside the labour market (André et al., 2013; Lomazzi, 2017; Lomazzi and Seddig, 2020). Individual attitudes concerning gender norms, however, are not static; on the contrary, they have evolved in parallel with societal modernisation (Inglehart, 1997). The theory of modernisation asserts that as a society attains significant levels of existential security, people become more receptive to adopting new social norms that match emerging post-materialist desires like autonomy and self-fulfilment (Inglehart, 2008; Inglehart and Norris, 2003).

Over the past seven decades, there has been a transition in Western societies from norms centred around survival to those emphasising individuality (Inglehart, 2018). The move toward individual-oriented norms has resulted in a progressive alteration of citizens' perceptions regarding gender roles. The rise of secularisation, increased educational attainment, greater participation of women in the workforce, and elevated divorce and single motherhood rates have catalysed a movement towards gender equality values (Inglehart and Norris, 2003). Consequently, the proportion of individuals embracing egalitarian or libertarian views, advocating against gender-based segregation within and outside the household, has significantly grown (Lomazzi and Seddig, 2020). In various national settings, this demographic now surpasses those adhering to conservative or authoritarian beliefs favouring gender-specific roles, where women are predominantly assigned caregiving responsibilities. In most Western European countries, this value shift became prominent mainly from the 1990s onward. Nevertheless, the pace and scale of this transformation differ among countries (e.g., Pavolini et al., 2021), and traditional gender norms—though having substantially weakened—have not disappeared and still coexist with the liberal ones (Giuliani, 2022, 2024).

Gender norms and the configuration of family policy instruments

The comparative literature has identified two macro-policy instruments within family policy (Leitner, 2003; Häusermann, 2018; Saraceno, 2022; Giuliani, 2024). On the one hand, familialising policy instruments intend to strengthen the family's caring function, primarily through passive measures (Leitner, 2003). Within this category, we find family allowances, child benefits, cash-for-care, and tax rebates for families. Regarding gender equality, these measures do not alter the uneven distribution of paid and unpaid work within the household (Jenson and Jacobzone, 2000). On the contrary, they may reproduce and reinforce the gendered specialisation of tasks, disincentivizing women to enter (or re-enter) the labour market (Duvander and Ellingsæter, 2016). For this reason, such instruments tend to be associated with the male breadwinner family model, where men work, and women care for children and the frail elderly.

On the other hand, de-familialising policy instruments facilitate care outside the family (Leitner, 2003; Saraceno and Keck, 2011). In this regard, childcare represents the main instrument for achieving this goal. Parental leave is a more complex policy instrument to classify. Empirical research has pointed out that relatively short and well-paid leave influences positively the mother's chances of re-entering the labour market (e.g., (Pylkkänen and Smith, 2004; Petersen et al., 2014; Ferragina, 2019)). Simultaneously, women's employment outcomes will likely improve when periods reserved for fathers are present and associated with high replacement rates (Kotsadam and Finseraas, 2011). De-familialising policy instruments thus foster the consolidation of the dual-earner family model, where both parents are fully employed (Häusermann, 2018). Regarding gender equality, childcare has a de-gendering effect since outsourcing care has a positive effect on the female employment rate (Korpi, 2000). However, it did not explicitly affect the redistribution of unpaid work per se. On the contrary, fathers' leave uptake is associated with a direct de-gendering goal since it helps to re-distribute caring tasks within the family between mothers and fathers in a more equal way. Fathers' leave is thus said to be associated with the earner–carer family model, where both parents are in the labour market and share caring responsibilities (Misra et al., 2007; Ciccia and Verloo, 2012).

The national configuration of the family policy consists of a mix of these policy instruments. The degree (weak vs. strong) to which both these policies are developed varies across welfare states; analysing their combinations, four different varieties of familialism emerge (Leitner, 2003; Häusermann, 2018). Explicit familialism strengthens the caring functions of the family but does not provide any alternative to family care. Optional familialism implies the parallel development of familialising and de-familialising policies; the family can be supported in its caring functions but alternatives are also offered. Implicit familialism provides neither de-familialisation nor active support for the caring tasks of the family. Finally, de-familialism is characterised by highly-developed de-familialising policies which unburden families from caring activity.

Overall, individuals showing pro-gender equality attitudes favour de-familialising measures, which foster the consolidation of the dual-earner family model and even the earner–carer one (Korpi, 2000; Pavolini et al., 2021). Those with more conservative/authoritarian attitudes prefer familialising policy instruments, which reinforce the male breadwinner model (Wood et al., 2023). Family policy reforms in a specific national context are expected to be affected by the gender norms endorsed by the most influential group. If libertarians are the majority, gender equality norms will be dominant, and de-familialising policy instruments will be expanded. On the contrary, if authoritarians are still in a powerful position, conservative gender norms remain dominant and familialising instruments

will be maintained.

Political party articulation of gender norms

Partisan politics operates under the assumption that a political party's welfare priorities are influenced by the interests of its voter base (e.g. Korpi, 1989). In the realm of family policy, this means that parties adjust their policy agenda based on the cultural inclinations of the electorate regarding gender norms (Pavolini et al., 2021). Generally speaking, on the left pole of the political spectrum, Social Democrats, the Radical Left, and the Greens usually rely on a cultural-progressive constituency favouring gender equality. Consequently, they are more inclined to support de-familialising measures such as childcare and fathers' parental leave (Häusermann, 2010, 2012).

On the right pole, the picture is more nuanced. Support for liberal parties has traditionally been rooted among libertarian electors (Häusermann, 2010). Thus, it is not surprising that the parties—especially in the Nordic countries—have historically supported de-familialising measures and the dual-earner family model (e.g., (Hiilamo and Kangas, 2009)). On the contrary, Christian democratic and conservative parties have been long associated with a traditionalist electorate. As a result, they have been said to promote familialising measures, such as cash-for-care (van Kersbergen, 2003; Leitner, 2010; Blome, 2016). However, societal modernisation has also impacted their voters in the last decade: in several Western European countries, centre-right supporters—first of all, working women—have gradually endorsed gender equality norms (Ferragina and Seeleib-Kaiser, 2015). As a consequence, in the post-Fordist era, the Christian democrats and conservatives have been increasingly incentivised to move away from a pure familistic agenda and to back de-familialising measures in order to avoid electoral punishment (Emmenegger and Manow, 2014).[1] In such a new modernised context, the preferences of authoritarian electors are increasingly often represented by the Radical Right parties, which, consequently, have adopted strict familistic stances (Giuliani, 2023; Meardi and Guardiancich, 2021). The relative strength of political parties at the subnational level can provide valuable information about the geography of gender norms and, therefore, about the individuals' current use of de-familialising policy instruments, such as fathers' parental leave. In those contexts where parties displaying strong support for de-familialisation receive a high electoral score, we can expect these instruments to be used to a great extent. On the contrary, in those areas where parties showing an explicit endorsement for familialisation obtain a high electoral success, we can expect a low uptake of the same policy instruments.

Parental leave in Finland

Finland has a generous social policy regime and a long tradition supporting the dual-earner family model. Promoting gender equality in terms of both parents having the right to both paid work and caring for their young children has been the focus of Finnish and other Nordic family policies for decades (Duvander et al., 2019; Huttunen and Eerola, 2015). As mothers started being employed at similar rates to fathers, the focus of gender equality shifted from enhancing women's employment to increasing fathers' participation in unpaid work.

[1] In Finland, the Christian Democrats are a religious-niche party with less than 5% support, mainly from fundamentalist Christians (Arter, 2022).

In the late 1970s, Finland and Norway were the first countries in the world to introduce paid paternity leave, followed by sharable parental leave soon after (Huttunen and Eerola, 2015). Despite this, the development of individual parental rights of father was slower in Finland than its Nordic counterparts, and fathers' uptake of solo parental leave has been a more recent phenomenon (Lammi-Taskula, 2017). In 2002, the leave scheme consisted of 18 weeks of paid maternity leave and 3 weeks of paid paternity leave (usually taken around birth), 43 weeks of sharable paid parental leave, and childcare leave with a low, flat-rate cash-for-care compensation available until 36 months. Most paid leave was (and still is) income-compensated at 70% up to a ceiling (around median income) and lower above that, with no upper limit. Parents with no or low income received a minimum flat-rate allowance. The eligibility to parental leave and childcare leave has been a universal benefit, with no requirements regarding employment or income, but only fathers married to or cohabiting with the mother were eligible before 2017.

Despite the sharable leave being available to fathers, it was almost always fully used by mothers (Kellokumpu, 2007). To encourage fathers to take solo parental leave, Finland followed its Nordic counterparts by introducing the first non-transferrable father's quota in 2003. At first, however, this "daddy month" was conditional on the mother's consent to share parental leave: the father got two bonus weeks after the parental leave period if they used at least two weeks of the sharable parental leave. Relatively few families took advantage of the daddy month, so a reform in 2007 increased its flexibility, allowing postponement of leave until 18 months, and another in 2010 increased the duration up to 54 days in total (3 weeks of birth-related leave and 6 weeks of solo parental leave). Still, less than a third of fathers used their quota leave before 2012. In 2013, Finnish fathers finally got an independent right to a six-week quota available until 24 months (Salmi and Närvi, 2017). In 2022, Finland introduced a fully gender-equal parental leave reform granting both (custodial) parents an independent right to 6.5 months of parental leave, of which 16 weeks is non-transferrable (Kinnunen et al., 2024).

The entitlement to early childhood education and care (ECEC) starts at nine months, although the majority of families (around 90 %) have used at least some childcare leave as an extension to parental leave (Österbacka and Räsänen, 2022). In 2015–2021, about a third of all 0–2-year-olds participated in ECEC (Statistics Finland, 2021).

During our study period, taking paternity leave was already common: For children born in 2009–2017, on average 76 percent of all fathers and 83 percent of employed fathers took at least some leave, typically the three-week paternity leave with the mother and their newborn. Taking father's quota was less typical, with 36 percent of all fathers and 40 percent of employed fathers taking father's quota, although the share was strongly increasing with the period, especially before the 2013 reform (increasing from 22% to 44%; own calculations). Fathers' share of recipients and leave days has been growing over the years, but leave use has remained gendered: about 89 percent of all parental leave days were paid to mothers in 2019 (Kela, 2020).

Hypotheses

Individual attitudes and gender norms also have a geographical dimension (Lomazzi, 2017; Uunk and Lersch, 2019). While at the national level, gender equality norms can be endorsed by the majority of the population—thus triggering de-familistic policy reforms—at the sub-national level, conservative norms could still be largely championed. As a consequence, though de-familialising instruments are de jure implemented and available

throughout the whole national territory—such as father's quota of parental leave—their uptake can substantially vary: in those territories where conservative gender norms are still dominant, the actual use of national de-familialising measures is expected to be lower. Thus, we hypothesise:

> *H1: Fathers' quota use increased faster in municipalities with more support for liberal gender norms.*

Finland has a Swedish-speaking language minority (about 5% of the population) that mainly lives in three regions: Uusimaa, Southwest Finland, and Ostrobothnia. The first two regions are bilingual, with Finnish being the dominant language in most municipalities, while Swedish is the dominant language in the Ostrobothnia region. Due to the shared language, the Swedish-speaking population has stronger cultural ties to Sweden and individuals feel strong affinity with the neighbouring country, especially in the Ostrobothnia region (Lindell, 2020). As Sweden has been the forerunner in gender-equality promoting parental leave, with high shares of fathers taking leave (Duvander et al., 2019), the Swedish-speaking population in Finland may have been acculturated to paternal leave-taking earlier than the Finnish-speaking population. We thus assume that the policy feedback effect—according to which welfare institutions shape citizens' welfare policy preferences (Svallfors, 1997; Pierson, 2001; Jordan, 2013)—triggered by the Swedish father's quotas has indirectly "crossed the border", therefore affecting fathers' leave uptake amongst the Swedish-speaking minority in Finland. Hence, we hypothesise:

> *H2: Fathers' quota use increased faster in municipalities with more Swedish-speaking residents.*

Finally, neighbouring municipalities are likely to share a common cultural heritage, including cultural norms. If care practices start shifting in one area, we expect it has a spill-over effect on the neighbouring area (e.g., Bertrand et al., 2000; Bobba and Gignoux, 2019) when families in close geographical proximity interact and share information. As such, we hypothesise:

> *H3: Fathers' quota use varied spatially so that uptake was more similar in neighbouring municipalities, even after accounting for their socioeconomic and cultural characteristics.*

Data and methods

Party Manifesto content analysis and family policy positions in Finland

Following the established tradition in comparative politics and comparative welfare state literature, to assess Finnish parties' family policy positions, we performed a content analysis of their manifestos of political programs issued during the 2019 elections (Budge, 2001; Enggist and Pinggera, 2021; Giuliani, 2024).

To identify the extent to which political parties support familialism and de-familialism, we recoded the data as quasi-sentences from the Comparative Manifesto Project Database (CMP). The quasi-sentences were assigned to three categories: Domain A, "Familialising Policy Instruments"; Domain B, "De-Familialising Policy Instruments"; and Domain C,

"Ambiguous Policy Instruments". We also coded whether the sentiment was positive (expanding) or negative (retrenching).

The quantitative results of the content analysis were used to construct a De-Familialisation/Familialisation Party Index (see Section A of Supplementary material, available on https://github.com/tihepasa/paternal_leaves). We plotted the values of the two indexes in a Cartesian graph, where it is possible to see the combined positions of the political parties (Figure 1).

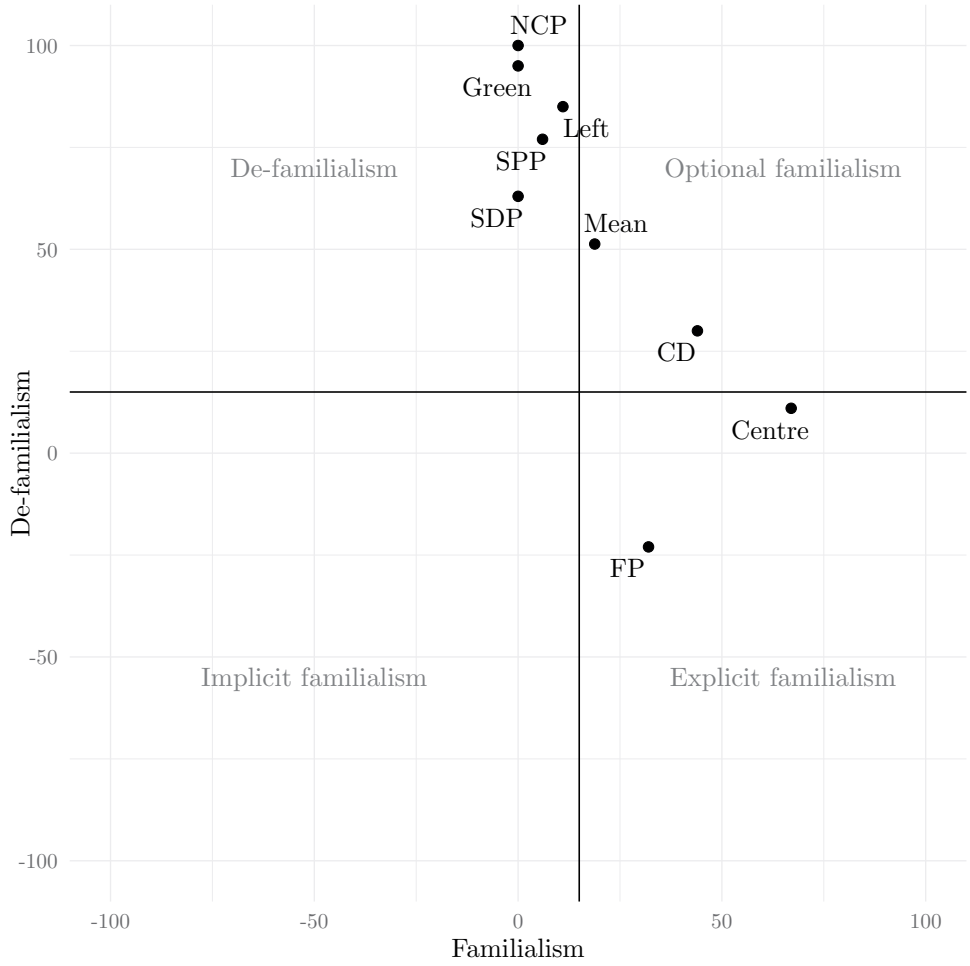

Figure 1: Parties' combined position in the multi-dimension of political conflict. In order to avoid bias in the calculation of the final positions, two threshold values were identified. Familialisation and De-familialisation Index values ranging from 0 to 15 respectively show very low support for familialising and de-familialising policy instruments and therefore are treated as neutral positions. The black horizontal and and vertical lines mark these thresholds. Source: Authors' own elaboration on Comparative Manifesto Project Data.

Overall, the Finnish party system in 2019 is biased toward support for de-familialising policy instruments. Indeed, all the Left/Green (Social Democrats, Left Alliance, Green Union) and Liberal (Swedish People's Party) parties are located in the "De-familialism" quadrant. In other words, they show low or no support for familialising measures and a moderate–high backing for de-familialising measures. Even the main centre-right party, the National Coalition, is located in this quadrant.

The Christian Democrats in Finland, on the other hand, are positioned in the "optional familialism" quadrant: the party both endorses familialising and de-familialising

measures. However, its position is more biased toward the former ones. Finally, the Finnish Centre (a former agrarian party) and the Finns Party (former True Finns; radical right) are located in the "explicit familialism" quadrant, indicating that they champion a familistic agenda. In particular, the Finns Party strongly oppose de-familialising measures and promote the male breadwinner family model.

We can draw two conclusions from the analysis of the Finnish party manifestos. First, at the national level, there is a sort of "de-familialism" consensus within the Finnish party system, suggesting that gender equality norms are dominant and that libertarian citizens are the majority. However, the political strength of familistic parties is not irrelevant: altogether, the Christian Democrats, the Finnish Centre, and the Finns Party obtained 35% of the votes. As a result, the total weighted mean falls just within the optional familialism quadrant. This suggests that conservative gender norms are still spread in the country—co-existing with the progressive ones—and that authoritarian citizens are a powerful minority. We expect that in those territories where the Christian Democrats, the Finnish Centre, and the Finns Party have obtained solid electoral success, the uptake of paternal leaves will be lower.

Administrative data

Our spatio-temporal data have been obtained from registers maintained by Statistics Finland and the Social Insurance Institution of Finland, Kela. The data consist of several variables describing the Finnish population on municipality level, based on the availability of parental leave data between 2009 and 2019. During this time period there were two legislative changes related to the parental leaves that entered into force, one in 2010 and another in 2013. All the data are presented on a yearly level using the municipal division as of 2023. The analysis considers only mainland Finland, excluding the self-governing autonomous Åland Islands—this results in 293 municipalities in total.

We used three different sources to prepare our data. First, the information about parental leave uptakes (share and number of fathers using father's quota, i.e., more than the three-week paternity leave), total number of fathers, and proportion of entrepreneurs were derived from licensed data modules from Statistics Finland and the Finnish Health Institute (THL). The averages over time of the share of fathers taking father's quota are visualised in Figure 3. The yearly maps of the original values can be seen in Supplementary Figure 6. Secondly, the spatial information about the proportions of registered mother tongues (variable code name in the database is 11ra), of low income (127y), and of unemployed (12tf), as well as the educational level (12bs) and the party support shares (13sw) are gained from the open database of Statistics Finland. Also all these variables are reported on a yearly level with the exception of the party support information which is based on the votes in the Parliamentary elections typically occurring every fourth year. Thirdly, the map of the municipalities is retrieved via R library geofi (Kainu et al., 2023) that returns the areal division used by Statistics Finland in their datasets.

Most of the data are available in the Supplementary material. However, due to privacy reasons we are unable to publish data on 2637 year-municipality combinations with small cell sizes (fewer than three quota users or non-users; about 9.5%). We conducted the main analysis using the full, uncensored data in a secure environment but also show replication codes and results using the censored data in the Supplementary material. The variables and descriptive statistics, apart from the voting statistics, can be found in Supplementary Table 7.

Constructing (de)familialisation scores for municipalities

Based on the Parliamentary election information, we know the proportions of votes for each party in each municipality in each election. The variables describing support to (de)familialising policies in given municipalities are derived from the party shares. To gain more robust approximations that are less sensitive to electoral volatilities, the shares are averaged locally over the years 2007, 2011, 2015 and 2019. The familialisation and de-familialisation scores in each municipality are calculated from the vote shares according to the transformation coefficients in Giuliani (2023) and supplemented with the coefficients for the Christian Democrats (derived using the same methodology). Explicitly, we have

$$\begin{aligned}\text{familialisation score} =& 32 \cdot \text{FP} + 0 \cdot \text{SDP} + 0 \cdot \text{NCP} + 67 \cdot \text{Centre} + 0 \cdot \text{Green} \\ &+ 11 \cdot \text{Left} + 44 \cdot \text{CD}, \\ \text{de-familialisation score} =& -23 \cdot \text{FP} + 63 \cdot \text{SDP} + 100 \cdot \text{NCP} + 11 \cdot \text{Centre} + 95 \cdot \text{Green} \\ &+ 85 \cdot \text{Left} + 30 \cdot \text{CD},\end{aligned}$$

where the letter abbreviations denote the average percentage of election votes for each party: FP = Finns Party (Perussuomalaiset in Finnish), SDP = Social Democratic Party (Suomen Sosialidemokraattinen Puolue), NCP = National Coalition Party (Kansallinen Kokoomus), Centre = Centre Party (Suomen Keskusta), Green = Green League (Vihreä liitto), Left = Left Alliance (Vasemmistoliitto), and CD = Christian Democrats (Suomen Kristillisdemokraatit). Votes for the Swedish People's Party (SPP) were not accounted for in the main analyses, despite being the majority party in several municipalities, as SPP supporters generally do not make their voting decisions based on culturally-oriented considerations but linguistic identity (see Section A in Supplementary material, and Supplementary Table 8 for descriptive statistics on (de)familialisation scores with and without the SPP votes).

Further inspection revealed that familialisation and de-familialisation scores have a strong correlation in the Finnish case ($\rho = -0.95$). Thus, we performed a principal component analysis via singular value decomposition on these two variables to gain uncorrelated variables. The original familialisation and de-familialisation scores, as well as the principal components, are illustrated in Figure 2.

The first principal component may be interpreted as a scale from explicit familialism to de-familialism (with higher values indicating higher support to de-familialising policies and lower support to familialising policies). This component alone explains 97.5% of the variance in the familialisation and de-familialisation scores. The second principal component can be interpreted as the left-over variation in the direction between optional familialism and implicit familialism (with higher values indicating higher preference toward the latter), and it explains the remaining 2.5% of the variation, i.e., very little.

Descriptive statistics

Figure 3 illustrates the shares of fathers' quota use and Swedish-speaking residents and the two principal components, showing an average across the study period. Overall, the fathers used quota more in and around big cities (especially the capital region, Turku, and Tampere) and in the Swedish-dominant Ostrobothnia region. Uptake was the lowest in rural areas further away from bigger cities. See Section B of Supplementary material for annual descriptive statistics.

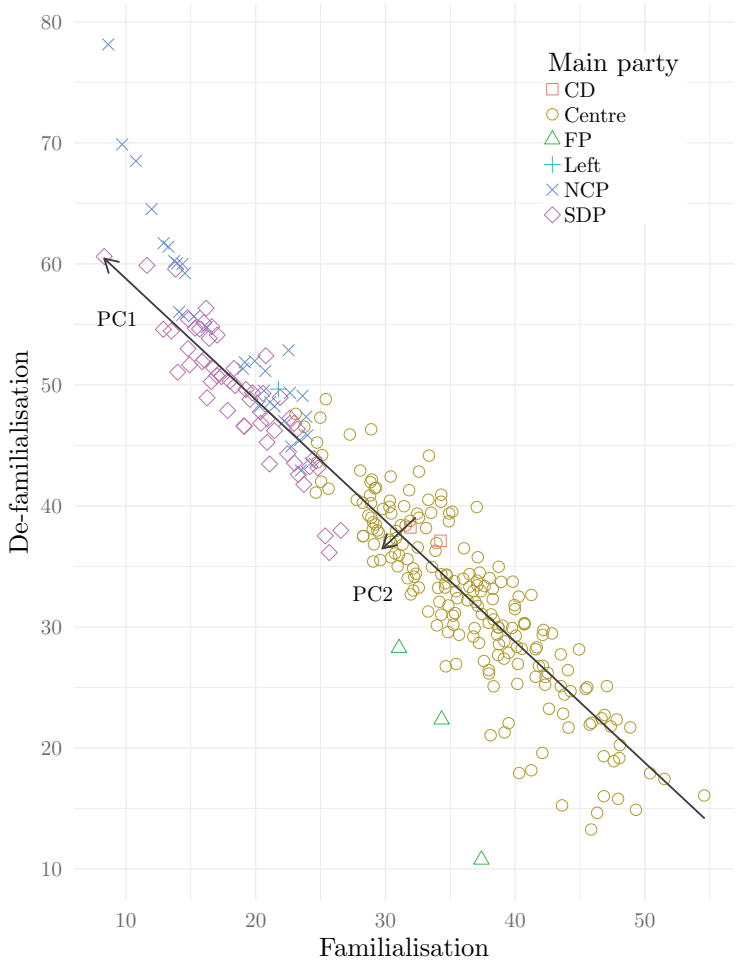

Figure 2: Scatter plot of the local familialisation and de-familialisation scores, and the principal component axes derived from them as black arrows. The length of the arrow reflects the share of variation in the scores the component captures. The appearance of the score observation indicates the main party in the municipality.

Regarding the first principal component, PC1, liberal municipalities with high preference for de-familialism (marked with yellow) were located in and around the biggest cities, while higher preference for explicit familialism (marked with dark blue) was predominately found in the smaller municipalities in the Finnish-dominant regions of South, Central, and North Ostrobothnia. Regarding the second principal component, PC2, which differentiates municipalities in their preference between optional familialism and implicit familialism, the map shows small differences between Lapland in the north (more optional familialism) and the rest of the Finland (less optional familialism), but overall the differences are negligible in comparison to the first principal component.

Model

The interest here is to model the uptake of father's quota taking into account the location and time. We denote each location with $i = 1, \ldots, N$ and each time point with $t = 1, \ldots, T$, where N and T denote the numbers of municipalities and years, respectively. As a response variable, $y_{i,t}$, we use the amount of fathers who have taken father's quota

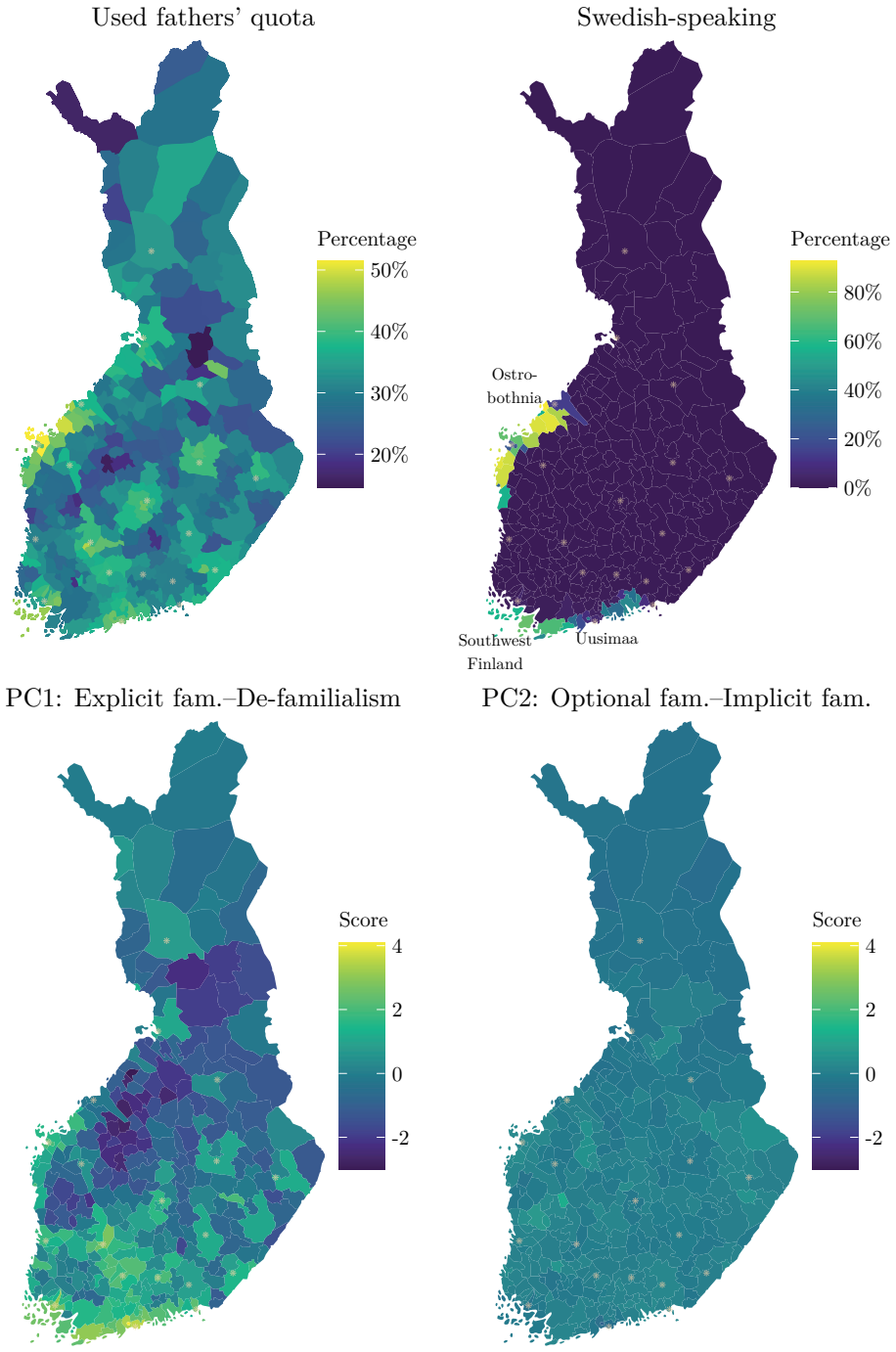

Figure 3: Descriptive statistics of the main variables of interest. The top panel shows municipal-level averages of the shares of fathers' quota use for child cohorts 2009–2017 and the shares of Swedish-speaking residents in 2009–2017. The lower panel shows principal components based on familialisation and de-familialisation scores (see the main text for more information). White spots show the locations of the main cities in each region.

of parental leave. They are modelled with binomial distribution and inverse logit link:

$$y_{i,t} \sim \text{Bin}(n_{i,t},\ p_{i,t}), \tag{1}$$

where the first parameter of the distribution, $n_{i,t}$, denotes the total number of fathers in the municipality i during the year t, and the second parameter denotes the probability

of taking father's quota in that municipality and year. The second parameter, or the probability of quota use, is our main interest and tool for answering the research questions, and it is defined as

$$p_{i,t} = \text{logit}^{-1}\left(c_t + \sigma_\varphi \varphi_{i,t} + \sum_{k=1}^{K} \beta_{k,t} x_{i,k,t}\right), \tag{2}$$

where K is the number of explanatory variables. This probability consist of three parts. First of all, there is the time dependent nationwide constant c_t, describing the average probability of taking father's quota. This parameter is constructed as a random walk $c_t \sim \text{N}(c_{t-1},\ \sigma_c^2)$.

Secondly, the spatial effect that cannot be explained with other included variables is captured by the term $\sigma_\varphi \varphi_{i,t}$. We assign an ICAR (Besag, 1974) structure for this component so that municipalities close to each other have more similar effects than those far from each other. As ICAR requires the knowledge of the neighbourhood of each site, we define that municipalities sharing borderline are neighbours. Based on this neighbourhood structure, we define a matrix W indicating the neighbourhood such that

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

and a diagonal matrix D collecting the numbers of neighbours of each location i as

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

Then, the ICAR structure for the spatial random variables at each time point can be written as $\boldsymbol{\varphi}_t \sim \text{N}(0, (D - W)^{-1})$. The symbol in bold denotes a vector of length N. This results in having for each year spatially dependent variables that are independent between the years.

Finally, the probability of taking father's quota is explained with a general regression term using the covariates available. The regression coefficients are let to depend on time monotonically in a similar way as in Bürkner and Charpentier (2020), so that the effect is decreasing or increasing across the years but it cannot alter between these during the time window. This is done with helper vectors $\boldsymbol{b}_k$ of length T, with the first element fixed to zero and the rest summing up to one, thus constructing a simplex for each covariate k. The regression coefficients at the first and the last time point, $\beta_{k,1}$ and $\beta_{k,T}$, are defined separately, in Bayesian context we set priors for them individually. Using the helper simplexes, the coefficients at the other time points are defined as

$$\beta_{k,t} = \beta_{k,1} + (\beta_{k,T} - \beta_{k,1}) \sum_{i=1}^{K} b_i$$

for all $k = 1, \ldots, K$, and $t = 2, \ldots, T-1$. In our case, all the explanatory variables are standardized based on their mean and standard deviation on each year individually. This facilitates the comparison between the covariates but also has to be taken into consideration when interpreting the results.

The model is fitted via Bayesian approach, which requires setting prior distributions

for the unknown parameters. The full listing of the priors we use is

$$
\begin{aligned}
c_1 &\sim \mathrm{N}(0,\ 1^2), \\
c_t &\sim \mathrm{N}(c_{t-1},\ \sigma_c^2),\ \text{for all } t = 2, \ldots, T, \\
\sigma_c &\sim \mathrm{N}(0,\ 1^2)[0,], \\
\sigma_\varphi &\sim \mathrm{N}(0,\ 1^2)[0,], \\
\boldsymbol{\varphi}_t &\sim \mathrm{N}(0, (D - W)^{-1}),\ \text{given } \sum_{i=1}^{N} \varphi_{i,t} = 0,\ \text{for all } t = 2, \ldots, T, \\
\beta_{k,1} &\sim \mathrm{N}(0,\ 1^2),\ \text{for all } k = 1, \ldots, K, \\
\beta_{k,T} &\sim \mathrm{N}(0,\ 1^2),\ \text{for all } k = 1, \ldots, K, \\
\boldsymbol{b}_{k,2:T} &\sim \mathrm{Dirichlet}(1_1, \ldots, 1_{T-1}),\ \text{for all } k = 1, \ldots, K,
\end{aligned}
\tag{3}
$$

where $\mathrm{N}(.,.)[x,]$ indicates that the distribution is left-truncated at x and the indexation $\boldsymbol{b}_{k,2:T}$ denotes the components from 2 to T of the vector $\boldsymbol{b}_k$. These priors are mostly weakly informative and merely set to enhance the computation (see Banner et al., 2020).

The model was estimated with Markov chain Monte Carlo (MCMC) using rstan (Stan Development Team, 2023), which is an R interface (R Core Team, 2023) for the probabilistic programming language Stan for statistical inference (Stan Development Team, 2024). The posterior samples were drawn using the NUTS sampler (Hoffman and Gelman, 2014; Betancourt, 2018) with four chains. Each chain consisted of 8,000 iterations, of which the first 1,500 were discarded as a warm-up. The estimation was done on Statistics Finland's Fiona remote access system, equipped with 2.3 GHz Intel Icelake processors and 64 GB of RAM. The estimation took approximately 30 minutes using four parallel chains. According to the MCMC diagnostics available in rstan (Vehtari et al., 2021), the model converged appropriately. All the $\hat{R}$ statistics were less than 1.005 and the effective sample sizes were approximately between 1,200 and 27,000.

Results

The upper panel of Figure 4 illustrates the nationwide regression coefficients of the local Swedish-speaking proportion and the principal component values describing the scales from explicit familialism to de-familialism and from optional familialism to implicit familialism. The lower panel shows the effects of changing the covariate values on the predicted probability of taking the fathers' quota, calculated as average marginal predictions.

For the two principal components, we compared the upper and lower quartiles of the observed covariates. The effects of the first principal component PC1 (explicit familialism–de-familialism) and the second PC2 (optional familialism–implicit familialism) were decreasing throughout the study period. The effect of PC1 decreased abruptly between 2012 and 2013, which coincides with the 2013 policy reform. The effect of PC2 is slightly decreasing but very small. Due to having very little variation in the data (see Figure 3), this dimension does not have any practical significance at the population level. Overall, we get partial support for hypothesis H1: Municipalities with higher support to de-familialism have a greater probability of using the fathers' quota, but only before the 2013 parental leave reform.

In the case of the Swedish-speaking population, we compared cases of 50 percent Swedish-speaking residents (approximately the proportion of Swedish-speaking residents

in the Ostrobothnia region) and 0.2 percent, which is the median of the Swedish-speaking residents across all municipalities. The effect of having more Swedish-speaking residents increased over time, suggesting a multiplicative effect of supportive policies and acculturation to paternal leave-taking and giving support to hypothesis H2.

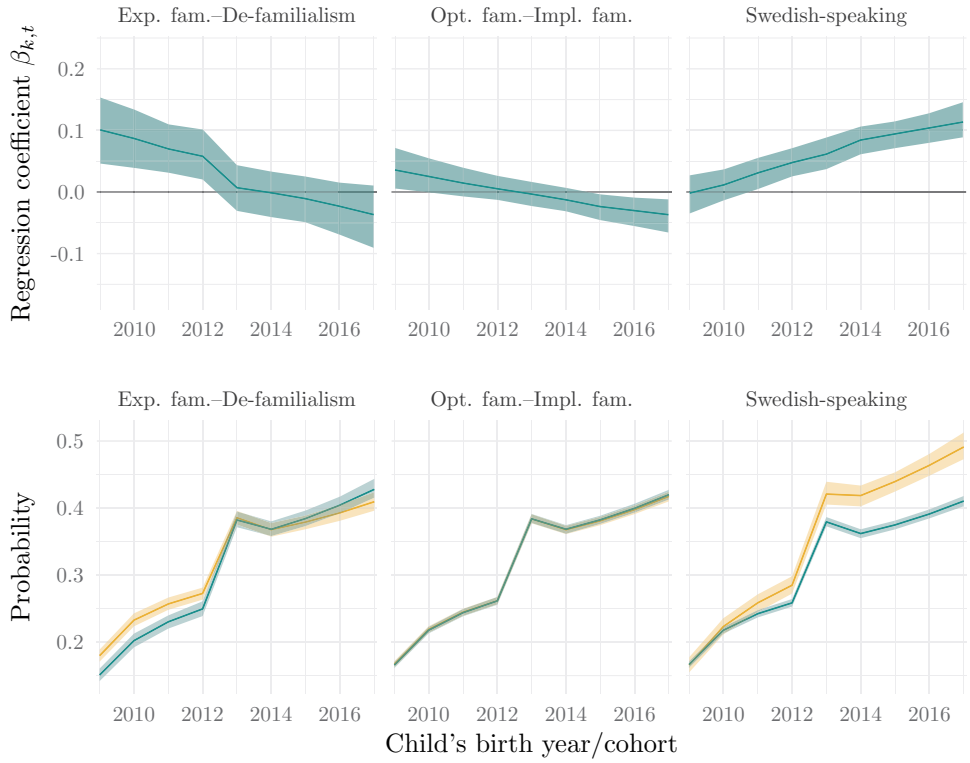

Figure 4: Regression coefficients (top) and predicted probabilities (bottom) for the proportion of fathers taking father's quota of parental leave in Finnish municipalities for child cohorts 2009–2017. In the upper panel, the dark lines show the posterior mean (with 95% posterior interval as shaded area) of regression coefficients $\beta_{k,t}$ for the two principal components and the proportion of Swedish-speaking residents. In the lower panel, the lines depict the average marginal probabilities of taking the fathers' quota comparing the effect of changing the values of the covariate being considered, and the shaded areas their 95% posterior intervals. The yellow lines represent the cases of higher values of the covariate (50% for Swedish, upper quartile for the other two) and turquoise the cases of lower values (median for Swedish and lower quartile for the other two).

The spatial components $\varphi_{i,t}$, describing the spatial variation not captured by the covariates, are illustrated in Figure 5. There seem to be clear spatial dependencies between neighbouring municipalities since the coefficients of adjacent areas are more similar than those of areas far apart. Overall, we find support to H3 about spill-over influences from neighbouring municipalities.

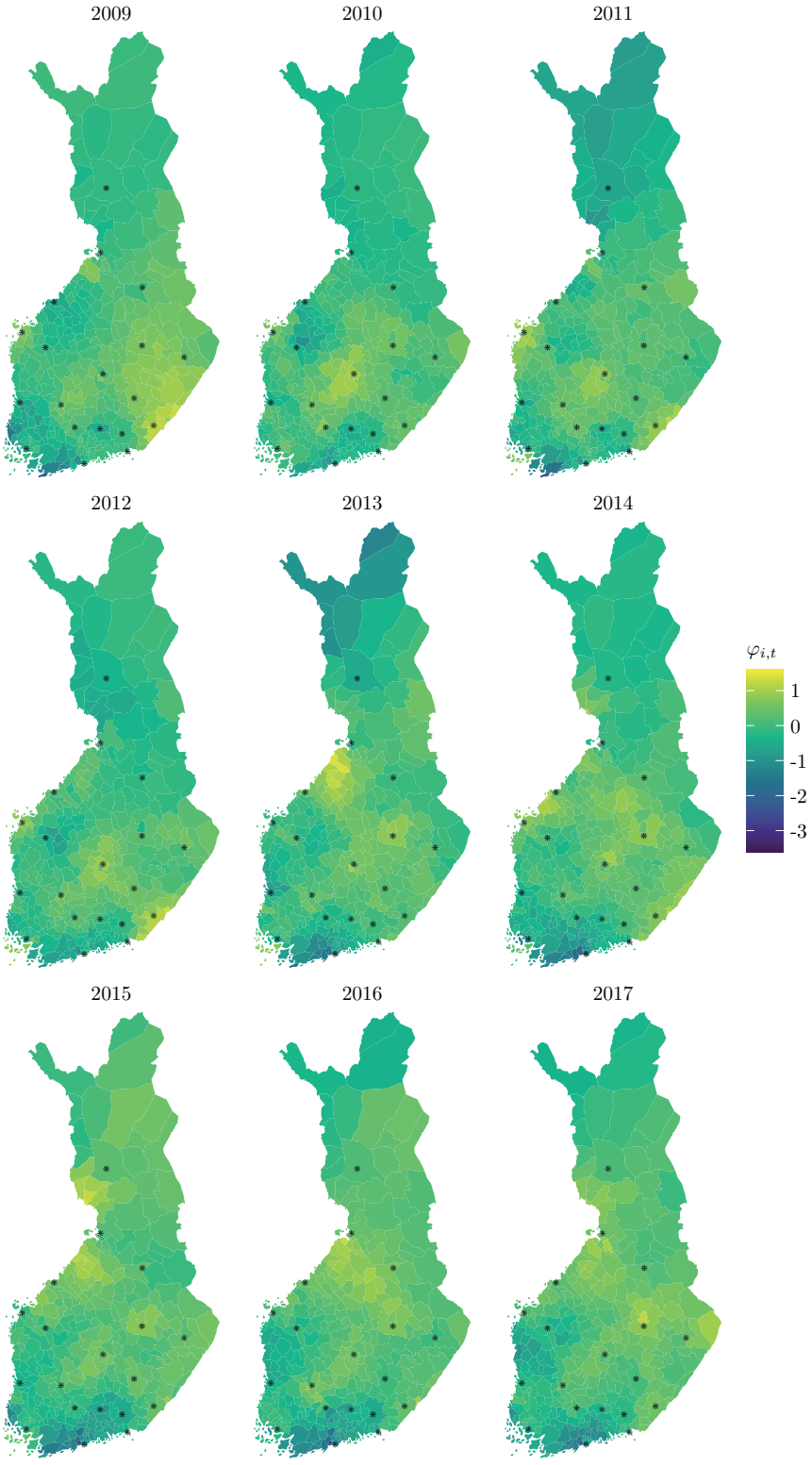

Figure 5: The posterior means of the spatial variables $\varphi_{i,t}$ showing the spatial dependency not explained by the covariates.

Predictions from the model correspond well to the observations (see Section C of the Supplementary material, which also includes detailed results of all the model parameters). We also estimated a number of additional models with differing specifications and conducted model comparisons. Due to resource limitations of the secure environment,

these were conducted using the censored data. Based on the comparison of the full data and censored data, both resulted in similar results. In order to verify the importance of the spatial component $\varphi_{i,t}$, we made comparisons between our main model and another model without the spatial component, based on which we determined that the model with the spatial component fits the data better. We also contrasted the binomial model with the beta-binomial model for accounting for potential overdispersion, but this did not improve the model fit, and the main conclusions from both distributional assumptions were identical. Further details about the model comparisons can be found from Section D of the Supplementary material.

We also tested how robust our results were to the exclusion of the SPP votes. This alternative specification resulted in a model with qualitatively similar conclusions but the effects were slightly smaller in magnitude for the main variables of interest. The results can be found from the Supplementary material Section D.

Discussion

The present work has examined to what extent gender norms at the local level and cross-border policy influences from Sweden were associated with fathers' parental leave uptake in Finnish municipalities in the 2010s—around the time when a policy reform provided fathers with an independent right to a six-week quota of "solo" parental leave. We formulated three interconnected hypotheses. First, we expected fathers' quota use to increase faster in municipalities with more support for liberal gender norms. Second, we hypothesised that the policy feedback effect triggered by the Swedish leave policy has indirectly "crossed the border": we thus expected fathers' quota use to rise speedier in municipalities with more Swedish-speaking residents, as they are more exposed to the influences of the Swedish leave practices through media. Third, we assumed that fathers' quota use varied spatially so that uptake was more similar in neighbouring municipalities, since we expected that families in close geographical proximity interacted and were more likely to adopt new practices when seeing them being used by others.

By using administrative data from Finnish municipalities and familialisation/de-familialisation indexes from a content analysis of Finnish party manifestos for approximating local gender norms, three main findings emerge from the empirical analysis. First, local support for de-familialising policies—which denotes more liberal attitudes concerning gender norms—predicted higher quota use, but only before the 2013 reform, i.e. when the father's bonus leave was conditional on the mother shortening her leave. After the 2013 reform, when fathers got an independent right to a six-week quota of parental leave, local gender norms lost importance. We can explain this result in two alternative ways: either the 2013 reform was effective in equalising leave-taking between families across the country or local gender norms may be more important in affecting the duration of maternal leave-taking than whether fathers take leave—at least during the time when father's quota was measured in weeks rather than months. Therefore, our first hypothesis was only partially confirmed: gender norms do likely affect fathers' parental leave uptake, but such an effect seems to decrease due to effective changes in the institutional arrangements. This result triggers new theoretical considerations concerning the role of gender norms in affecting citizens' behaviour in the family policy field and how these norms interact with policy reforms. Future research can explore these interaction dynamics at the theoretical and empirical levels.

Second, a higher proportion of Swedish-speaking residents was found to be increasingly

important factor over the years, especially after the 2013 reform. Our second hypothesis was thus confirmed. We interpret this result as a sign of cross-border policy influence from Sweden: the combination of supportive policies and exposure to behavioural changes seem to produce a multiplicative effect boosting paternal leave-taking. Such results suggest that foreign national policy feedback effects can cross the border, heading to those close regions inhabited by culturally and linguistically-related minorities, and thus can contribute to shaping these populations' preferences even before a similar policy is enacted at the national level. This finding triggers new theoretical considerations regarding the role of national institutions in affecting welfare state preferences at the local level, thus potentially sparking new avenues of research within the literature on the comparative welfare state and Historical Institutionalism.

Interestingly, the region where uptake increased the fastest is Ostrobothnia, which is predominately Swedish-speaking but is also the region with the most support for conservative values among the Swedish-speaking population (Lindell, 2020). Within the Swedish-speaking population, residents of Ostrobothnia feel more affinity with Sweden (Lindell, 2020) and consume Sweden-Swedish media more often than Swedish speakers in other regions of Finland (see Section A of Supplementary material). This finding of a conservative region being the driver of paternal leave-taking suggests that in some circumstances, family policy preference formation is more institutionally-driven than culturally-driven. Furthermore, it raises the question of whether (relatively) more conservative families are inherently more resistant to leave-sharing, or simply tend to be slower to adapt, e.g. due to observing fewer examples of cultural change or due to facing more barriers.

Third, there are spatial dependencies between neighbouring municipalities, which supports our "spillover hypothesis": the interactions among families nearby likely lead them to adopt new practices. This result stresses the importance of community socialisation, and new research could better investigate the dynamics of the spill-over effect, especially in the most deprived areas.

To conclude, fathers' parental leave uptake is a highly timely topic, as, since 2022, all European Union member states have been compelled to offer at least four months of parental leave to each parent, of which at least two must be non-transferable (Directive 2019/1158). Across Europe, a considerable proportion of parental leave is now reserved for fathers (or social parents) on a "use-it-or-lose-it" basis (Blum et al., 2023), even though several countries—especially in Southern and Eastern Europe—continue to lag behind in the uptake of paternal leaves. Unveiling the factors affecting the variation in fathers' parental leave uptake and adopting a local-level perspective thus appear increasingly important in order to analyse the effectiveness of policy reforms.

Acknowledgements

We thank Statistics Finland and THL for providing access to parental leave data (decisions TK/2182/07.03.00/2024 and THL/3141/6.02.00/2022). This work was funded by the Research Council of Finland (decision numbers 331816, 331817, 355153, 345546), the Strategic Research Council (364374), and Emil Aaltonen Foundation. We also wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

We thank Dr Julia Hellstrand, who provided valuable information regarding the Swedish-speaking culture in Ostrobothnia.

Data availability

Data are collected from different administrative registers and are available in Supplementary Material on https://github.com/tihepasa/paternal_leaves, where also all the model codes and other supplementary material can be found. Local parental leave statistics are censored for municipalities that had fewer than three eligible quota-users or non-users in a given year.

References

André, S., Gesthuizen, M., and Scheepers, P. (2013). Support for traditional female roles across 32 countries: Female labour market participation, policy models and gender differences. *Comparative Sociology*, 12(4):447–476.

Arter, D. (2022). Can a religious-niche party change—or was Kirchheimer right? Analysing the Finnish Christians' search to become a catchall electoral party. *European Politics and Society*, 25:115–131.

Banner, K. M., Irvine, K. M., and Rodhouse, T. J. (2020). The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods in Ecology and Evolution*, 11(8):882–889.

Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv*. [Referenced: 11.10.2024].

Bicchieri, C., Muldoon, R., and Sontuoso, A. (2023). Social norms. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Blome, A. (2016). *The Politics of Work-Family Policy Reforms in Germany and Italy*. Routledge.

Blum, S., Dobrotić, I., Kaufman, G., Koslowski, A., and Moss, P. (2023). 19th international review of leave policies and related research 2023. Available at: http://www.leavenetwork.org/lp_and_r_reports/.

Bobba, M. and Gignoux, J. (2019). Neighborhood effects in integrated social policies. *The World Bank Economic Review*, 33(1):116–139.

Budge, I. (2001). *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*. Oxford University Press.

Bürkner, P.-C. and Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, 73(3):420–451.

Ciccia, R. and Verloo, M. (2012). Parental leave regulations and the persistence of the male breadwinner model: Using fuzzy-set ideal type analysis to assess gender equality in an enlarged Europe. *Journal of European Social Policy*, 22(5):507–528.

Daly, M. (2020). Conceptualizing and analyzing family policy and how it is changing. In Nieuwenhuis, R. and Van Lancker, W., editors, *The Palgrave Handbook of Family Policy*, pages 25–41. Springer International Publishing, Cham.

Dubois, N. (2002). Autour de la norme sociale. *Cahiers de Psychologie Politique*, 2(2).

Duvander, A.-Z. and Ellingsæter, A. L. (2016). Cash for childcare schemes in the Nordic welfare states: diverse paths, diverse outcomes. *European Societies*, 18(1):70–90.

Duvander, A.-Z., Eydal, G. B., Brandth, B., Gíslason, I. V., Lammi-Taskula, J., and Rostgaard, T. (2019). Gender equality: Parental leave design and evaluating its effects on fathers' participation. In Moss, P., Duvander, A.-Z., and Koslowski, A., editors, *Parental Leave and Beyond*, pages 187–204. Policy Press.

Emmenegger, P. and Manow, P. (2014). Religion and the gender vote gap: Women's changed political preferences from the 1970s to 2010. *Politics & Society*, 42(2):166–193.

Enggist, M. and Pinggera, M. (2021). Radical right parties and their welfare state stances–not so blurry after all? *West European Politics*, 45(1):102–128.

Eydal, G. B. and Rostgaard, T. (2023). Childcare by fathers in the context of active father-oriented policies. In Daly, M., Pfau-Effinger, B., Gilbert, N., and Besharov, D. J., editors, *The Oxford Handbook of Family Policy Over The Life Course*. Oxford University Press.

Fernández-Cornejo, J. A., Escot, L., Del-Pozo, E., and Castellanos-Serrano, C. (2016). Do fathers who took childbirth leave become more involved in their children's care? The case of Spain. *Journal of Comparative Family Studies*, 47(2):169–191.

Ferragina, E. (2019). Does family policy influence women's employment?: Reviewing the evidence in the field. *Political Studies Review*, 17(1):65–80.

Ferragina, E. and Seeleib-Kaiser, M. (2015). Determinants of a silent (r)evolution: Understanding the expansion of family policy in rich OECD countries. *Social Politics: International Studies in Gender, State & Society*, 22(1):1–37.

Giuliani, G. A. (2022). The family policy positions of conservative parties: A farewell to the male-breadwinner family model? *European Journal of Political Research*, 61(3):678–698.

Giuliani, G. A. (2023). Investigating the radical right's family policy agenda: Evidence from six European countries. *Italian Political Science Review/Rivista Italiana di Scienza Politica*, 53(2):179–200.

Giuliani, G. A. (2024). *The Mainstream Right and Family Policy Agendas in the Post-Fordist Age*. Emerald Publishing Limited.

Hiilamo, H. and Kangas, O. (2009). Trap for women or freedom to choose? The struggle over cash for child care schemes in Finland and Sweden. *Journal of Social Policy*, 38(3):457–475.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Huttunen, J. and Eerola, P. (2015). Finland. In Lenz, K. and Adler, M. A., editors, *Father Involvement in the Early Years: An International Comparison of Policy and Practice*, pages 29–60. Bristol University Press.

Häusermann, S. (2010). *The Politics of Welfare State Reform in Continental Europe: Modernization in Hard Times*. Cambridge Studies in Comparative Politics. Cambridge University Press, Cambridge.

Häusermann, S. (2012). The politics of old and new social policies. In Bonoli, G. and Natali, D., editors, *The Politics of the New Welfare State*. Oxford University Press.

Häusermann, S. (2018). The multidimensional politics of social investment in conservative welfare regimes: family policy reform between social transfers and social investment. *Journal of European Public Policy*, 25(6):862–877.

Inglehart, R. (1997). Modernization, postmodernization and changing perceptions of risk. *International Review of Sociology*, 7(3):449–459.

Inglehart, R. (2008). Changing values among western publics from 1970 to 2006. *West European Politics*, 31(1-2):130–146.

Inglehart, R. (2018). *Cultural Evolution*. Cambridge University Press, Cambridge.

Inglehart, R. and Norris, P. (2003). *Rising Tide: Gender Equality and Cultural Change Around the World*. Cambridge University Press.

Jenson, J. and Jacobzone, S. (2000). Care allowances for the frail elderly and their impact on women care-givers. OECD Labour Market and Social Policy Occasional Papers 41, OECD Publishing.

Jordan, J. (2013). Policy feedback and support for the welfare state. *Journal of European Social Policy*, 23(2):134–148.

Kainu, M., Lehtomaki, J., Parkkinen, J., Miettinen, J., Kantanen, P., Vesanen, S., and Lahti, L. (2023). *geofi: Access Finnish Geospatial Data*. R package version 1.0.8.

Kangas, O. and Rostgaard, T. (2007). Preferences or institutions? Work–family life opportunities in seven European countries. *Journal of European Social Policy*, 17(3):240–256.

Kela (2020). Kelan lapsiperhe-etuustilasto 2019 | FPA-statistik, Förmåner till barnfamiljer 2019. Suomen virallinen tilasto, Kela, Helsinki.

Kellokumpu, J. (2007). Perhevapaiden kehitys 1990–2005: Isillä päärooli uudistuksissa, sivurooli käyttäjinä. *Palkansaajien tutkimuslaitoksen raportteja, 10.*

Kinnunen, A., Lammi-Taskula, J., Miettinen, A., Närvi, J., and Saarikallio-Torp, M. (2024). *Perhevapaat ja työn ja perheen yhteensovittaminen muuttuvassa työelämässä*. Kela, Helsinki.

Korpi, W. (1989). Power, politics, and state autonomy in the development of social citizenship: Social rights during sickness in eighteen OECD countries since 1930. *American Sociological Review*, 54(3):309.

Korpi, W. (2000). Faces of inequality: Gender, class, and patterns of inequalities in different types of welfare states. *Social Politics: International Studies in Gender, State & Society*, 7(2):127–191.

Koslowski, A., Duvander, A.-Z., and Moss, P. (2019). Parental Leave and beyond: Recent international developments, current issues and future directions. In Koslowski, A., Duvander, A.-Z., and Moss, P., editors, *Parental Leave and Beyond: Recent International Developments, Current Issues and Future Directions*, pages 353–370. Bristol University Press.

Kotsadam, A. and Finseraas, H. (2011). The state intervenes in the battle of the sexes: Causal effects of paternity leave. *Social Science Research*, 40(6):1611–1622.

Lammi-Taskula, J. (2017). Fathers on leave alone in finland: Negotiations and lived experiences. In O'Brien, M. and Wall, K., editors, *Comparative Perspectives on Work-Life Balance and Gender Equality*, volume 6, pages 89–106. Springer International Publishing, Cham.

Leitner, S. (2003). Varieties of familialism: The caring function of the family in comparative perspective. *European Societies*, 5(4):353–375.

Leitner, S. (2010). Germany outpaces Austria in childcare policy: The historical contingencies of 'conservative' childcare policy. *Journal of European Social Policy*, 20(5):456–467.

Lewis, J. (2009). *Work-Family Balance, Gender and Policy*. Edward Elgar.

Lindell, M. (2020). Vad tycker Svenskfinland? Technical report, Institutet för samhällsforskning (Samforsk), Åbo Akademi University.

Lomazzi, V. (2017). Gender role attitudes in Italy: 1988–2008. a path-dependency story of traditionalism. *European Societies*, 19(4):370–395.

Lomazzi, V. and Seddig, D. (2020). Gender role attitudes in the international social survey programme: Cross-national comparability and relationships to cultural values. *Cross-Cultural Research*, 54(4):398–431.

Meardi, G. and Guardiancich, I. (2021). Back to the familialist future: the rise of social policy for ruling populist radical right parties in Italy and Poland. *West European Politics*, 45(1):129–153.

Misra, J., Budig, M. J., and Moller, S. (2007). Reconciliation policies and the effects of motherhood on employment, earnings and poverty. *Journal of Comparative Policy Analysis: Research and Practice*, 9(2):135–155.

O'Brien, M. and Wall, K. (2017). *Comparative Perspectives on Work-Life Balance and Gender Equality: Fathers on Leave Alone*. Life Course Research and Social Policies. Springer Nature, Cham.

Omidakhsh, N., Sprague, A., and Heymann, J. (2020). Dismantling restrictive gender norms: Can better designed paternal leave policies help? *Analyses of Social Issues and Public Policy*, 20(1):382–396.

Ottosen, M. H. (2014). The long-term impacts of early paternal involvement in childcare in Denmark: what happens after nuclear family dissolution. In Eydal, G. B. and Rostgaard, T., editors, *Fatherhood in the Nordic Welfare states: Comparing care policies and practice*. Policy Press.

Pavolini, E., Scalise, G., et al. (2021). Following different paths of modernization: The changing socio-cultural basis of southern Europe. In Luigi Burroni, E. P. and Regini, M., editors, *Mediterranean Capitalism Revisited. One Model, Different Trajectories*. Cornell University Press.

Petersen, T., Penner, A. M., and Høgsnes, G. (2014). From motherhood penalties to husband premia: The new challenge for gender equality and family policy, lessons from Norway. *American Journal of Sociology*, 119(5):1434–1472.

Pfau-Effinger, B. (2014). Explaining differences in child care and women's employment across six European 'gender arrangements'. In León, M., editor, *The Transformation of Care in European Societies*, pages 83–103. Palgrave Macmillan UK, London.

Pfau-Effinger, B. (2023). Theorizing the role of culture and family policy for women's employment behavior. In Daly, M., Pfau-Effinger, B., Gilbert, N., and Besharov, D. J., editors, *The Oxford Handbook of Family Policy Over The Life Course*, pages 224–242. Oxford University Press.

Pierson, P. (2001). *The New Politics of the Welfare State*. Oxford University Press.

Ponthieux, S. and Meurs, D. (2015). Gender inequality. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, volume 2 of *Handbook of Income Distribution*, pages 981–1146. Elsevier.

Pylkkänen, E. and Smith, N. (2004). The impact of family-friendly policies in Denmark and Sweden on mothers' career interruptions due to childbirth. Available at SSRN 522282.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rossier, C., Brachet, S., and Salles, A. (2011). Family policies, norms about gender roles and fertility decisions in France and Germany. *Vienna Yearbook of Population Research*, pages 259–282.

Salmi, M. and Närvi, J. (2017). Perhevapaat, talouskriisi ja sukupuolten tasa-arvo. Report 4, Terveyden ja hyvinvoinnin laitos, Helsinki, Finland.

Saraceno, C. (2022). *Advanced introduction to family policy*. Edward Elgar Publishing.

Saraceno, C. and Keck, W. (2011). Towards an integrated approach for the analysis of gender equity in policies supporting paid work and care responsibilities. *Demographic Research*, 25:371–406.

Sjöberg, O. (2004). The role of family policy institutions in explaining gender-role attitudes: A comparative multilevel analysis of thirteen industrialized countries. *Journal of European Social Policy*, 14(2):107–123.

Stan Development Team (2023). RStan: the R interface to Stan. R package version 2.32.3.

Stan Development Team (2024). Stan modeling language users guide and reference manual, version 2.34.

Statistics Finland (2021). Early childhood education and care [online publication]. [Referenced: 13.11.2024]. Access method: https://stat.fi/en/publication/ckwd9j2c8fbpq0c53mhrob1za.

Svallfors, S. (1997). Worlds of welfare and attitudes to redistribution: A comparison of eight western nations. *European Sociological Review*, 13(3):283–304.

Tamm, M. (2019). Fathers' parental leave-taking, childcare involvement and labor market participation. *Labour Economics*, 59:184–197.

Uunk, W. and Lersch, P. M. (2019). The effect of regional gender-role attitudes on female labour supply: A longitudinal test using the BHPS, 1991–2007. *European Sociological Review*, 35(5):669–683.

van Kersbergen, K. (2003). *Social Capitalism*. Routledge, London.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16(2):667–718.

Wood, J., Marynissen, L., and Van Gasse, D. (2023). When is it about the money? Relative wages and fathers' parental leave decisions. *Population Research and Policy Review*, 42(6).

Österbacka, E. and Räsänen, T. (2022). Back to work or stay at home? Family policies and maternal employment in Finland. *Journal of Population Economics*, 35(3):1071–1101.