UNIVERSITY OF LEEDS

Analysing spatio-temporal networks with continuous-time models

Sarah Catherine Gadd

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Faculty of Earth and Environment

School of Geography

February 2021

Intellectual Property and Publication Statements

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 6 has appeared in publication as:

Gadd SC, Tennant PWG, Heppenstall AJ, Boehnke JR, Gilthorpe MS (2019) Analysing trajectories of a longitudinal exposure: A causal perspective on common methods in life-course research. PLoS ONE 14(12): e0225217. https://doi.org/10.1371/journal.pone.0225217

Contributions: MSG conceived ideas for this paper. SCG conducted the investigation and formal analysis with methodology planned by SCG, MSG and PWGT. SCG wrote the original draft which was reviewed and edited by PWGT, WJH, JRB and MSG. MSG, AJH and PWGT provided supervision to SCG.

Chapter 7 has been accepted for publication as:

Gadd SC, Comber A, Gilthorpe MS, Suchak K, Heppenstall AJ (In Press) Simplifying the interpretation of continuous-time models for spatio-temporal networks. Journal of Geographical Systems.

Contributions: SCG, AJH and AC contributed to the conceptualisation and planning of the study. KS provided quality assurance for analysis methods while some of the supervi-

sory team were unable to contribute. SCG carried out the investigation and analysis and wrote the original draft. All authors contributed to editing and review. AJH and MSG provided supervision to SCG.

Chapter 8 is ready for submission for publication to the journal Computers, Environment and Urban Systems with the following authors:

Gadd SC, Comber A, Tennant PWG, Gilthorpe MS, Heppenstall AJ.

Contributions: SCG, AJH and MSG contributed to the conceptualisation and planning of the study. SCG carried out the investigation and analysis and wrote the original draft. All authors contributed to review and editing. AJH, MSG and PWGT provided supervision to SCG.

Acknowledgements

First and foremost I wish to thank my supervisors, Alison Heppenstall, Mark Gilthorpe and Peter Tennant, for their guidance and support during the completion of this work. Thank you also to Alexis Comber who provided useful guidance and feedback throughout the project. My gratitude extends to the Economic and Social Research Council and White Rose Doctoral Training Partnership for the funding opportunity to undertake my studies.

Many colleagues in the Leeds Institute for Data Analytics have generously offered support during my studies. I particularly wish to thank Ning Lu and the Data Science Society for recruiting me into their welcoming and friendly committee.

Thank you to all my friends in Leeds, especially the swimming women, who have made me feel at home. Finally I wish to thank my family, especially my mother, Suzanne, and my partner, Alex, for their support and encouragement throughout the last four years.

Abstract

In the analysis of spatio-temporal networks, many of the commonly used analysis methods and models use discrete specifications of time – these divide time into a series of discrete steps. Discrete time specifications require irregularly measured data to be temporally aggregated into the discrete steps, which introduces error. They also require careful choices regarding the length of time steps. Using continuous-time models would avoid these issues.

This thesis examines the application of continuous-time models, specifically multilevel models, to examine temporal patterns of edge properties in spatio-temporal networks. Following a scoping review of the extent to which continuous-time methods are used in spatio-temporal network literature, a spline-based multilevel model specification suitable for analysing temporal patterns of edge properties, along with two methods for simplifying the interpretation of results from this model, are introduced. These include extracting temporal 'pattern features' (and estimates of uncertainty), which can be further analysed, and generating continuous temporal functions of whole-network properties (with estimates of uncertainty) which is a novel contribution. These methods are applied to real and simulated data, and the ability of the model to extract temporal pattern features accurately and with appropriate precision was tested using simulations.

Simulated data were generated using a widely applicable simulation tool developed as part of this research. The thesis also includes a simulated illustration of one of the potential pitfalls of analysing temporal patterns in data – this was identified as an issue in early chapters but had not been extensively covered in the literature, so this illustration and recommendations of how to avoid this issue are an important contribution of this research.

Overall, this thesis identifies that continuous-time models have potential applicability in spatio-temporal network analysis, especially in the presence of irregularly measured data.

Further investigation into different model types and specifications to improve accuracy is recommended.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Geographers have been evolving the methods they use to understand the world for decades. In the past, the availability of large quantities of detailed data and powerful computers to process it was limited. This meant that models and methods focused on aggregate and high-level relationships between variables. In the present day, data concerning the world around us are regularly collected for commercial or other purposes (Higham et al. 2017). For example, railway systems might use smart cards as tickets and collect data on customer journeys in order to charge them the correct fare; weather data is automatically collected at a range of monitoring sites for forecasting purposes; remote sensing systems enable collection of geographical data on a large scale. Greater quantities of data with finer detail are available which enables more detailed modelling of finer-scale processes.

One area of interest to geographers is the movement of people around cities. This field of interest comprises complex and individual movement, thus requiring researchers to look in greater detail than aggregate variable relationships across a whole city. For example, they may focus on collecting data on individual movements, or on the movements between selected areas or locations. One method that can be used to study movement is network analysis (Newman 2018). When examining movement around cities, network analysis represents locations as vertices/nodes and adds connections (edges/links) between locations if there is travel between them. Networks can also be used in other contexts, for example, to study infectious disease spread, or to study ecological connections between species (food webs, for example) (Anderson and Dragicevic 2018; Iotti et al. 2017).

Network analysis can be used to produce aggregate statistics describing the network of movements around a city – for example, whether there are clusters of locations that people tend to remain within and rarely leave, or whether people take a wide range of paths throughout the network or only use a few of the available options to move between locations. Descriptions of specific locations (vertices) or specific edges (links between areas) and how they interact with other parts of the network can also be produced. Furthermore, network connections can be used to inform other analyses and models. Many spatial analyses aim to investigate or account for spatial autocorrelation – a degree of similarity between locations or objects due to their proximity in space (Dubin 2009; Tobler 1970). Proximity of vertices in a network, for example being directly linked by a single edge, rather than indirectly linked via another vertex, may also lead to autocorrelation. Analysing the connections or movement between locations as a network allows us to represent network proximity numerically and analyse or account for it in models.

Many representations of networks are static, that is they do not change over time. In many situations, however, we might expect a network to change over time (Blonder, Wey, et al. 2012). For example, a network representing movement of public transport buses around a city is likely to appear different on a bank holiday than on a normal weekday, as fewer bus services will be running. Temporal variation in the network can take place in the network topology (the presence of edges and vertices in the network) and the network properties (for example, edge weights representing the strength of connections, e.g. the frequency of buses connecting two locations; or the location of vertices might change over time) (Blonder, Wey, et al. 2012). These changing properties can be of interest to researchers. For example, they can be used to predict traffic flow in a network, to examine ecological changes over time, or to identify times of high travel demand in a public transport network (Carlsson and Kindvall 2001; T. Cheng et al. 2014).

Many current methods that analyse spatio-temporal networks tend to specify time as a series of discrete time intervals. This is in contrast to the true, underlying continuous nature of time, which can be split into infinitely small intervals. While specifying time as a discrete variable is convenient for the specification of some models, it can introduce bias to the results of analyses and lead to the loss of information where data are not measured exactly at these discrete time intervals (de Haan-Rietdijk et al. 2017; Hwang

2000; Rossana and Seater 1995; W. Wei 1981; Weiss 1984). Using continuous representations of time would avoid these issues, which is why this research aims to examine the use of continuous-time models to examine spatio-temporal networks as an alternative to the more familiar discrete time specifications. Consequently, this research focuses on the application of continuous-time statistical methods to spatio-temporal networks, examining temporal patterns of network properties.

The aims of this thesis are to:

- Review existing analysis methods for spatio-temporal networks as used in the recent literature. This will give a clearer picture of the range of methods used (including their specification of time), and in what contexts spatio-temporal networks are applied.
- Identify continuous-time models that could be applied to examine temporal network property patterns in spatio-temporal networks and critically analyse them to identify the most appropriate model development strategies.
- Develop an analytical procedure that applies these models to spatio-temporal network data and simplifies their interpretation in order to provide clear 'real-world' inferences.
- Evaluate the accuracy and precision of this analytical procedure in specific circumstances, using simulations, to investigate its limitations.

Chapter 2 reviews methods used to analyse spatio-temporal networks in scientific literature and discusses their limitations. Chapter 3 provides an overview of continuous-time models used to analyse temporal data that could potentially be used to analyse spatio-temporal networks. Chapter 4 outlines the model framework applied to spatio-temporal networks in this research and introduces methods for simplifying the interpretation of results within this framework. Chapter 5 introduces some real data and tools for producing simulated data used in further analysis of the model framework discussed in Chapter 4. Chapter 6 provides a more detailed analysis of the problems with one of the methods introduced in Chapter 3 – this had not previously been covered in detail in the literature, even though the methods are widely used, particularly in the field of health sciences. Chapter 7 shows an example application of the model framework introduced in Chapter 4 to both real and

simulated data. Chapter 8 includes a simulation evaluating the accuracy and prevision of the model framework introduced in Chapter 4 when applied to spatio-temporal network data. Chapter 9 summarises and discusses the preceding material with respect to strengths and weaknesses and provides suggestions for future work.

This thesis is submitted in the alternative format: some of the chapters included are published or written for publication in peer reviewed journals. These papers make up Chapters 6, 7 and 8.

Chapter 2

Scoping Review of Spatio-temporal Network Methods

2.1 Introduction

This chapter includes a scoping review of methods used to analyse spatio-temporal networks in the literature preceded by background information on space-time methods and network analysis. The key focus for this review is the specification of time as a continuous or discrete variable.

2.2 Background

2.2.1 Space-time methods

Many problems and questions have both a spatial and a temporal dimension, that is, they seek to examine variation in some property over both time and space simultaneously. Unlike many standard statistical analyses, methods applied to spatio-temporal data must account for interdependence of observations due to both spatial and temporal autocorrelation (Dubin 2009; Fitzmaurice and Ravichandran 2008). A range of geographical methods are used for spatio-temporal analyses: space-time cubes, prisms, and paths; geographically and temporally weighted regression; space-time filtering; space-time autoregressive integrated moving average (STARIMA) models and their relatives; and spatio-temporal Kriging. These methods are briefly introduced below before a critical review is presented.

Space-time cubes were originally developed by Hägerstraand as a component of his 'time geography' (Hägerstraand 1970). They provide a way of visualising spatio-temporal information over three axes. They are often used to display paths of movement over space and time, or the potential space that an object could move to, given a starting point and time limit. These are known as space-time paths and potential path spaces, respectively (Kraak 2003). Such approaches can be used in studies of accessibility in which the calculation of potential path spaces with a range of restrictions (for example, activity schedules or transport restrictions) is of importance (Kraak 2003; Kwan 2013).

Geographically and temporally weighted regression is a special case of local regression that allows for heterogeneity in model parameters across space and time. The model for a relationship between two (or more) variables at a given space-time location is estimated using both data from that location, and data from nearby locations which are weighted according to their distance from the location of interest (Fotheringham et al. 2015). This approach to modelling spatio-temporal variation is non-parametric, that is, variation in the relationships between variables over space and time does not have to follow a particular functional form. The user must specify which nearby locations are used to estimate a local relationship of interest and how much each one contributes to the estimation. This is specified using a kernel function that weights each observation's contribution and includes a maximum bandwidth – a distance from the point of interest outside of which observations are not used to estimate the relationship (Huang et al. 2010). Kernel functions can take a wide range of forms. Spatio-temporal kernel functions include both space and time dimensions separately, rather than having them operate together as if representing a three-dimensional space (Huang et al. 2010).

Space-time filtering is a semi-parametric approach to modelling data with spatial and temporal autocorrelation. It uses a 'filter' model to remove spatial and temporal variation in the data. The residual data can then be modelled using standard methods that assume independence of observations (Griffith 2012; Patuelli et al. 2006). This method essentially separates the data into different components which are modelled separately.

STARIMA integrates autoregressive and moving average model structures, in which the current value of a variable depends on previous values (autoregressive) or previous model error values (moving average). These structures exist in both spatial and temporal dimen-

sions (Pfeifer and Deutsch 1980). Spatial relationships between observations are specified using a spatial weight matrix – this is often based on neighbour relations of locations rather than distance-based relationships. Neighbour relationships specify whether two locations are 'first neighbours', for example if they share a boundary, or 'second neighbours', if they share a first neighbour *et cetera* (Dubin 2009). A maximum distance over which locations can affect each other through autoregressive and moving average processes is specified. Similarly, a maximum time lag over which previous observations affect current ones is specified (Pfeifer and Deutsch 1980). Autoregressive and moving average processes that deal with time lags tend to treat time as a series of discrete steps, rather than a truly continuous process (Oud et al. 2012).

Spatio-temporal Kriging is a geostatistical method that uses explicit modelling of spatio-temporal covariance structures. It aims to obtain more precise predictions of variables at different points in space and time. This method models a mean spatio-temporal trend in addition to explicitly modelling spatio-temporal random variation (Gneiting et al. 2006). Modelling of the random variation is carried out using a variogram model. Different variogram models incorporate space and time dimensions in different ways; specifically 'separable' variogram models combine spatial and temporal models as if time were an extra spatial dimension, for example, by multiplying them together, whereas 'non-separable' models are designed specifically for examining space and time together and allow space-time interactions (Gneiting et al. 2006). Separable models are more computationally efficient, but the assumption of no space-time interactions is not always appropriate (Gneiting et al. 2006).

2.2.2 Networks and analysis methods

Networks are used in a wide range of subject areas to represent or analyse connections between objects. They can be used in many different contexts to represent different types of structures using two structural components: edges (or links) representing the connections between vertices (or nodes) which represent the objects (Newman 2002).

Edges and vertices can represent both tangible and abstract objects. For example, networks are commonly used in the transport literature. In this area, one might use a network to represent a railway network. Vertices might represent stations and edges the physical

rail line connections between them – these are both tangible objects.

In some contexts, vertices might represent tangible objects, but edges might represent more abstract connections between them. For example, social network analysis is often used to represent social connections between people (Scott 1988), which might be thought of as a more abstract concept than a physical rail line.

In geography, some spatial structures can be thought of as network connections. For example, when examining geographical areas, boundary sharing is often used as a way of representing spatial connections between areas. Boundary sharing can be thought of as a network connection between areas, with each area representing a vertex and each shared boundary representing an edge between the areas in question (Dubin 2009).

The use of networks is not limited to the examples above – network structures are found in a range of areas. For example, they might be used to represent economic trading networks (S. Liu et al. 2018), urban structures like road networks (Colak et al. 2013), ecological environments (Neeson et al. 2012), ecological connections between organisms such as food webs (Anderson and Dragićević 2018), communication networks (Ramachandran et al. 2007a), or the spread of a contagious disease through a network of contacts (Iotti et al. 2017).

The following paragraphs include some background information on network structures, statistics used to describe them and the treatment of time in network analysis. They also discuss the possibility of networks forming a context for analyses where they are not the main focus.

Networks are graph structures formed of vertices (or nodes) connected by edges (or links). They can be represented in data as an origin-destination matrix, with rows and columns representing each vertex. A zero entered in row i and column j shows there is no edge from vertex i to vertex j, and a one can be used to represent an edge between vertices (Newman 2003).

Networks can be directed or undirected. In directed networks, the edges between vertices have a direction. This indicates that the connection, or flow of something along the connection between vertices, is unidirectional (Newman 2003). For example, an edge may go from vertex i to vertex j but not vice versa. To include a two-way connection, two

edges would be included. In an undirected network, edges are inherently bidirectional and indicate a connection from vertex i to vertex j and vice versa. For undirected networks, the origin-destination matrix is always symmetrical, but this is not necessarily the case for directed networks (Newman 2018).

Edges in networks can be weighted to indicate, for example, the intensity of a connection between two vertices. In a weighted network, a range of numbers (not just one) may be used in the origin-destination matrix to indicate the presence of a connection and its weight (Newman 2003). Other properties may be assigned to vertices and edges, too. For example, vertices may be given a location. In a road network, edges might be classified by the type of road surface or structure.

It is possible to calculate all possible paths between two vertices on the network along edge connections. This can be useful, for example, when finding possible routes for moving through a transport network. Often the shortest of the potential paths (in terms of the number of edges traversed) between two vertices is used to represent how closely they are connected (Saberi et al. 2017). This is called the shortest path distance. The longest distance between any two vertices in the network is called the network diameter. This whole-network property gives an indication of the network distance spanned by the graph (Saberi et al. 2017).

There are other network properties that can be used to summarise network structure at both a local and whole-network level. Network properties are descriptive statistics that are used to give information about the structure of a network, either across the whole graph or at the level of individual vertices and edges, or subgroups of them.

Centrality is one such property – it is intended to give an indication of the importance of different vertices within the network (Newman 2002). For example, in a rail network where vertices represent stations and edges represent rail connections between them, centrality measures aim to identify which stations are the most important for the functioning of the rail network. When assessed locally, centrality indicates how 'important' or 'well-connected' a vertex is within a network (L. Freeman 1978; Opsahl, Agneessens, et al. 2010). A simple way to measure this is vertex degree, the number of edges leaving or entering a vertex. In directed networks, this can be subdivided into in-degree (edges *to* the vertex in question) and out-degree (edges *from* the vertex). In weighted networks, the

sum of the weights of edges leaving or entering a vertex can be used to represent centrality. This is called vertex strength (Opsahl, Agneessens, et al. 2010).

Vertex degree and strength only take into account the network structure immediately surrounding the vertex in question (Opsahl, Agneessens, et al. 2010). Other measures can be used to assess vertex centrality while accounting for wider network structure. For example, closeness centrality is the inverse of the sum of distances from a vertex to every other vertex in the network (L. Freeman 1978). The distances are calculated as the number of network edges traversed on the shortest path from one vertex to another. This measures how closely connected a vertex is to all other vertices in the network, rather than just the number of immediate connections. Betweenness is another measure of centrality. This method calculates all the shortest paths between each pair of vertices in the network. The betweenness centrality of a vertex is the number of these shortest paths that it appears in (L. Freeman 1978). This gives an indication of how important a particular vertex is in maintaining connectivity across the whole network.

Centrality can also be measured for a whole network to give an indication of the extent to which centrality is concentrated in a small number of vertices. The distribution of vertex degrees is often of interest as it can give some indication of the processes underlying network formation (Newman 2003). The average degree can also be calculated. Whole-network closeness centrality can be calculated from individual vertex centrality measures if they are scaled to a range of zero to one. The resulting measure gives an indication of the extent to which centrality is concentrated in a small number of vertices versus spread evenly across the whole network (L. Freeman 1978).

Edge density is another important whole-network property. This gives an indication of how many edges there are in a network, relative to the number of vertices. In a rail network, a high edge density would indicate a lot of connections between all the rail stations, whereas a rail network with low edge density would have few connections linking them. Edge density is the ratio of the total number of edges in the network, to the number of possible edges (Horvath 2011). For weighted networks, the total number of edges is replaced with the sum of all edge weights. This gives an indication of whether connections in the network are rare or abundant in relation to the number of vertices.

Sometimes, networks contain clusters of vertices that are more densely connected to each

other than to other vertices. A clustering coefficient gives an indication of the extent to which clustering exists in a network (Saberi et al. 2017). There is more than one way to measure clustering, but one common clustering coefficient is based on identifying triplets in a network – these are sets of three vertices with two or three edges connecting them. The clustering coefficient is the proportion of triplets with three edges connecting them (Saberi et al. 2017). If a network has high clustering, vertices that share a neighbour are more likely to be directly connected themselves. This will give a higher proportion of triplets with three edges, rather than two. There are also several methods that aim to identify cluster membership of individual vertices in a network. These are often algorithms that aim to divide the network graph into sections while optimising a property related to clustering, similar to clustering algorithms used on data without a network structure (Luptáková and Pospichal 2018). There are alternative clustering methods for networks (such as 'walktrap clustering' (Pons and Latapy 2005)) that are based on simulated flow of some object, for example information or vehicles, between vertices.

Networks can be temporally dynamic, that is, the structure of the network might change over time. This change might take place in vertex or edge properties or the presence of vertices and edges in the network (Blonder, Wey, et al. 2012). Changes in network properties may take place as a result of flow of some object, property or substance through the network, for example vehicles or information. When network structure changes dynamically over time, network analysis carried out on aggregated time windows or cross-sections may misrepresent the network structure, resulting in inappropriate inferences (Blonder, Wey, et al. 2012). Time-ordered networks are a way of representing temporally dynamic networks with changing connections. Graphically, these show each vertex in the network as they move through the time dimension, with arrows indicating the start and end times for connections between different vertices. Mathematically, several 'temporary vertices' are used to represent each vertex in the network at the start and end times for each connection. Edges between the temporary vertices are included at their appropriate start and end times, and directed edges are used to connect temporary vertices that represent the same vertex in time order (Blonder, Wey, et al. 2012). This method is useful for looking at changes in binary properties over a continuous time period, for example, the presence or absence of connections, but cannot represent continuous-time changes in continuous properties, for example, continuous changes in flow or spatial location of vertices. The

graphical representation of time-ordered networks could be extended to include a spatial dimension in a similar way to the space-time cube, however this may result in a crowded graph that is difficult to interpret, especially for large networks.

Networks can also form a context for analyses where they are not the direct subject of investigation. When analysing data that is associated with a network it is important to account for this underlying structure. Connections between observational units, for example, due to spatial or temporal proximity, must often be accounted for in analyses as these connections violate the assumption of independence of observations made by many simpler statistical methods (Dubin 2009; Fitzmaurice and Ravichandran 2008). Network connections between observational units may also lead to a violation of this assumption and must be accounted for in the analysis (T. Cheng et al. 2014). As with spatial correlations, choices must be made about how to include network correlations in an analysis – these could be based only on immediate connections (e.g. boundary sharing, or sharing a single edge), or on the number of edges between two vertices (their geodesic distance) (Ermagun and Levinson 2018; Leenders 2002). It is also possible to include spatial information, for example, the distance along a road network between two junctions.

2.2.3 Method types for spatio-temporal networks

There are a range of different method types that can be used to examine spatio-temporal networks. In this review they are categorised as stochastic models, statistical models, machine learning methods, discrete-time network analysis, continuous-time network analysis and visualisation methods.

Stochastic models are models of a system that inherently include some amount of randomness (Morris et al. 2019; Zoppou 2001). For example, a model for the occurrence of an event may include the probability of an event happening in a given period. The output of the model will vary each time it is run because the model contains uncertainty as to whether the event will happen or not (Zoppou 2001). As well as discrete events, stochastic models can model properties of a system or parts of a system. These might be discrete properties (e.g. the colour of a car) or continuous (e.g. the amount of rainfall). The probability of properties taking certain values, or of events occurring, is specified using a probability distribution which might be estimated using real data (Morris et al. 2019).

Often a statistical model estimates the relationship between one or more input variables and one or more outcome variables. Unlike stochastic models, statistical models fit a deterministic relationship between variables (Zoppou 2001); if used to generate data, the relationship would be the same every time. Randomness is incorporated as error or residual variation, often in the outcome variable(s), and is constrained to follow a particular probability distribution (Zoppou 2001).

The aim of machine learning is to provide computers with 'experience' from which they 'learn' through the use of algorithms (Bi et al. 2019). In the context of the papers reviewed, the 'experience' comes in the form of data and the computer 'learns' by identifying patterns in the data. This is often done with the aim of prediction, rather than inference (Bi et al. 2019). The boundary between machine learning and statistical methods is sometimes unclear – some model selection algorithms and methods like logistic regression are considered machine learning methods as well as being common statistical methods (Muthukrishnan and Rohini 2016; Wu et al. 2010). There is some overlap between machine learning methods and network analysis methods, in that some machine learning methods are designed specifically for networks.

Network analysis methods are a set of analysis tools designed specifically for network structures. They often include metrics for summarising local or global properties of a network, or analysis methods aiming to find paths through a network. Various network and vertex properties were discussed in Section 2.2.2. Network analysis methods might aim to treat time as a discrete or continuous variable, however it is much more common for time to be aggregated into discrete windows with static representations of the network examined in each window (Blonder, Wey, et al. 2012).

Visualisation methods are often important for enabling interpretation of analyses and communication of information to a wider audience. However, visualising variation in a property over both time and space, in the context of a network, is challenging. Often, spatio-temporal visualisation is approached by aggregating data into time windows and plotting static network maps. These methods mean the reader only sees temporal patterns as snapshot images that can be difficult to link together and may leave out important temporal pattern features occurring in between snapshots. As the interpretation of spatio-temporal information is so difficult, alternative visualisation methods to discrete-time

snapshots are of particular interest for this review.

2.2.4 Specification of time

Most of the spatio-temporal methods in Section 2.2.1 and methods for examining temporal networks use a discrete specification of time. That is, instead of treating time as a continuous variable, they treat it as a series of discrete time steps or windows. This can be convenient for modelling complex systems but is contrary to the true nature of time which is continuous and can be divided into infinitely small units. Treating time as a series of discrete steps requires the researcher to choose the length of these steps carefully based on their knowledge of the underlying system (Comber and Wulder 2019) – too long or short a step length can mask variation in temporal patterns and introduce bias (J. Freeman 1989; W. Wei 1981; Weiss 1984). In many cases, however, the minimum size of time step is determined by the frequency of measurements in the data, rather than knowledge of the underlying system (J. Freeman 1989). In some cases, data will not be measured at regular intervals – in this situation the measurements are often aggregated into time steps. Temporal aggregation introduces error into the time measurement for each observation and can introduce bias to autoregressive and other time series models (Hwang 2000; Rossana and Seater 1995). Using continuous-time models in spatio-temporal data analysis and temporal network analysis would circumvent the issues associated with the choice of time step length and temporal aggregation.

2.3 Review aims

This scoping review aims to survey the methods used to analyse spatio-temporal networks in the published literature, with a particular focus on the treatment of time (discrete or continuous) by these methods. Often the methods chosen for an analysis are shaped by the research question posed which is in turn shaped by the context of the analyses (the subject area, structure represented by the network, temporal and spatial dimensions). Therefore, this review will examine the context of the analyses, research question type (for example, descriptive or forecasting) and analysis methods used in relevant papers, including the specification of time (discrete or continuous) and methods used to visualise or present spatio-temporal networks.

2.4 Methods

A scoping review was used to identify papers analysing spatio-temporal networks and survey the methods and models they utilised. The SCOPUS and Web of Science databases were searched for papers published on any date up to February 2021 (inclusive), in English, with a title, abstract or keywords matching one or more of the following phrases (where '?' represents a wild-card character): 'space?time' or 'spatio?temporal' preceding either 'network' or 'graph'; or 'dynamic' preceeding 'spatial', 'space?time' or 'spatio?temporal', and the word 'network'.

Papers were screened for relevance based on their title, abstract, introduction and methods sections, and then full text. A range of information was recorded about the papers, including their subject area, research question or aim, methods applied, time period, spatial scale, network level of input and output information and specification of time (discrete or continuous). Once recorded, much of this information was categorised as shown in Table 2.1.

The aims of the papers fell into four categories: 1) hypothetical 'what if?' scenarios, for example, assessing potential policies; 2) description of the structure of a system, for example, examining the structure of a social network; 3) quantifying relationships between variables, for example, examining the relationship between two properties of objects connected in a network; and 4) forecasting, for example, traffic flow prediction.

The 'network level' characteristic is not commonly used and may be unfamiliar for readers. In this review, the papers examined worked with information at three different levels: whole-network information, for example, properties relating to the whole of the network; information about network components, for example, properties of individual edges and vertices in the network; or properties of actors that moved through the network, for example in a road traffic network, vehicles can be considered actors that move through the edges (road segments) and vertices (junctions) of a road network.

The same object might be represented at different network levels in different analyses depending on how it is characterised. For example, in a social network analysis, people often form the vertices of a network with their social connections forming the edges. In this context, there are no actors moving through the network. However, in the context of

Information	Definition	Categories
Subject area	Subject area in which the analysis was carried out	Transport, Economics, Ecology, Sociology, Network Analysis Methods, Communications, Geography, Urban Planning History, Epidemiology
Research question/aim	The type of research question being asked	Hypothetical scenario, Describe network structure, Quantify relationships, Forecasting
Methods applied	The type of method used to analyse a spatio-temporal network	Stochastic Simulation, Statistical Model, Discrete-time network analysis, Machine Learning, Visualisation, Continuous-time network analysis
Time specification	Whether the analysis uses discrete or continuous time specification	Discrete, continuous
Time period	The approximate time period the data analysed covered – for some papers this was an abstract number of time steps or unspecified period	Hours, Days, Months, Years, Abstract/Unspecified
Spatial scale	The size of the spatial area covered by the data – in some examples, particularly simulations, the area covered is not specified	Small area, Partial City of town, City-wide, Regional, Multi-region, Country wide, Multi-country, Unspecified
Network level (input and output)	Level on the network structure which input and output information for the analysis is recorded	Whole-network, Network components (edges or vertices), Actors (objects moving through the network)

Table 2.1: Characteristics recorded about reviewed papers

a public transport network, people may be represented as actors that move through the edges and vertices of a network formed by train lines or bus connections.

The 'input network level' detailed whether the data used as an input for the analysis pertained to actor, component or whole-network properties. The 'output network level' detailed whether the outcome of interest for the analysis related to actor, component or whole-network level properties. For both input and output, it was possible for one analysis to include information on more than one level, for example, a simulation of animals moving through a hypothetical river network might include information about the animal behaviour (actor level) and the structure of the river network (whole-network level) as an input.

2.5 Results

538 unique references were initially identified and screened for relevance based on title, followed by abstract, introduction and then full text. The screening process is summarised as a flow diagram in Figure 2.1. 87 relevant papers remained. Two of these were not duplicated references but detailed the same work by the same authors and were reviewed as one paper – leaving 86 papers (Abdelghany et al. 2001; Albers et al. 2018; Anderson and Dragicevic 2018; Anderson and Dragićević 2018; Anderson and Dragićević 2020a; Anderson and Dragićević 2020b; Basile et al. 2018; Baybeck and Huckfeldt 2002; Burnett and X. Zhao 2017; Cai et al. 2014; Carlsson and Kindvall 2001; Carter et al. 2020; Caruso 2007; L. Chen et al. 2020; S. Chen, Ilany, et al. 2015; S. Cheng et al. 2018; T. Cheng et al. 2014; Colak et al. 2013; Diao et al. 2019; Donfouet et al. 2018; Fan et al. 2019; S. Fang et al. 2019; Y. Fang et al. 2019; Feng et al. 2020; Ferenc et al. 2016; Garbin et al. 2019; D. Guo et al. 2013; S. Guo et al. 2019; Y. Guo et al. 2020; Hall et al. 2018; Han et al. 2019; He et al. 2020; Iotti et al. 2017; Jacoby et al. 2012; Jeong and L. Lee 2020; Jia et al. 2019; G. Jin et al. 2020; Y. Jin et al. 2019; Jovanovic et al. 2019; Kuersteiner and Prucha 2020; Lengyel et al. 2020; G. Li et al. 2020; T. Li and Liao 2016; W. Li et al. 2021; X. Li and Griffin 2013; Y. Li et al. 2008; Z. Li et al. 2020; Lin et al. 2020; C. Liu et al. 2018; S. Liu et al. 2018; Ma et al. 2020; Mateus-Anzola et al. 2019; Mendoza-Cota and Torres-Preciado 2019; Neeson et al. 2012; Neto et al. 2016; Ou et al. 2020; Parent and LeSage 2010; Peng et al. 2020; Perez and Dragicevic 2009; Perry and F. Lee 2019; Qu

et al. 2020; Ramachandran et al. 2007a; Ramachandran et al. 2007b; Ryser et al. 2019; Semboloni 2000; Srivastava and Salapaka 2020; B. Sun et al. 2021; G. Sun et al. 2017; K. Tian et al. 2020; X. Wang et al. 2016; Wanzenböck and Piribauer 2018; V. Wei et al. 2020; Xu and Z. Gao 2009; B. Yang et al. 2020; Yao et al. 2019; H. Yu et al. 2019; W. Yu et al. 2019; Zambrano-Monserrate et al. 2020; C. Zhang et al. 2019; H. Zhang et al. 2020; W. Zhang, Z. Tian, et al. 2019; X. Zhang et al. 2019; Z. Zhang et al. 2019; N. Zhao et al. 2020; X.Y. Zhao and Pu 2018; Zheng et al. 2020; F. Zhou et al. 2020).

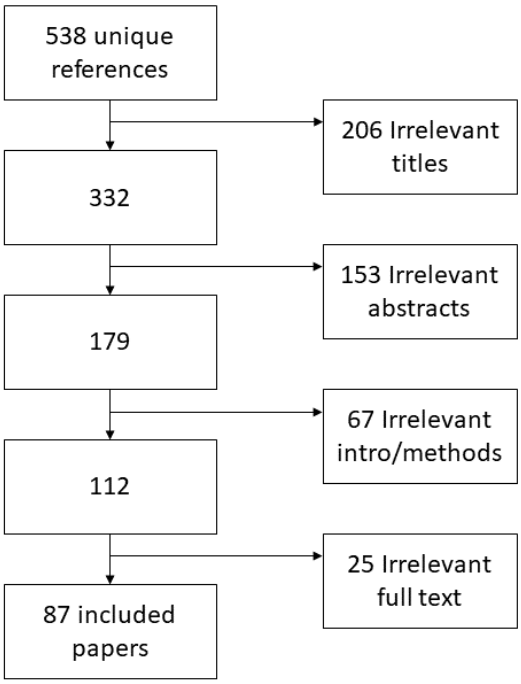


Figure 2.1: Flow diagram showing the process of identifying relevant papers.

A summary of the different characteristics of interest for these papers is shown in Table 2.2. The sum of counts for some variables is greater than 86 – this occurs if some papers fit into more than one category. Individual information for the papers is shown in Appendix A. The most common subject area was transport followed by ecology and economics. The least common subject area was history. The most common method was stochastic simulation, followed by machine learning, which was very common amongst more recent papers. Discrete network methods, i.e. carrying out network analysis methods on discrete time slices or periods of data, were also commonly used, though only one paper sought to use continuous-time network methods. Only four papers treated time as a continuous variable in their analysis, and one of these also included a discrete specification alongside continuous information. These included two statistical papers and one using stochastic

methods, alongside the continuous-time network methods paper.

Most papers used data that spanned multiple months or years. Many machine learning papers used months of data in order to train their models. The most common spatial level was the town or partial city. This was often used in machine learning traffic prediction papers. Many papers also had an abstract spatial setting – this was used frequently in papers developing new models or looking at effects of network structure on various outcomes.

Figure 2.2 shows the correspondence between the subject area, aim, method type and time specification of papers. Three subjects make up more than half of the papers: transport, ecology and economics. A large proportion of transport papers involve forecasting and these make up the majority of forecasting papers. Ecology and economics make up a large proportion of the hypothetical scenario papers, with the rest being transport, network methods, communications, urban planning and epidemiology, the latter of which exclusively includes papers with hypothetical aims. Sociology papers only included aims to do with description and quantifying relationships; geography papers also only aimed to quantify relationships.

Hypothetical scenario papers exclusively employed stochastic methods, with some forecasting and descriptive papers also using these methods. Most of the papers aiming to quantify relationships between variables used statistical methods and most aiming to describe the structure of a network used discrete or continuous-time network analysis and visualisation methods. Forecasting papers were the only papers to use machine learning, and almost all of these were in the transport literature (with some in the communications area), however stochastic and statistical methods were also used for forecasting aims.

Figure 2.3 shows the different levels of the network information used as inputs and outputs for papers with different method types. Only papers using stochastic methods incorporated information at the level of whole network structure as an input. These papers either looked at the effect of changing network topology on the outcome of the stochastic simulation (Carter et al. 2020; Y. Li et al. 2008; Neeson et al. 2012) or used information about a network (for example a transport network) to govern the movement of agents (Garbin et al. 2019; Perez and Dragicevic 2009). The majority of these types of papers used simulated networks with an undefined spatial area (Garbin et al. 2019; Neeson et al. 2012), but

Information	Category	Count
Subject area	Transport	39
	Ecology	14
	Economics	13
	Network Analysis Methods	8
	Sociology	8
	Communications	4
	Epidemiology	4
	Urban planning	3
	Geography	2
	History	1
Aim	Forecasting	29
	Hypothetical scenario	25
	Describe structure	23
	Quantify relationships	13
Method	Stochastic sim	28
	Machine learning	23
	Statistical model	20
	Discrete network methods	18
	Visual	3
	Continuous network methods	1
Time period	Hours	9
	Days	12
	Months	26
	Years	24
	Abstract or Unspecified	16
Time spec	Continuous	4
	Discrete	84
Spatial level	Small area	3
	Town or partial city	20
	City wide	18
	Regional	13
	Multi-region	1
	Country wide	6
	Multi-country	10
	Abstract	19
Input network level	A	34
	C	64
	WN	5
Output network level	A	4
	C	55
	WN	46

Table 2.2: Summary of characteristics of reviewed papers

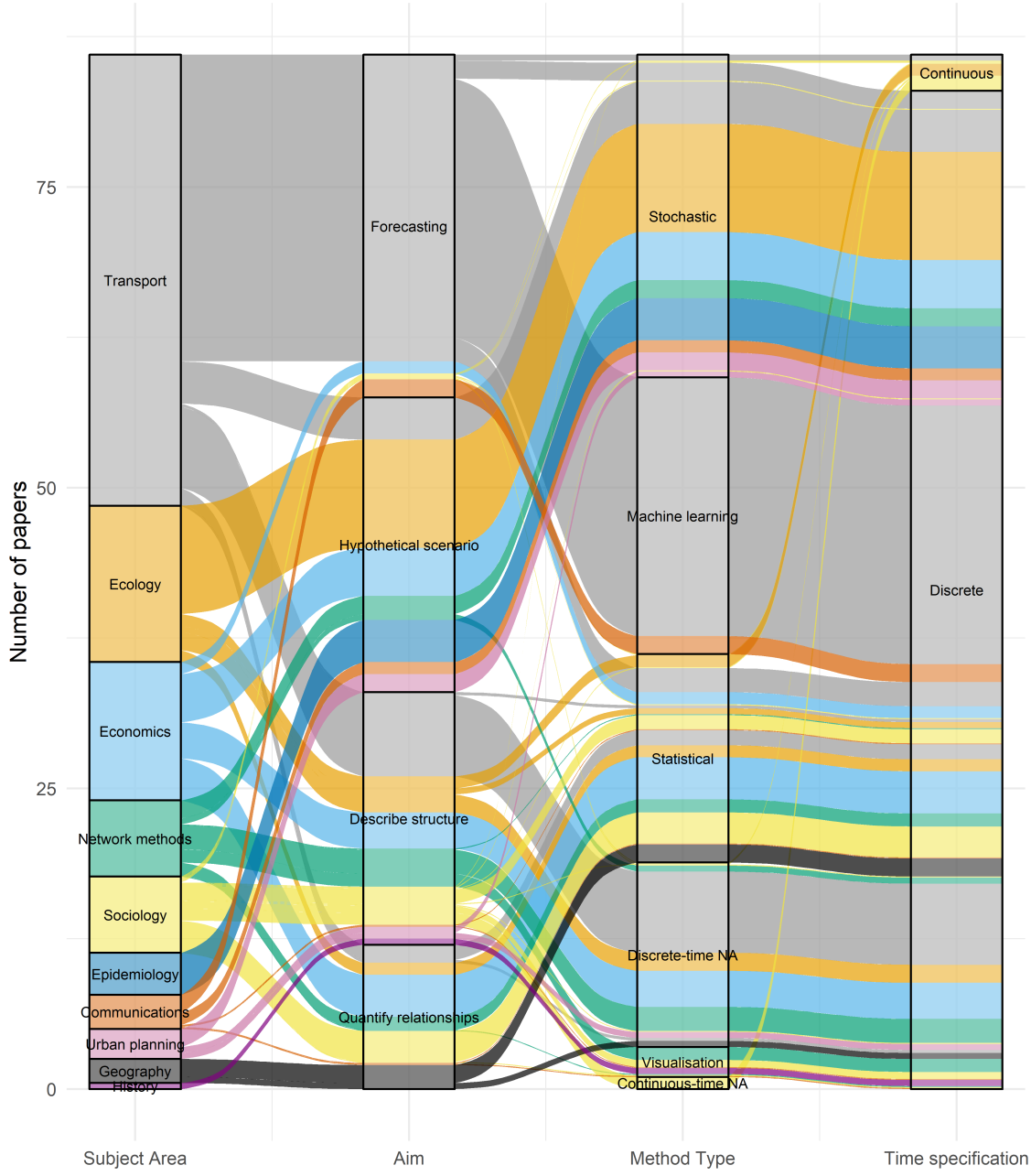


Figure 2.2: Correspondence between subject area, aim, method type and time specification of reviewed papers

some, like Carter et al. (2020), used real spatial settings. Stochastic methods were also the only category that included information about the actors in the network as an output – this included prediction of individual level traffic behaviour and individual infection in an epidemic model embedded in a spatial transport network (Ferenc et al. 2016; Y. Guo et al. 2020; Perez and Dragicevic 2009).

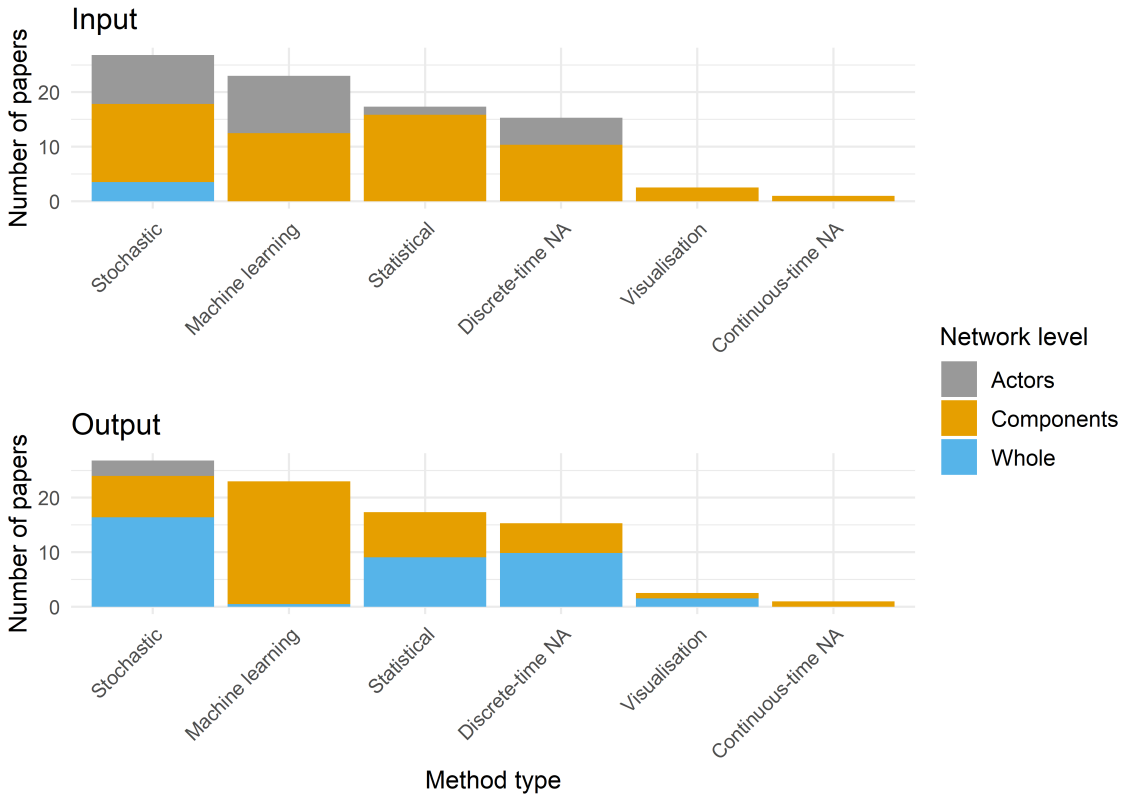


Figure 2.3: Relationship between method type and the network level of input and output information in the reviewed papers

Figure 2.4 outlines the correspondence between the network level of input information in the analysis and the network level of output information. More low level information (actors and components) was used as an input, with a greater proportion examining whole-network information as an output. Of those papers beginning with low level information about actors moving through the network, most had an output only one level higher – at the component level. These papers looked at how the properties of individual objects moving through the network affected the properties of individual links and vertices in the network. For example, how the behaviour of lamprey fish and their larvae moving through a river network affects the properties of individual reaches of the river network (Neeson et al. 2012). Very few papers began with whole-network level information. Those that did

often used this in conjunction with other lower level information. Frequently this was to assess the impact of structural changes in the network on network processes occurring at a lower level.

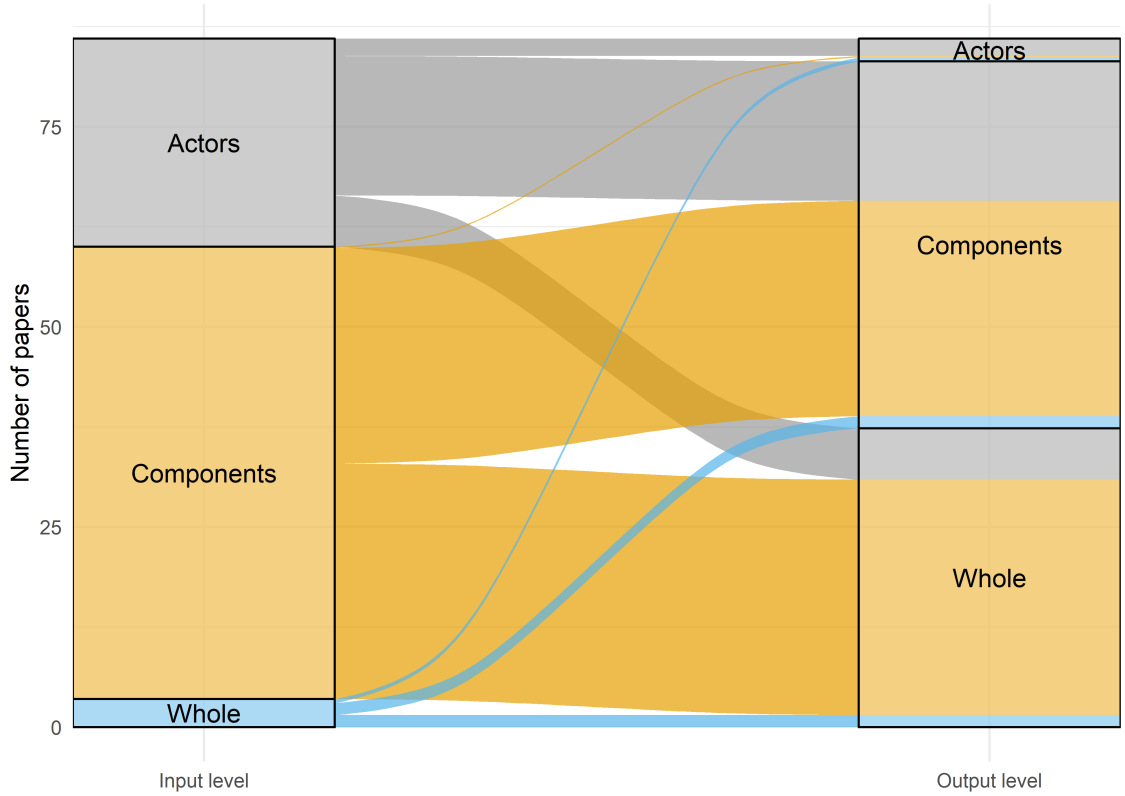


Figure 2.4: Correspondence between levels of input and output information in the reviewed papers

2.5.1 Methods used by reviewed papers

The most common type of method used was the stochastic model, followed by machine learning then statistical models. There are a range of different types of models within each of the categories described.

The reviewed papers used a range of stochastic methods. Two common types were agent-based models and microsimulation (Anderson and Dragicevic 2018). These methods model the properties of 'agents' in a system using information captured on a micro-scale to build up a larger picture of how the system may respond under different circumstances. Agent-based models allow interactions between agents and microsimulations do not (Crooks and Heppenstall 2011). These models were often used to examine the movement of actors within a network.

Some stochastic models were used to model the formation of networks. For example, Iotti et al. (2017) presented a new model to mimic the behaviour of social networks. Semboloni (2000) presented a model for the growth of an 'urban cluster' of cells and roads. Both these models use a series of probability-based rules to govern changes at the level of network components (edges and vertices) at each unit advance in time.

Anderson and Dragićević (2018) introduced the concept of geographic network automata for modelling the evolution and changes in an ecological network with fixed vertices over time. This also included a set of rules that governed edge formation based on connection 'costs' related to distance. Stochastic models are often based on a set of rules at the level of small components within a system. While this could be said to mimic how the properties of real systems develop, poor specification of properties at a low scale can compound to larger inaccuracies and departures from reality as models are run over time (Ward et al. 2016).

23 papers used machine learning methods, with many of these focusing on traffic speed or flow prediction. Generally these used deep learning or graph convolutional neural network approaches (Diao et al. 2019; S. Guo et al. 2019; Yao et al. 2019). Many of the papers presented new models for prediction in spatio-temporal networks which accounted for local and long distance dependencies between locations related to the network structure (L. Chen et al. 2020; S. Fang et al. 2019; Feng et al. 2020).

The most common statistical model in the papers reviewed was the dynamic spatial panel data model. This includes an autoregressive component, as well as effects of an explanatory variable and spatially and temporally local effects. Terms with explanatory variable effects are lagged over space and/or time (Basile et al. 2018). Spatial relationships are specified using a spatial weight matrix. Other models based on autoregressive and moving average processes, such as STARIMA and space-time autoregressive models, were also included in the reviewed papers (Burnett and X. Zhao 2017; T. Cheng et al. 2014). Instead of modelling temporal patterns as a continuous function of time, these models estimate the effect of previous and nearby measures of a variable (or model errors) on the current measure in a given location. These models can include network-based spatial relationships (Basile et al. 2018).

Another common method was multilevel or mixed models. These model data in a hierar-

chical format where observations are grouped in 'clusters' between which model coefficients are allowed to vary randomly (Goldstein 2011). These models can be extended to incorporate complex correlation structures (Goldstein 1994; Goldstein et al. 1994). They are able to model temporal data with a continuous representation of time but are limited to fitting parametric temporal patterns.

Two papers use stochastic differential equations – a type of continuous-time statistical model. These papers examined the persistence of populations of organisms in a river network (Y. Jin et al. 2019) and the adoption of a social media platform in Hungary (Lengyel et al. 2020). Stochastic differential equations can be thought of as a continuous-time equivalent of autoregressive models in which the slope of a model is regressed on its current value (and other covariates) (Oud et al. 2012).

Additional statistical methods included were the gravity model, exploratory spatial data analysis and statistical comparison of observed network properties to those in a random network. Gravity models relate interactions between populations (which could be thought of as vertices) to their size and the amount of separation between them (G. Sun et al. 2017). This separation can be representative of geographical or network space, or other properties. G. Sun et al. (2017) also includes linguistic and cultural similarity, for example. Exploratory spatial data analysis was used to investigate correlations of children's activity types within geographical space (within a classroom) and within a social space (if children who frequently interact take part in the same activity types) (X. Li and Griffin 2013). Jacoby et al. (2012) directly compared a network property derived from shark movement through a network of habitat patches to a distribution of this property from 10000 randomly generated networks representing shark movement.

Papers that used network analysis methods to summarise network properties aggregated observations of temporal networks into time windows, in each of which network properties were summarised (S. Chen, Ilany, et al. 2015; S. Cheng et al. 2018; Jacoby et al. 2012; C. Liu et al. 2018; S. Liu et al. 2018; X. Wang et al. 2016; Wanzenböck and Piribauer 2018; W. Zhang, Z. Tian, et al. 2019). Some papers used methods dealing with clusters or communities within networks. Analyses pertaining to these features included summary statistics describing the degree of clustering within a network (S. Chen, Ilany, et al. 2015), community detection or clustering of network vertices based on connections between them

(C. Liu et al. 2018; X. Wang et al. 2016) and community search, which identifies the community or cluster to which a particular vertex belongs. There are a variety of clustering algorithms used in these papers, both hierarchical and partitional (C. Liu et al. 2018; X. Wang et al. 2016) The community search algorithm proposed takes account of movement of vertices in the network over time – this is discussed in more detail in Section 2.6.4 (Y. Fang et al. 2019).

One paper used an unusual approach to summarising flow (within time windows) through a very densely connected network that was difficult to summarise and visualise: vector coarse graining (W. Zhang, Z. Tian, et al. 2019). This method used an algorithm to determine the dominant direction of flow in to and out of each vertex, which reduced the information that needed to be displayed.

Two papers used a process to assign trips or calculate potential paths through a network based on constraints and activity schedules. In one paper this was used as part of a stochastic simulation and the expected time to complete the trip was also calculated (Abdelghany et al. 2001). In the other, a set of potential routes through a public transport network was generated based on an origin, destination, time budget and the current state of the transport network (S. Cheng et al. 2018).

Most papers approached visualising spatio-temporal information by aggregating data into time windows and plotting static network maps, which can leave out important temporal features that occur between snapshots. Some papers included more unusual or novel methods for visualising spatio-temporal networks. These include alluvial diagrams to visualise temporal clustering (C. Liu et al. 2018) and spatio-temporal link segmentation (T. Li and Liao 2016). These are discussed in more detail in Section 2.6.4.

2.6 Discussion

2.6.1 Context

The papers included covered a range of research questions across a broad variety of subject areas, ranging from more traditional applications of network methods such as transport and social networks to areas such as epidemiology and history. Networks often provided a context for analyses to be carried out in, for example, they were used to describe rivers and

other ecological settings in which animal or plant species completed life cycles (Neeson et al. 2012) and to describe an urban area in an epidemic model (Perez and Dragicevic 2009). The integrated approach of analysing individuals in the context of their environment, which may be represented as a network, allows for the impact of the changing environment on individual behaviour and vice versa (Anderson and Dragicevic 2018; Anderson and Dragićević 2018; Carlsson and Kindvall 2001; Neeson et al. 2012).

In the papers included, network structures appear in several different forms. Many papers use networks to represent tangible structures, for example, public and road transport networks, river networks and electrical supply networks (Neeson et al. 2012; Neto et al. 2016; Perez and Dragicevic 2009). These often determine the possible paths of movement of units of interest, for example vehicles in a transport network. In some such papers, information at the level of the actors moving through the network was also included, for example, in agent-based models simulating the movement of lamprey and their larvae through a river network.

Another tangible form of network that appeared in the papers is boundary sharing of geographical areas (Dubin 2009). Often this is included with a view to describing spatial proximity of areas but can be thought of as including areas as vertices with shared boundaries forming edges between them. Many papers using spatial panel data models included this form of correlation to allow 'spill-over effects', i.e. influence of one geographical area on its neighbours and vice versa (Parent and LeSage 2010). Using a stochastic analysis, Semboloni (2000) used geographical 'cells' as an alternative representation of space to simulate the growth of an urban area and its road network. Cells had a network vertex at their centre and could potentially be connected to all adjacent cells by edges. The model advanced in discrete time steps. At each of these steps, rules were used to determine: 1) whether a road was built along an edge, becoming permanently fixed in the model and 2) if a new cell was generated in the centre of a road connecting two existing cells. The propagation of these components influenced each other and lead to the growth of a cluster of cells and roads representing an urban area (Semboloni 2000).

Some papers used network structures to represent non-physical structures – particularly this included connections made between network components by some communication or interactive behaviour between them. For example, connections made by social contact

(Carter et al. 2020; Han et al. 2019), economic trading activity (S. Liu et al. 2018), or connection through the internet or telephone system (Ramachandran et al. 2007a; Ramachandran et al. 2007b). In these networks, the network components are often performing some kind of activity, behaving in certain ways or physically moving to generate connections. For example, S. Chen, Ilany, et al. (2015) investigated the spatial distribution of social interactions among cows, which could move around a large pen and come into close contact with each other in different areas. S. Liu et al. (2018) examined the properties of trade networks between China and Arab States, investigating the properties and subgroups of networks formed by trade relations between different countries in different periods.

2.6.2 Question

The reviewed papers addressed all the different question types outlined in Section 2.4. Forecasting questions were the most commonly addressed and were mostly used in transport literature, making up over half of this subject area. Other subjects covered by forecasting aims were communications, economics and sociology. Predicting traffic speed and flow in transport networks is a common analysis aim and it is not surprising that forecasting questions made up a large proportion of the aims of transport papers.

Many papers asked hypothetical-type research questions or aimed to describe the structure of a network. Both of these types of questions were posed in papers from most of the different subject areas with only sociology, geography and history not including a hypothetical scenario, and epidemiology and geography not including papers that described network structure.

Papers in geography only aimed to quantify the relationships between variables and history papers only aimed to describe the structure of a network. Epidemiology papers only dealt with hypothetical scenarios to do with infection spread (Anderson and Dragićević 2020a; Iotti et al. 2017; Mateus-Anzola et al. 2019; Perez and Dragicevic 2009).

2.6.3 Methods

Different methods were generally used to address different types of questions. Stochastic simulations lend themselves well to 'what if?' scenarios, and this type of aim was almost exclusively addressed by this type of method. For example, Neeson et al. (2012)

used a stochastic model to represent the life cycle and movement of lamprey and their larvae through river systems. They used this simulation with different hypothetical river network structures to investigate the effect of river network structure on the movements and spatial distribution of lampreys. Other topics investigated using hypothetical scenarios included invasive species management (Albers et al. 2018), economic behaviour (Neto et al. 2016), infection dynamics (Perez and Dragicevic 2009) and traffic movement (Xu and Z. Gao 2009). Papers with aims involving hypothetical scenarios often included input information at several different network levels. They frequently included information about a hypothetical setting, for example network component properties or whole-network structure, with information governing behaviour of lower level agents, for example actor movement through a network or the formation of network edges or vertices. This led to some papers, like Neeson et al. (2012), including information about the whole network as an input, which was not seen in papers with other types of aims.

Exploring and describing the structure of systems was often carried out using network analysis methods, most commonly with a discrete-time specification. Frequently, information about networks over time was aggregated into discrete time periods and network properties for these time periods were calculated (S. Chen, Ilany, et al. 2015; S. Cheng et al. 2018; Jacoby et al. 2012; C. Liu et al. 2018; S. Liu et al. 2018; X. Wang et al. 2016; Wanzenböck and Piribauer 2018; W. Zhang, Z. Tian, et al. 2019). Some papers also went on to use statistical methods to explore these network properties or relate them to other variables (addressing a different type of aim) (Han et al. 2019; Jacoby et al. 2012). Visual analyses were also used to describe the structure of network systems. These were most frequently used as part of a discrete-time network analysis, but one paper used only visual analysis to describe the spread of a particular method of schooling across Spain (Caruso 2007). Two other papers also introduced novel visualisation methods to describe temporal properties of a network (T. Li and Liao 2016; G. Sun et al. 2017).

Papers investigating the relationships between variables mostly used statistical methods. Often the aim of these analyses did not directly involve a network but required network relationships between units of analysis to be accounted for. For example, Basile et al. (2018) examined the relationship between export diversification and economic development in countries, accounting for the effect of the properties of each country on their neighbours

(i.e. countries with whom they share a border). This paper used a spatial dynamic panel data model which was a common type of statistical model in the papers reviewed. Other types of analyses included stochastic differential equations (Y. Jin et al. 2019; Lengyel et al. 2020) exploratory spatial data analysis with social weights (X. Li and Griffin 2013) and the correlation of temporal profiles of vertex properties with local land use information (C. Liu et al. 2018).

Forecasting questions were addressed in the papers by stochastic, statistical and machine learning methods. Of the 29 forecasting papers, 26 aimed to predict traffic flow, speed or movement through various transport networks (Abdelghany et al. 2001; L. Chen et al. 2020; T. Cheng et al. 2014; Diao et al. 2019; S. Fang et al. 2019; Feng et al. 2020; Ferenc et al. 2016; S. Guo et al. 2019; He et al. 2020; G. Jin et al. 2020; G. Li et al. 2020; W. Li et al. 2021; Z. Li et al. 2020; Lin et al. 2020; Ou et al. 2020; Peng et al. 2020; Qu et al. 2020; B. Sun et al. 2021; K. Tian et al. 2020; B. Yang et al. 2020; Yao et al. 2019; C. Zhang et al. 2019; X. Zhang et al. 2019; Z. Zhang et al. 2019; Zheng et al. 2020; F. Zhou et al. 2020), with the remaining three predicting electricity prices, cellular (telephone) network traffic, and adoption of a new social media platform (Burnett and X. Zhao 2017; Lengyel et al. 2020; N. Zhao et al. 2020). Machine learning methods were the most commonly used for forecasting aims, and also went unused in papers with other aims. They focused on deep learning methods such as graph convolutional neural networks (Diao et al. 2019; S. Guo et al. 2019; Yao et al. 2019). A number of papers aimed to capture dynamic spatial relationships between network locations, including many of the machine learning papers and T. Cheng et al. (2014) who introduced a variant of the STARIMA model with a dynamic spatial weight matrix that depended on current traffic conditions.

2.6.4 Time

Discrete or continuous specification

A large majority of papers used analyses that treated time as a series of discrete steps. The number and size of these steps varied considerably between papers. Papers using stochastic simulation methods often tended to use many small time steps, whereas discrete-time network analyses tended to use data aggregated into a few much larger time periods. The majority of the statistical models included used discrete time specifications.

There were four papers that dealt with time in a continuous specification. Abdelghany et al. (2001) used a stochastic simulation with discrete time steps but also incorporated some continuous-time information. This was in the form of travel times between two points in the system, estimated for a particular agent at a specific time. The travel time depended on the current state of the network at the departure time.

It is possible for stochastic simulations to use continuous-time information. Stochastic simulations often move time forward in discrete steps with transition probabilities between states defined for steps of a constant size (P. Jones 2017). For example, a simple Markov chain with four possible states (A, B, C and D) is defined by a set of 16 transition probabilities. For each state A-D, these describe the probability of changing to state A, B, C or D after one time step (P. Jones 2017). A continuous-time Markov process is specified differently and allows the time between state changes to vary continuously, rather than being restricted to a whole number of time steps. This defines the transition between states in two ways: the probability of transitioning at time t (after entering the current state) and the probabilities of the transition being to each of the other possible states (Liggett 2010). The difference between the specification of time in these models can be thought of in the following terms: continuous-time Markov processes answer the question "when did the state change take place?" whereas discrete-time Markov chains check the state at regular intervals and answer "has the state change occurred yet?".

While the specification of continuous-time in a Markov process is not overly complex, the example above is extremely simple. Agent-based models and microsimulations are far more complex stochastic simulations. If the time intervals for the stochastic simulation are very small, the difference between the answers to the questions "when did the state change take place?" and "has the state change occurred yet?" may be extremely small. In cases such as these, adding a further layer of complexity to an already complex simulation may not provide enough benefit to be justified.

Y. Fang et al. (2019) used a continuous-time approach to spatial aware community search. Community search algorithms aim to find the community or cluster to which a particular vertex of interest in a network belongs, based on the structure of the network surrounding it. Spatial aware community search includes information on the geographical location of vertices, as well as the edges that connect them (Y. Fang et al. 2019). This process

aims to identify a community for a vertex in which all vertices satisfy a minimum degree requirement (i.e. all have more than a certain number of connections) and the spatial circle containing all the vertices is minimised (Y. Fang et al. 2019). In Y. Fang et al. (2019), network of connections and location of vertices can change over time. At each change, the community for a specific vertex may change. Y. Fang et al. (2019) describes an algorithm that approximates running a spatial aware community search algorithm at every change to the network or location of vertices, resulting in a list of all communities to which a vertex of interest belonged within a given time period. While the approach here aims to treat time continuously (i.e. identify new communities at each change, rather than at set intervals), the outcome is aggregated over the time period of interest, so does not give an indication of how community membership changes over time (Y. Fang et al. 2019).

Lengyel et al. (2020) and Y. Jin et al. (2019) made use of stochastic differential equations to describe and predict changes in networks or populations existing within network structures. Stochastic differential equations relate the slope or change in a variable to its value an other covariates at a given time (Meredith and Tisak 1990). They can be used as a continuous-time alternative to autoregressive models but can be complex to specify and interpret (Niezink et al. 2019).

Discrete-time models may be easier to specify than continuous-time models. The division of time into a discrete variable makes it easier to specify non-parametric temporal variation in properties and also in model parameters – for example, localised (L)-STARIMA allows the spatial weight matrix to vary over time (T. Cheng et al. 2014). However, in reality time functions in a continuous fashion – it can be broken down into infinitely small intervals. When applying discrete-time models, we must aggregate observations into time windows. The effect of this is most noticeable where data are measured irregularly in time, rather than following a strict schedule. In this case, to aggregate data into discrete time intervals means we lose some accuracy and detailed information in our measurements of time (J. Freeman 1989; Weiss 1984). Aggregation of data into temporal units can mask high frequency temporal variation (J. Freeman 1989). It has also been shown that aggregation can mask lower-frequency variation in autoregressive integrated moving average models (Rossana and Seater 1995). Aggregation can result in biased results from autoregressive

models which may fail to capture the temporal dynamics of the process of interest. (de Haan-Rietdijk et al. 2017; Hwang 2000).

For any analysis using discrete time intervals, the length of these intervals must be chosen. Ideally, this decision would be based on pre-existing knowledge of the process being analysed, but where data are measured regularly this interval is often dictated by the measurement interval (J. Freeman 1989). When using secondary data, there is no possibility of choosing this interval before data collection. It has been demonstrated that periodicity and autocorrelation parameters change in response to the size of temporal intervals (Hawawini 1978; Hwang 2000). If the intervals reflect the measurement process more than the processes of analytical interest, they may be less adept at investigating them (Comber and Wulder 2019). In addition, several papers have identified that systematic sampling over time can introduce bias to results from autoregressive models, in a similar way to temporal aggregation (W. Wei 1981; Weiss 1984).

While continuous-time models can be difficult to fit and specify, including time as a continuous variable would circumvent these issues by allowing temporal processes to be specified independently to measurement intervals and any temporal unit choices, and remove the need for temporal aggregation. This may allow them to better reflect the processes of interest.

Visualisation of temporal properties

Visualisation is often an important way of communicating information and can greatly aid its interpretability. Most of the papers that visualised information about temporally dynamic networks did so using a series of snapshots – images of the network at different, cross-sectional time points. This is a useful and easily interpretable way of showing network evolution, however, it does exclude a large amount of information available about the network – that from any time between the snapshots shown – in the same way that discrete-time network analysis does (Blonder, Wey, et al. 2012).

Several papers used different methods to visualise temporally dynamic properties of networks. G. Sun et al. (2017) aimed to visualise the diffusion of information across a social network over time and space. To this end, it introduced a new visualisation tool, 'SocialWave', that aimed to combine spatial and temporal visualisations. This tool produced

a multipaneled 'dashboard' showing the static spatial network diagrams for several time snapshots and temporal trends of a whole-network property in a different plot. The network diagrams included a method for visualising common words associated with the topic of interest at each location – this combined word clouds and modified circular bar charts (G. Sun et al. 2017). These particular methods, however, still visualised spatial and temporal patterns separately.

Z. Zhang et al. (2019) created a three-dimensional visualisation of the predicted speed on different road segments in a network over time. This visualisation was confined to only a ring road (a loop structure), rather than a more complex network, but still some information for particular segments was obscured. This highlights one of the shortcomings of this type of three-dimensional spatio-temporal visualisation method: it often results in obscured information for parts of the space or network of interest.

C. Liu et al. (2018) used an alluvial diagram to show changes in the community structure of a network of rail stations over one day. This is the same type of diagram used in Figures 2.2 and 2.4. In C. Liu et al. (2018), data were aggregated into hourly periods and community detection was carried out for each hour. The alluvial diagram colours 'flows' by their cluster in the first time period and shows how proportions of stations belonging to each initial cluster move between different clusters in the subsequent hours. This gives an overview of changes in the number of clusters over time and helps identify communities that remain fairly constant throughout the whole day.

T. Li and Liao (2016) introduced another visualisation tool that aimed to enable visualisation to temporally dynamic edge properties. This was carried out using 'link segmentation' in which the links drawn to represent edges in a static network graph were divided into time segments. Each segment was coloured to represent the value of an edge property at a given time. To represent multiple edge properties at the same time, segments were further subdivided, with each subdivision representing one property. This paper shows an interesting and potentially useful visualisation idea, but it has some shortcomings. In a spatial context, it is likely that the length of edges of links in a network diagram with vary. This means that when these are segmented, the scale for a unit of time will differ for different edges. This may make comparing temporal patterns between different edges difficult. In addition, the subdivision of each link segment to represent multiple properties

is carried out in the same axis as the temporal link segmentation. This may produce further confusion if it is not clear whether a boundary between segments separates two time segments or two different properties during the same time segment.

Colak et al. (2013) also includes an alternative representation of temporal properties of network edges. In this case, the paper monitors the traffic volume on different road segments in a stochastic simulation of the movement of vehicles through a road network. Each road segment has a particular capacity and the paper defines congestion as occurring when a road reaches 80% capacity. Instead of showing the whole temporal pattern of traffic volume for each road segment in a map of the network or showing traffic volume for each segment at different time points, Colak et al. (2013) represents the time at which congestion occurs for each road segment on the map using different colours. Defining a specific feature of the temporal pattern of traffic volume for each road segment that is of interest, i.e. when traffic volume reaches 80% of capacity, produces a single piece of numeric information that can be represented simply, interpreted easily and that provides information of interest about the temporal properties of network components. An approach that focuses on a specific event of interest as part of a temporal pattern may encourage researchers to fix specific aims for analyses before carrying them out – this may discourage practices like p-hacking which are associated with a lack of clear research aim (Head et al. 2015; Simmons et al. 2011).

All of the visualisations of spatio-temporal properties discussed here used discrete representations of time. Note however that the method in Colak et al. (2013) could be used to represent the results of continuous-time models. Only two papers, discussed in Section 2.6.4, used continuous representations of time in analysis.

2.7 Summary

This review has identified that analyses of spatio-temporal networks appear in many subject areas – networks can be used to represent a large range of processes and systems with different applications. A variety of different methods were used, mostly to address different types of questions.

Overwhelmingly, the analyses in the reviewed papers specified time in a discrete form,

with only two papers including continuous-time information. Using discrete-time analyses requires the specification of time steps of a certain length by the researcher – this is an important choice and should reflect the underlying process being examined to avoid introducing bias with too long or too short time steps (Comber and Wulder 2019; J. Freeman 1989). However, often the time steps are chosen based on the data measurement scheme, or irregularly measured data are aggregated into discrete temporal units which introduces additional bias (J. Freeman 1989; W. Wei 1981; Weiss 1984). There are continuous-time models that could potentially be used in spatio-temporal network context, for example, multilevel models or stochastic differential equations, but these did not appear in the reviewed papers. Future work should explore the application of continuous-time models to spatio-temporal networks as they provide a more realistic representation of the continuous nature of time and remove the need for temporal aggregation or choosing of temporal unit lengths.

Visualisation of spatio-temporal networks was mostly conducted using a series of 'snapshot' static network maps at discrete time intervals. Some papers introduced novel visualisations, but these were not always clear to understand. One paper visualised the timing of an 'event' or 'pattern feature' which produced a concise and clear plot (Colak et al. 2013). This approach may be a useful method for simplifying visualisation of spatio-temporal patterns and may encourage the focus of analyses on a particular question, discouraging practices such as p-hacking that are associated with a lack of clear analysis aims (Head et al. 2015; Simmons et al. 2011).

Chapter 3

Overview of Continuous-Time Longitudinal Analysis Methods

3.1 Introduction

This chapter describes a range of models used in epidemiological and geographical literature with the aim of capturing temporal patterns of variables observed repeatedly in individuals, objects, or systems. For example, these methods may be used to study growth patterns (over time) in a group of children. In this situation, the children are the units of observation and each child's growth pattern would be the summary temporal pattern of interest. In a network, the temporal pattern of interest may be the summary pattern of a variable associated with a network edge (for example, flow), or it may be the summary pattern of the properties of a network vertex (for example, the number of passengers in a rail station).

An overview of each method is included, along with a discussion of some strengths and weaknesses. The methods examined are multilevel models, methods that condition on the outcome, latent growth curve modelling, functional data analysis and stochastic differential equations. Further investigation into the methods that condition on the outcome is outlined in Chapter 6.

These methods are all capable, or presented as being capable, of examining time as a continuous variable. Most of the papers examining spatio-temporal networks identified

in Chapter 2 used methods that treated time as a series of discrete steps. However, in reality, time is a continuous process, that may be broken down into infinitely small units. Analysing time as a discrete variable circumvents many of the challenges with parameterising complex continuous temporal patterns. However, if observations are not measured at strict intervals, then it is necessary to aggregate into discrete time windows (J. Freeman 1989), resulting in a loss of information and accuracy that could affect model results, as discussed in Section 2.6.4 (de Haan-Rietdijk et al. 2017; Hwang 2000; Rossana and Seater 1995; W. Wei 1981; Weiss 1984).

3.2 Multilevel models

3.2.1 Single level regression models

Regression models are a common tool for statistical analysis. They are used to quantify the relationship between two variables, potentially while controlling for one or more other variables (Wolf and Best 2014a).

For example, to model temporal patterns of growth in children, one might use data on the height of ten children, measured in centimetres eight times between the ages of four and 15 years. A simple regression model for height regressed on age for these data might be specified as in Equation 3.1.

$$\begin{aligned} h_i &= \beta_0 + \beta_1 t_i + \epsilon_i \\ \epsilon_i &\sim N(0, \theta^2), i.i.d. \end{aligned} \tag{3.1}$$

In this model, each observation is indexed by i. h_i represents height and t_i represents the age of the child for each observation. β_0 is the intercept term. This represents the average value of height at age 0. It is possible to transform the data so β_0 represents the average height at a different age by recentring t_i – this would be useful in this case as age zero is not included in the data. We may centre t_i (and set the model intercept) at age 4, by subtracting 4 years from each age recorded. β_0 would then tell us the average height of a child aged 4 years old (Wolf and Best 2014a).

β_1 is the coefficient for height. This tells us how much height increases on average, for

each unit increase in age (Wolf and Best 2014a). For example, if β_1 is equal to 4, this means the children grow an average of 4cm per year. In this model, the increase in height over time is set to be a straight line – in reality perfect linear growth is very unlikely, as children do not grow at a constant rate. It is possible to fit more complex models to account for this – these will be discussed in Section 3.2.4.

The final term, ϵ_i, is the residual term for observation i. This is the residual difference between the height predicted by the model at age t_i and the observed height h_i (Wolf and Best 2014a). The residual term captures all possible reasons why the model did not perfectly predict height, including all exogenous causes of height, random variation, imperfect measurement, and imperfect model fit. A number of assumptions are made about the residual terms for a robust single level regression model: they follow a normal distribution with mean zero, the standard deviation of this distribution is not related to age, and each residual term is independent of all other residual terms (Wolf and Best 2014c). The abbreviation *i.i.d.* indicates that the residual terms are independent and identically distributed, that is, they are unrelated to each other and all from the same distribution.

In order to examine the temporal patterns of a variable in some unit of observation (e.g. individuals), that variable must be measured repeatedly for each unit. However, repeated measurements within the same unit of observation will not be independent, since repeated measures within the same unit will be more similar to each other than to measurements outside that unit. Therefore, the assumption of independent residual terms made by simple regression methods does not hold for data containing repeated measures within the same unit (Ugrinowitsch et al. 2004). Using simple regression methods to analyse such data results in underestimation of the standard errors of parameter estimates, resulting in misleadingly small uncertainty for each estimate (Ugrinowitsch et al. 2004). In these cases, we should use a modified regression modelling approach to account for the similarity of repeated measures within observational units.

3.2.2 Basic multilevel models

Multilevel models are a type of regression model that accounts for clustering of observations, making it a useful tool for modelling temporal patterns within observational units.

This is achieved by grouping the data into a hierarchical structure so that the repeated measurements within each unit of observation are grouped into unit-level 'clusters' (Goldstein 2011). Parameters for the model can be allowed to vary randomly between the clusters (units of observation), with one set of residuals recording the deviation of the mean parameter for one unit of observation from the overall estimated mean parameter for all units. A second set of residuals record the deviation of each observed measurement from the estimated value for their particular observational unit (Goldstein 2011). The simplest form of multilevel model that could be specified for these data is known as a 'random intercept model' (Goldstein 2011). This allows the intercept of the model (β_0) to vary randomly between units. This model is outlined in Figure 3.1 and Equation 3.2, with the fixed effects and random effects grouped by parenthesis for clarity.

$$\begin{aligned} h_{ij} &= (\beta_0 + \beta_1 t_{ij}) + (u_{0j} + \epsilon_{0ij}) \\ u_{0j} &\sim N(0, \theta^2_{u_0}), i.i.d. \\ \epsilon_{0ij} &\sim N(0, \theta^2_{\epsilon_0}), i.i.d. \end{aligned} \tag{3.2}$$

In our example, j in Equation 3.2 represents an index for each child and i an index for each measurement within each child. The overall intercept is represented by β_0, and each child's average deviation from this intercept is represented by the residual term u_{0j}. The slope is fixed to the same value, β_1, for each child. Finally, all additional deviation between the height predicted for child j at age t_ij and the observed height h_ij is represented by the residual ϵ_{ij}. This model allows for correlation between observations for the same child due to them having a higher or lower average height than other children, but not due to them having a different growth rate. To allow for this, we can also allow the slope to vary between children by adding a residual term for the age coefficient, β_1, as detailed in Equation 3.3 (Goldstein 2011).

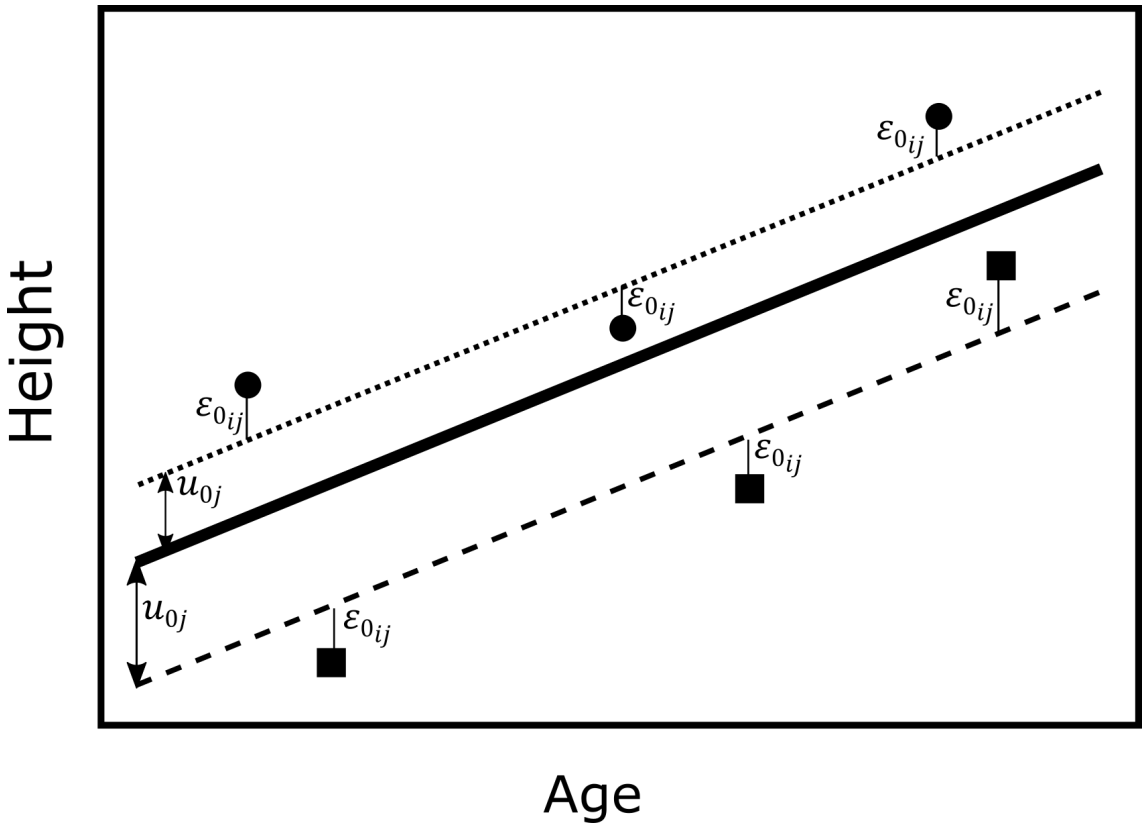


Figure 3.1: Graphical representation of a random intercept model

$$
\begin{aligned}
h_{ij} &= (\beta_0 + \beta_1 t_{ij}) + (u_{1j} t_{ij} + u_{0j} + \epsilon_{0ij}) \\
u_{1j} &\sim N(0, \theta^2_{u_1}), i.i.d. \\
u_{0j} &\sim N(0, \theta^2_{u_0}), i.i.d. \\
\epsilon_{0ij} &\sim N(0, \theta^2_{\epsilon_0}), i.i.d.
\end{aligned}
\tag{3.3}
$$

In this model the growth rate for each child is comprised of two parts: β_1, which represents the average growth rate for all children; and u_{1j}, which represents the difference between child j's growth rate and the average growth rate for all children. This allows observations for the same child to be correlated with each other due to them having a higher or lower average height than other children and a higher or lower growth rate than other children.

3.2.3 More complex multilevel models

Multiple levels

The multilevel models above include a relatively simple situation where measurements of height are clustered within individual observational units, in this case children. This model assumes that the observational units are themselves independent of each other. This would not hold, for example, if some of the children were siblings. In cases like this, further levels can be added to the multilevel model to allow clustering of the observational units within higher level groups (Goldstein 2011). In the example above, observations would be grouped within children and some children would be grouped together within mothers. Level 0 represents the measurements of height, level 1 represents the individual children and level 2 represents the mothers shared by siblings. Such a model would be specified as in Equation 3.4. The index i represents each observation within each child, j, within each mother, k. The average intercept and slope are represented by β_0 and β_1. Random variation from the average intercept and slope associated with different mothers is captured in the coefficients v_0 and v_1, respectively. Random variation associated with the child is captured in coefficients u_0 and u_1. ϵ represents observation level residuals.

$$
\begin{aligned}
h_{ijk} = (\beta_0 + \beta_1 t_{ijk}) + v_{1k} t_{ijk} + u_{1jk} t_{ijk} + v_{0k} + u_{0jk} + \epsilon_{0ijk} \\
v_{1k} \sim N(0, \theta^2_{v_1}), i.i.d. \\
v_{0k} \sim N(0, \theta^2_{v_0}), i.i.d. \\
u_{1jk} \sim N(0, \theta^2_{u_1}), i.i.d. \\
u_{0jk} \sim N(0, \theta^2_{u_0}), i.i.d. \\
\epsilon_{0ijk} \sim N(0, \theta^2_{\epsilon_0}), i.i.d.
\end{aligned}
\tag{3.4}
$$

This type of nested model is capable of capturing random variation where individuals belong to only one group at each level. Observations of height for the same child are allowed to covary due to them having an average height higher or lower than other children, and a growth rate higher or lower than other children. Observations of height for different children with the same mother (siblings) are also allowed to covary due to the siblings having a higher or lower average height than children of other mothers and a growth rate

higher or lower than children of other mothers.

Cross-classified models

In some situations, the clusters or groups that observations fall into do not have a clear hierarchy. Continuing with the example of measuring the height of children over time, we may wish to group children within residential areas and schools. Schools may contain children from various areas and areas may contain children who attend different schools. This means that areas and schools do not simply fit into a nested heirarchy in the same way as children and mothers do. Models that deal with this type of clustering are called cross-classified models (Fielding and Goldstein 2006; Goldstein 1994). Specifying this type of model involves the inclusion of random effects associated with both school and area, as shown in Equation 3.5.

$$
\begin{aligned}
h_{ijkl} = (\beta_0 + \beta_1 t_{ijkl}) + z_{1l} t_{ijkl} + v_{1k} t_{ijkl} + u_{1jkl} t_{ijkl} + z_{0l} + v_{0k} + u_{0jkl} + \epsilon_{0ijkl} \\
z_{1l} \sim N(0, \theta^2_{z_1}), i.i.d. \\
z_{0l} \sim N(0, \theta^2_{z_0}), i.i.d. \\
v_{1k} \sim N(0, \theta^2_{v_1}), i.i.d. \\
v_{0k} \sim N(0, \theta^2_{v_0}), i.i.d. \\
u_{1jkl} \sim N(0, \theta^2_{u_1}), i.i.d. \\
u_{0jkl} \sim N(0, \theta^2_{u_0}), i.i.d. \\
\epsilon_{0ijkl} \sim N(0, \theta^2_{\epsilon_0}), i.i.d.
\end{aligned}
\tag{3.5}
$$

Here k indexes area and l schools; i indexes observations within children, j. There are random intercept and slope coefficients associated with school (z_0, z_1), area (v_0, v_1) and child (u_0, u_1). In this model, observations of height are allowed to covary if:

1. They are for the same child – the child can have an average height higher or lower than other children and a growth rate higher or lower than other children.

2. They are from different children who share the same school – the school can have an average height higher or lower than other schools and a growth rate higher or lower than other schools.

3. They are from different children who share the same neighbourhood – the neighbourhood can have an average height higher or lower than other neighbourhoods and a growth rate higher or lower than other neighbourhoods.

Some observations from different children will share a school and a neighbourhood and these will covary more strongly than observations that just share a school or neighbourhood.

Complex residual structures

The models listed in the previous sections all specify that residuals at each level follow a normal distribution with mean zero and variance to be estimated by the model (Browne, Draper, et al. 2002; Goldstein 2011). This variance is typically specified to be constant with respect to the explanatory variables in the model (homoskedastic) and residuals for each observation are typically assumed to be independent (Browne, Draper, et al. 2002; Goldstein 2011). It is possible, however, to relax these assumptions by specifying that the model residuals follow a model of their own. One common method for this is to set the variance of residuals to follow a function of time – this introduces heteroskedasticity into the model (Browne, Draper, et al. 2002). Another common specification is that residuals follow a temporal autoregressive structure within observational units, for example, the residual for the current observation of height of a child specified to be a function of the previous measurement and the time elapsed since the previous measurement (R. Jones and Boadi-Boateng 1991). This allows observations that are closer in time to be more similar to each other than observations further apart, which also reflects the underlying data generating process. In addition, covariance of residuals (for example, those describing random effects) can be constrained to be related to some relationship between different units of observation – for example, random coefficients can be allowed to covary according to the spatial proximity of observational units (Besag, York, et al. 1991).

3.2.4 Multilevel models fit parametric shapes

The multilevel models described above fit linear temporal patterns. These are patterns represented by a straight line on a graph of height, for example, over time. Multilevel models can be used to fit more complex shapes, but they must be specified in a parametric form. One option to fit more complex shapes is to use polynomial terms. Polynomial

terms allow the outcome variable to be related to some function of time, for example, time squared or cubed (Wolf and Best 2014b). Polynomial terms can be included in models to build up curvilinear patterns, but their flexibility is somewhat limited because they are restricted to a fixed set of curve shapes. Another alternative is to use fractional polynomials. Fractional polynomial models can include logarithmic terms and non-integer powers (for example, $time^{1.5}$) and provide increased flexibility compared to standard polynomial terms (Royston and Altman 1994).

A further option is to use spline functions. Spline functions divide the time-axis into a series of sections. One or, occasionally, several functions (of the same form with different parameters) are specified that take, for example, a polynomial form in these sections and a value of zero elsewhere. The polynomial curves are combined to form a continuous function with a large amount of flexibility (Perperoglou et al. 2019). There are a range of different types of spline functions – some of these are discussed in more detail in Section 3.6.1 and Section 4.2.5.

When the relationship between time and an outcome is specified in a polynomial and/or spline form, the coefficients become difficult to interpret and relate to a 'real-world' meaning. This means that although these models may fit the data better, they can be extremely hard to interpret and explain, especially to non-statistical audiences.

3.2.5 Estimating multilevel models

In statistical programming packages, simple regression models are often estimated using maximum likelihood estimation. This is a maximisation process that attempts to identify the set of parameters with the highest likelihood of being correct given the observed data and model parameterisation (Browne and Draper 2006). This approach can be used to fit multilevel models, but the maximisation process may fail to converge on a maximum likelihood if the model is too complex, e.g. with cross-classified models.

A common alternative approach is to use Bayesian Markov chain Monte Carlo (MCMC) estimation procedures. Bayesian estimation has different aims to frequentist statistics. While a frequentist approach aims to estimate the 'true' population parameter from a small sample, Bayesian estimation does not assume there is one 'true' population parameter, but a distribution specifying what parameter values are most probable. Bayesian estimation

processes aim to estimate this distribution based on model specification, data provided and prior beliefs about the data, specified in the form of prior distributions for each parameter in the model (Zyphur and Oswald 2015). Many algorithms for Bayesian estimation focus on the probability distributions for parameter values, conditional on the other parameter values (Zyphur and Oswald 2015). They work through a series of samples of parameter values. In each sampling iteration, the value sampled is conditional on values of other parameters in the previous sample (S. Geman and D. Geman 1984; Zyphur and Oswald 2015). The first sample is taken from the prior distributions for each parameter. After a chain of many samples has been taken, a number of samples from later in the chain (after a 'warm-up' or 'burn-in' period) are used to estimate the distribution properties of each parameter (Zyphur and Oswald 2015). Multiple chains with different starting values can be sampled.

There are a range of different algorithms for choosing the next sample in a chain, given the current sample, for example Gibbs sampling (S. Geman and D. Geman 1984; Zyphur and Oswald 2015) or No-U-Turn sampling (Hoffman and Gelman 2014). The choice of sampling algorithm may influence the number of iterations before a chain converges on a stable distribution – if this number is reduced the estimation procedure will be quicker. Often, however, the choice of programming environment will determine the sampling algorithm used. There are a range of programs capable of estimating multilevel models using MCMC procedures: MLWin, MPlus, JAGS, BUGS and STAN (Bürkner 2017; Lunn et al. 2009; L. Muthén and B. Muthén 2010; Plummer 2003; Rabash et al. 2020; Stan Development Team 2019). Each of these has different capabilities in terms of sampling procedure specification, which will thus influence their usefulness when particular model structures need to be specified. MCMC estimation techniques are not limited to use in multilevel models and are frequently used with other modelling strategies.

3.3 Models conditional on the outcome

A number of analysis methods used in the literature attempt to characterise individual exposure trajectories and relate them to a later outcome, but inadvertently capture 'average patterns' for levels of the outcome variable, i.e. conditional on the outcome (Færch et al. 2013; Hulsegge et al. 2017; Oka et al. 2017; Tabák et al. 2009; Tu, Tilling, et al. 2013).

This is problematic because individual trajectories are no longer being modelled, though this is not often realised. Modelling the exposure variable conditional on the outcome also goes against the natural order of events in time and makes it impossible to interpret the results as a cause and effect. The resulting modelled 'trajectories' are often interpreted as patterns of the exposure that lead to (i.e. cause) certain outcome values; however, as detailed below, they constitute nothing more than a series of cross-sectional analyses, i.e. a series of correlations, that derive their existence conditional on each outcome value, and therefore cannot be interpreted as either 'trajectories' or causes of the outcome.

A common graphical analysis involves dividing the data into two groups based on the outcome value; for example, high and low outcome levels. The average exposure values within these two groups are then plotted for each time point at which the exposure was measured (Barker et al. 2005; Færch et al. 2013). This does not use information about individual patterns of the exposure but instead analyses the exposure at each time point separately, generating a series of cross-sectional analyses. This series of analyses cannot be viewed as summarising each individual pattern and therefore cannot be interpreted as a longitudinal pattern of the exposure that leads to (i.e. causes) certain outcome levels (Tu, Tilling, et al. 2013). An equivalent statistical analysis to the z-score plots might be to plot a bivariate correlation between each measurement occasion and the outcome, or the regression coefficient from the model of the outcome on each measurement occasion (perhaps additionally accounting for earlier measurements). In either event, this is explicitly conducting a series of cross-sectional procedures; plotting either the resulting correlations or regression coefficients from each separate evaluation of the outcome for each exposure measurement, does not yield a summary of individual exposure patterns (Tu, Tilling, et al. 2013).

Multilevel models are also sometimes used in a way that conditions on the outcome, with the aim of accounting for clustering of measurements within individuals. (Goldstein 2011). While multilevel models can be correctly used to characterise patterns of the exposure for individuals (Howe et al. 2013), some analyses include the outcome variable as a covariate when modelling the exposure, along with an outcome-time interaction term (Barker et al. 2005; Færch et al. 2013; Oka et al. 2017; Tabák et al. 2009). In addition to reversing the flow of time in this regression approach, the interaction term stratifies the

data into groups based on the outcome value, which is another form of conditioning on the outcome – the same procedure and hence the same problem when plotting the average exposure values in outcome groups. This method, therefore, also cannot be interpreted as generating a pattern of the individual exposures that cause certain outcome values, as it does not represent longitudinal individual pattern characteristics that emerge ahead of the outcome occurring. Problems with this methodology have not been discussed in depth in the literature and are investigated in more detail in Chapter 6.

3.4 Latent growth curve models

Latent growth curve models are used particularly to describe patterns of longitudinal variables over time, for example, patterns of weight throughout childhood (Bollen and Curran 2005). The longitudinal variable under consideration must be split into a series of cross-sectional variables – one for each time it was measured (Meredith and Tisak 1990). For example, if the longitudinal variable is weight measured at birth and at ages 1, 2, and 3 years, four weight variables would be used: birth weight, weight at aged 1 year, weight at aged 2 years and weight at aged 3 years. The longitudinal variable is represented as a function of two or more continuous latent summary variables that together describe the shape of the growth curve (Meredith and Tisak 1990). These might be, for example, intercept and slope variables. These represent the same concepts as the intercept and slope in a standard linear regression model, but the values are allowed to vary between individuals; the intercept represents individual values of the longitudinal variable when time is equal to zero (or at a different time point if this is chosen) and the slope represents how much the longitudinal variable increases or decreases for one unit of time (Bollen and Curran 2005). Latent growth curve models, therefore, can be mathematically equivalent to multilevel models (regression models that account for grouping of data) under appropriate assumptions, but this excludes the use of some particularly flexible forms of the latent growth curve model (Johnson 2014; Tu, D'Aiuto, et al. 2009).

Time need not be included in latent growth curve models as a variable in the traditional sense. Instead it may be used to inform factor loadings that specify the effect of the latent variables on each of the longitudinal variable measures (Bollen and Curran 2005; Tu, D'Aiuto, et al. 2009). For example, if weight from birth to three years is the longitudinal

variable and an intercept and slope variable are included, linear growth can be specified by making the factor loadings from slope to each weight measure equal to the time each measure was taken. This model is represented in Figure 3.2 in the form of a path diagram. In path diagrams for structural equation modelling, measured variables are represented in boxes and latent variables in ovals or circles. Arrows between them show regression models, with the outcome at the point of the arrow. Bidirectional arrows indicate covariance between two variables. The effect of the intercept is set to one for all variables and a residual term also makes up each longitudinal measure. This means that a measurement for individual i at time t is represented by the sum of that individual's intercept value, that individual's slope value multiplied by t and that individual's residual term value at time t. This can be represented in Equation 3.6, which is analogous in form to that for a simple linear regression model (Bollen and Curran 2005; Johnson 2014).

$$w_{it} = \beta_{0i} + \beta_{1i}t + \epsilon_{it} \tag{3.6}$$

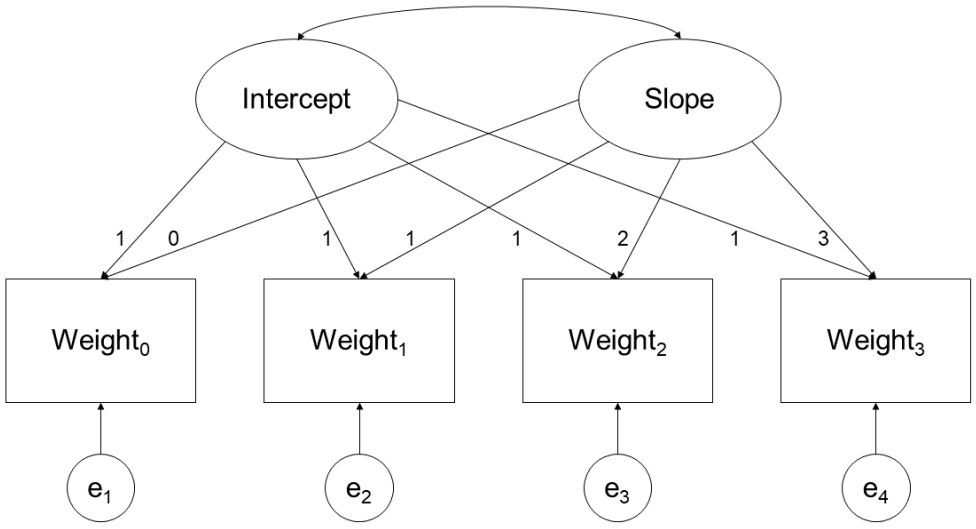


Figure 3.2: Structural equation modelling diagram showing a latent growth curve model

Here t refers to time, w_{it} refers to weight for individual i at a time t, β_{0i} refers to the value of the intercept variable for individual i, β_{1i} refers to the value of the slope variable for individual i, and ϵ_{it} refers to the error term for individual i at a time t. By default, latent

growth curve models typically allow the intercept and slope variables to covary with one another (Byrne and Crombie 2003). In addition, covariance between the residual terms over time can be introduced to represent further autocorrelation of repeated measures or clustering of individuals within a higher level structure, such as within geographical areas (Bollen and Curran 2005). A second variable may also be included in a further part of the model, for example an exposure variable may be allowed to affect the slope variable, or an outcome variable may be a function of it (Tu, D'Aiuto, et al. 2009).

Factor loadings can be changed to some function of time to represent differently shaped growth curves. For example, quadratic growth could be represented by introducing a second slope variable. Factor loadings for the effect of this on each weight measurement would be set to the square of each measurement time (Meredith and Tisak 1990). This can be extended to include splines and other more complex non-linear structures (Bollen and Curran 2005). Latent growth curve models can also be specified so that they estimate a value for some of the factor loadings included, providing two or more are specified as anchor points (Meredith and Tisak 1990). The estimated factor loadings represent the estimated time at which a measurement would have been taken if it were to fit the linear pattern set by the anchor points, i.e. the model 'fits' a linear relationship with respect to the latent slope parameter, but this linear relationship is stretched or compressed in terms of the 'time' between the anchor point and other variables in order to emulate a nonlinear change pattern in linear time. Thus, the time depicted by a factor loading may differ from the time at which the measurement was actually taken, and this difference indicates the amount of deviation from a linear trajectory (Bollen and Curran 2005). This can be used to represent nonlinear patterns of a longitudinal measure without need for complex parameterisation. This type of model has been given various names: shape-factor (Sterba 2014), freed-loading (Bollen and Curran 2005), completely or fully latent (Aber and McArdle 1991; Meredith and Tisak 1990), unspecified (Duncan et al. 2006) or latent-basis latent growth curve model (Grimm et al. 2011). In this thesis, it will be referred to as a latent-basis latent growth curve model.

As described above, latent growth curve models require the division of a longitudinal measure into separate cross-sectional variables (Meredith and Tisak 1990). Often, the assumption is made that these measurements were taken at the same time for all individ-

uals and that the intervals between successive measurements are equal for all individuals, even if successive intervals vary (Aydin et al. 2014). In reality, this is rarely the case. For example, in a longitudinal cohort study, a measurement schedule might be specified to be the same for all individuals, but the timing of the actual measurements might vary somewhat around the proposed schedule. This results in *wave data*, where measurement times are grouped around the times specified by the schedule. It may be that the time interval between each measurement is approximately the same for each individual in the study, for example, if each individual is measured annually, but on different dates to others. Where such intervals are not identically-spaced across all individuals, some error is introduced, though – depending on the patterning – this need not lead to bias. Such variation in measurement intervals between individuals will be referred to as *interval heterogeneity*. Data may also be collected in a fashion that has no measurement schedule, where intervals vary for every pair of measurements for most if not all individuals in the study. For example, routinely collected data from medical records. This will be referred to as *irregularly measured data* in which there is no grouping of measurement times as seen in wave data. Interval heterogeneity associated with wave data and irregularly measured data may be respectively considered as more modest and more severe forms of *sampling time variation*.

Latent growth curve models have been adapted to deal with sampling time variation, and this has been implemented in a range of software (L. Muthén and B. Muthén 2010; Neale et al. 2003). The process involves the use of *definition variables* which specify factor loadings at each measurement for each individual, rather than requiring them to remain the same for all (Bollen and Curran 2005; Mehta and West 2000; Neale et al. 2003). However, descriptions of this have been limited to simple parameterisations of latent growth curve models (Mehta and West 2000). It has been suggested that it would not be possible to fit a latent-basis latent growth curve model on data with sampling time variation as this would require estimation of too many parameters (Blozis, Conger, et al. 2007). However Sterba (2014), describes the process of fitting latent-basis latent growth curve models with individually varying time-points, which can be accounted for using definition variables as factor loadings (Bollen and Curran 2005; Sterba 2014).

When latent-basis latent growth curve models are used, the estimated factor loadings de-

note the time at which a measurement would have been taken, were it to fall on the linear pattern specified by the two points with fixed factor loadings. When the estimated factor loading differs from the real measurement time, this indicates a departure from linearity (Bollen and Curran 2005). Sterba (2014) describes the use of latent-basis latent growth curve models with individual varying measurement times by using a variant parameterisation of the latent-basis model which is outlined in Figure 3.3. The observed data are grouped into time windows throughout the observation period – data within each time window is represented as a variable. The factor loadings for these variables, which would normally be completely freely estimated as part of a latent-basis model, are split into the sum of two parts: a) the real measurement time and b) the departure from linearity for each loading. Part (a) remains fixed while part (b) is estimated. To adapt this to a model with individually varying time scores, part (a) may be specified by a definition variable, and part (b) is estimated while constrained to be the same across all individuals. The anchor loadings may also be specified by definition variables. This allows the estimation of a latent-basis model with individually varying time scores without increasing the number of parameters estimated (Sterba 2014).

The author states that this model can only be used with wave data, which would limit its applicability when analysing routinely collected data, for example. However, a footnote refers to a simulation study that examined the effects on the results from this model when examining the degree of variation of measurement times within windows of grouped data points, and suggests that only a small amount of bias was incurred even when measurement windows overlapped. No further information about this simulation is included in the paper (Sterba 2014). While this model is an exciting development in the application of latent-basis latent growth curve models to data with sampling time variation, the assumption that all departures from linearity are constant within each wave may be unrealistic, as it seems reasonable to expect departures from linearity to vary at the different observation times recorded for different individuals within waves. Keeping this departure constant within each wave does not allow for this and may result in a poorly fitting model for some observational units.

While latent growth curve models with no latent-basis component treat time as a continuous variable, the departures from linearity in a latent-basis latent growth curve model

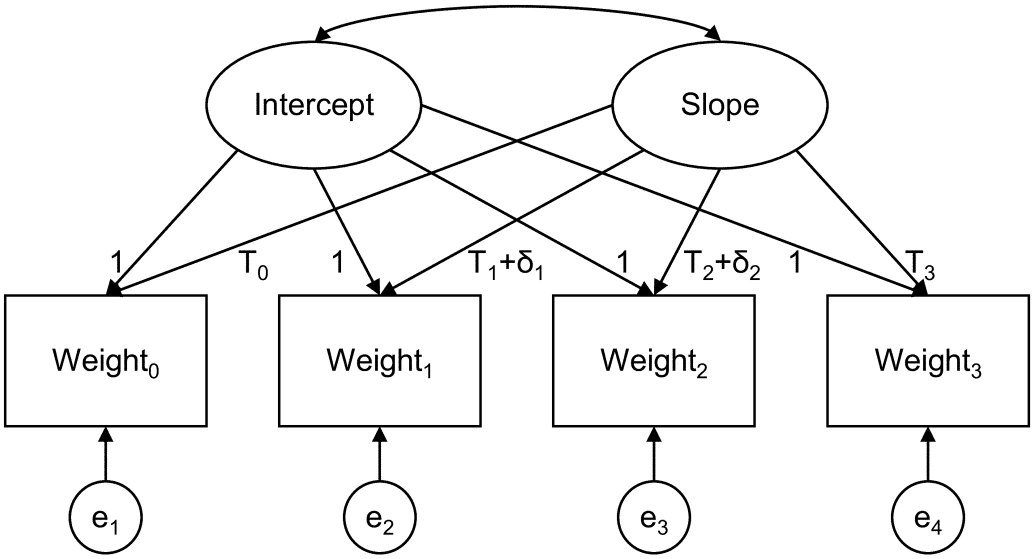


Figure 3.3: Structural equation modelling diagram showing a latent-basis model that accounts for sampling time variation. The variables T_{0-4} are definition variables representing the measurement times and the variables δ_1 and δ_2 are the estimated departures from linearity.

do not follow any function of continuous time – they are non-parametric. Therefore, while latent growth curve models could be used to analyse temporal patterns while treating time as a continuous variable, latent-basis latent growth curve models do not strictly do this – it would not be possible to interpolate between observations from a latent-basis model, as departures from linearity are not generally known, except at each observation point.

3.5 Autoregressive latent trajectory models

There are a number of extensions to latent growth curve models that can be used to represent exposure patterns. Autoregressive latent trajectory models can be used to take account of autoregressive structures in the repeated exposure measurements not captured by the latent parameter variables in standard latent growth curve models, e.g. serial autocorrelation (Bollen and Curran 2004). Autoregressive models are a type of model commonly used for time series data that models each successive observation as a function of the previous observation, plus some residual term. Autoregressive latent trajectory models incorporate this autoregressive structure into the observed variables of the latent

growth curve model as described in Figure 3.4 (Bollen and Curran 2004). The arrows between each successive weight measurement describe the autoregressive structure. Note that the first weight measurement is not, as in the latent growth curve model in Figure 3.2, a function of the latent slope and intercept variables, but correlated with them. This treats the initial measurement of weight as pre-determined which ensures that the autoregressive component of the model is separated from the effect of the latent slope and intercept variables, and the two parts are not conflated (Bollen and Curran 2004).

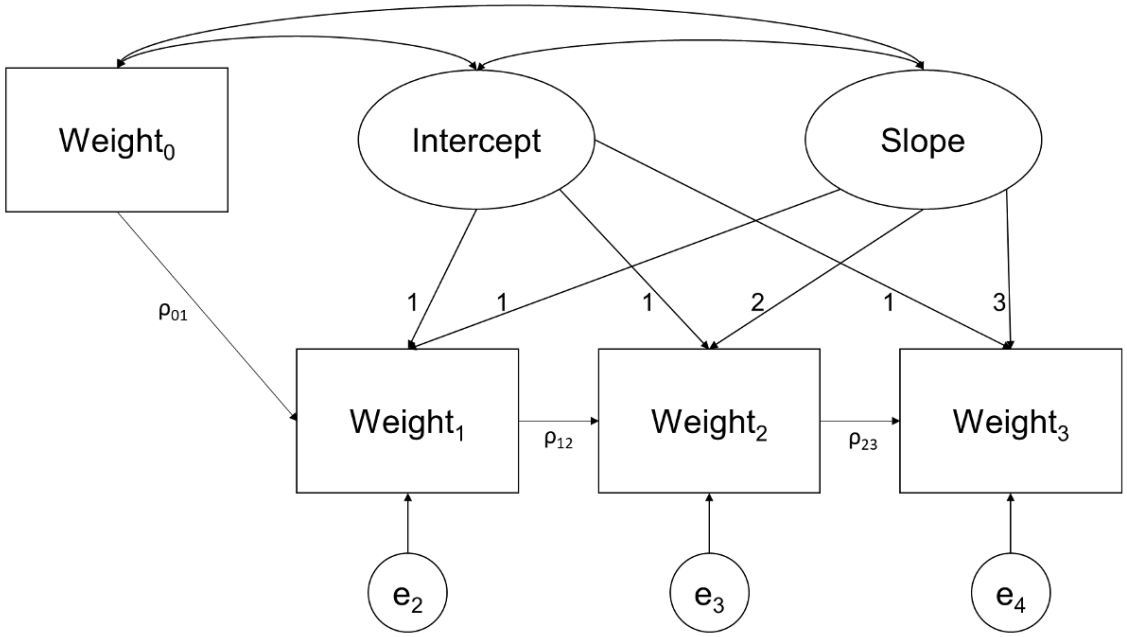


Figure 3.4: Structural equation model of an autoregressive latent trajectory model

3.6 Functional data analysis

Functional data analysis, aims to represent individual repeated measures within a dataset in a 'functional form' i.e. as a model equation. This method uses a series of 'basis functions' as a starting point for these representations. Individual patterns are described by linear combinations of the basis functions that best fit their data (Ramsay and Silverman 1997). The form of basis functions and the length and number of periods for which they are positive must be chosen by the user. This method can capture complex curves in longitudinal data, but the functional form may not have such a clear interpretation as latent parameter variables in latent growth curve model, for example (Ramsay and Silverman

1997). The functional representations of data produced by functional data analysis can be used to study in depth variation in functions between observational units, departures of measured data from the functional forms, or further analysed using methods such as functional regression and functional principle components analysis (J.-L. Wang et al. 2016). Functional data analysis often makes use of derivative functions as a method of displaying variation in functions that might not be apparent without undergoing this transformation (Ramsay and Silverman 1997).

3.6.1 Basis functions

There are a range of possible basis functions used in functional data analysis, each suitable for different purposes. The choice of basis function is a key aspect of functional data analysis – basing this choice on the characteristics of the data is likely to lead to a better representation of the functions of interest (Malaba et al. 2017; Ramsay and Silverman 1997). Two commonly used basis functions are Fourier basis systems and b-spline basis systems.

Fourier basis systems are suitable for capturing functions with a periodic element, for example, seasonal patterns in temperature at weather stations over a number of years (Ramsay and Silverman 1997). They include a series of sine and cosine functions of decreasing period (Ramsay and Silverman 1997). The periodicity of the functions can be chosen by the user. Fourier basis systems perform best when used to capture stable functional forms with limited local departures from the regular periodic shape (Ramsay and Silverman 1997).

Spline basis systems are suitable for capturing non-periodic patterns. Spline basis functions include a series of polynomial functions that are positive in a short section of the observation period and equal to zero in the rest (Perperoglou et al. 2019; Ramsay and Silverman 1997). The length and number of the positive sections is defined using a series of 'knots'. The order of the polynomials must also be chosen, for example, whether they are quadratic, cubic or quartic. At each knot, the polynomials and their derivatives (up to a certain order) must join smoothly (by taking equal values at the knots that join them) (De Boor 2001). The number of derivatives that must join smoothly is the order of the polynomials minus two. For example, if the polynomials chosen are of order four (cubic),

then the first and second derivatives must join smoothly (De Boor 2001).

B-spline basis systems are a popular form of spline basis. Rather than being positive only in between two knots, b-spline basis functions overlap. For a b-spline basis system of order n, each basis function is positive over a maximum of n adjacent intervals between knots, with a reduced number of intervals covered by functions at the tail ends of the basis system (De Boor 2001; Perperoglou et al. 2019). At the ends of the basis system, instead of placing a single knot, a b-spline basis system places n knots (De Boor 2001; Perperoglou et al. 2019). Placing multiple knots at a single point makes the spline function discontinuous (De Boor 2001; Perperoglou et al. 2019). For, example, if two knots are placed in one location in a cubic spline system, the second derivative of the spline function will no longer be continuous at this knot location. Placing n knots means that the spline function itself is no longer continuous, meaning that it does not extend continuously beyond the boundaries of the data (De Boor 2001; Perperoglou et al. 2019).

The choice of knot positions is important in accurately capturing the functional form of the data. A common choice is equally spaced knots, but this can cause problems if data are not equally spaced (for example, throughout time) as coefficients for some spline functions will be estimated using very little data (Ramsay and Silverman 1997). Alternatively, knots can be placed based on the location of observations. For example, if observations are made at different time points, they might be placed at the quartiles or quintiles of the observed time points. This includes an equal number of measurements for each spline section (Ramsay and Silverman 1997). Knots can also be placed based on underlying knowledge of the data generating process of a system, for example, when a particular event that affects the observed data is expected to occur (Ramsay and Silverman 1997).

3.6.2 Analysing functional data

Once a basis system is chosen, the linear combination of basis functions to best fit the data for each observational unit is computed (De Boor 2001; Perperoglou et al. 2019; Ramsay and Silverman 1997). This can be done using several methods, for example, by applying roughness penalties or through least squares estimation (De Boor 2001; Ramsay and Silverman 1997). It is possible to fit models to the residuals estimated along with the smooth functions to allow for heteroskedasticity or residual autocorrelation (J.-L. Wang

et al. 2016).

The coefficients from functional data analysis do not necessarily carry a real-world interpretation. However, functional data can be further analysed using methods such as functional linear regression or functional principle components analysis (Locantore et al. 1999; J.-L. Wang et al. 2016). These methods seek to examine the variation in functions between observational units and how this relates to some exposure characteristic(s). In addition, analyses that allow similarity in functional data curves associated with spatial proximity have been developed. For example, Delicado et al. (2010) present an analysis in which an average functional curve of temperature for a country is calculated based on a number of functional data curves at weather stations, taking into account their spatial locations.

3.7 Stochastic differential equations

Stochastic differential equations are a type of model used to analyse temporal patterns of properties. They have been described as the "continuous time version of autoregressive models for time series" (Niezink et al. 2019). Autoregressive models describe the relationship between pairs of sequential observations in a time series. This is inherently linked to the time difference between the two observations which means that coefficients from the model are specific to the time steps between observations observed (or chosen, if data are aggregated). This reduces the extent to which models with different intervals can be compared (Niezink et al. 2019; Voelkle et al. 2012).

Stochastic differential equations are based on ordinary differential equations. In particular they use differential equations that, for a property x at time t ($x(t)$), regress the first derivative of $x(t)$ with respect to time ($\frac{dx(t)}{dt}$) on the current value of $x(t)$ and other potential covariates of interest, represented in Equation 3.7 by b (Niezink et al. 2019).

$$\frac{dx(t)}{dt} = ax(t) + b \tag{3.7}$$

The solution to an ordinary differential equation of this type is a function (or set of

functions) of time, $x(t)$, that satisfy the relationship specified by the coefficients a and b.

Stochastic differential equations also include a stochastic error term which adds random variation around a function like that in equation 3.7. The error term is defined by a Weiner process at the level of $X(t)$, which differs from the distribution-based error terms found in multilevel models (Bjork 2004; Evans 2013). The Weiner process is a continuous-time stochastic process in which the difference between two realisations of the process at two different times, $W(t_2) - W(t_1)$, is an independent random draw from a normal distribution with mean zero and variance a function of the time difference between the realisations (Evans 2013). In Equation 3.8 this is represented by $W(t)$. The Weiner process cannot be differentiated with respect to time, hence the altered representation Equation 3.8 compared to the previous Equation 3.7. The first line of Equation 3.8 shows a simple, informal representation of the second line.

$$\begin{aligned} dX(t) &= [aX(t) + b]dt + gdW(t) \\ X(t) &= X(0) + \int_0^t [aX(u) + b]du + \int_0^t gdW(u) \end{aligned} \tag{3.8}$$

The result of including the stochastic error term means that $X(t)$ is now a random variable and there are many potential realisations of the model in Equation 3.8 (Bjork 2004; Niezink et al. 2019). The expected value of $X(t)$ at a given time is equal to the solution of the underlying ordinary differential equation (Niezink et al. 2019). One advantage of stochastic differential equations is the relatively interpretable (compared to spline models) 'real-world' meaning of their coefficients. The interpretation of Equation 3.8 would be that for a small interval δt, $X(t)$ changes by an amount equal to an independent random draw from the distribution $N([aX(t) + b]\delta t, g^2\delta t)$ (Niezink et al. 2019). Like multilevel models, stochastic differential equations can be modified to include autocorrelation effects for temporal property patterns associated with multiple observational units (Niezink et al. 2019). They can also include random effects (Ditlevsen and De Gaetano 2005).

Stochastic differential equations do not directly model the relationship between time and $X(t)$. It has been suggested that this reflects a belief that time does not directly cause change but can be used as in index to represent change over time that is caused by other

processes. However, the solution to a stochastic differential equation does still involve a function of time, $X(t)$, as an expected value, which is in many ways analogous to the temporal functions underlying multilevel models and functional data analysis which model a 'true' function with additional error.

Notably, the relationship between the value of the function $X(t)$ and its differential ($\frac{dX(T)}{dt}$) remains constant over time unless the stochastic differential equation explicitly includes time-varying covariates. In many cases this means that if the value of $X(t)$ is the same at two different times, the slope of $X(t)$ will also be the same at these two times. This limits the flexibility of the temporal patterns captured by stochastic differential equations.

3.8 Summary

This chapter has outlined a range of methods that aim to analyse temporal patterns for multiple observational units with continuous time specifications. Ultimately, this project aims to apply continuous-time methods to analyse temporal patterns of observational units within a network structure. Multilevel models appeared to be suitable candidates for modelling continuous temporal patterns, particularly when extending this to a network context due to their ability to account for complex variance structures. Estimation of multilevel models, however, can be time consuming and the results can be difficult to interpret and relate to a real-world context. Care must also be taken when using multilevel models for repeated measures data. Methods that condition on the outcome, including the use of multilevel models in a way that conditions on the outcome, have a fundamental issue in their failure to properly address their purported aim. This problem has not been extensively discussed or explored in the literature and should be addressed (Tu and Gilthorpe 2010; Tu, Tilling, et al. 2013) – hence a deeper discussion and analysis of issues with these methods is included in Chapter 6. They would not be capable of capturing continuous temporal patterns of network properties.

While simple latent growth curve models are able to treat time as a continuous variable (Sterba 2014), the most flexible form of this model, the latent-basis latent growth curve model, does not strictly treat time as a continuous variable; non-parametric 'stretching' of the time-axis to produce deviations from linear patterns results in the model only being defined at discrete, observed time points.

Like multilevel models, functional data analysis is capable of capturing complex curves using spline functions. It is possible to model residual structures to account for temporal dependence, and further analyses of the functions can investigate spatial patterns (R. Jones and Boadi-Boateng 1991). However, the software for easily fitting complex residual and covariance structures is much more established in the field of multilevel modelling.

Stochastic differential equations provide an interesting way of modelling changes over time while aiming to shift the focus of an analysis away from functions of time and towards the generation process underlying the data. While stochastic differential equations and functional data analysis both present potentially useful methods for modelling continuous temporal patterns, multilevel models were chosen as the focus for further investigation in this thesis due to their ability to include spatial covariation in random coefficients and the wide availability of software to fit complex covariance structures (Besag, York, et al. 1991). The details of multilevel model application in this context and approaches to producing interpretable summaries are discussed in Chapter 4.

Chapter 4

Methods

4.1 Introduction

Chapter 3 outlined a range of methods for analysing temporal patterns of variables within individuals and other units of observation, for example growth curves within children. Of these, the most suitable method for application to spatio-temporal networks was multilevel models. This chapter outlines a multilevel model that can be applied to model continuous temporal patterns of edge properties in a spatio-temporal network. This is discussed in the context of an example application: modelling 'flow' of passengers between stations in a rail network. The rail stations are represented by vertices in a network and the direct connections (rail lines) between stations represented by edges. 'Flow' is an abstract edge property, used purely for illustrative purposes, that varies over time giving some indication of the rate at which passengers are travelling along an edge (between stations).

4.2 Multilevel model for edge properties in a directed spatio-temporal network

4.2.1 Requirements for the model

The aim of a multilevel model for patterns of flow along edges of a rail network is to capture temporal patterns of flow for each edge. Temporal patterns for different edges will be correlated with each other due to spatial and network relations between edges. Successive measurements of flow for each edge will be correlated due to temporal proximity. Models

for the temporal patterns of flow should account for the spatial, temporal and network structure of the data.

4.2.2 Model specification

The specification of the proposed model for temporal patterns of flow in a rail network is described in Equation 4.1 in which $from$ and to index the origin and destination stations, respectively, for edges in the network; t refers to the time at which each measurement was made; $b_n(t)$ refers to the value of basis function n at time t; k refers to the number of basis functions specified; $e_{i,from,to}$ represents the residual term for the i^{th} observation for a particular edge; u represents aspatial random effects, indexed by origin or destination station (and basis function for u_n); v represents spatial random effects following an intrinsic Gaussian conditional autoregressive (CAR) distribution, indexed by origin and destination station (and basis function for v_n); α represents the coefficient for the autoregressive residual structure estimated by the model and w_i the independent residual term in this structure.

$$\begin{aligned} y_i &= \beta_{0,from,to} + \sum_{n=1}^{k} \beta_{n,from,to} b_n(t_i) + e_{i,from,to} \\ \beta_{0,from,to} &= \beta_0 + u_{0,from} + u_{0,to} + v_{0,from} + v_{0,to} \\ \beta_{n,from,to} &= \beta_n + u_{n,from} + u_{n,to} + v_{n,from} + v_{n,to} \\ e_{i,from,to} &= \frac{\alpha e_{i-1,from,to}}{t_i - t_{i-1} + 1} + w_{i,from,to} \end{aligned} \tag{4.1}$$

4.2.3 Spatial structure

Each edge is indexed by its origin and destination stations. A random intercept coefficient is included for each station, as an origin and destination, along with a random coefficient for each basis function in a spline function – see Section 3.2.4 (Goldstein 1994). This allows edges with the same origin to be more similar to each other than edges with different origin stations. The same applies to destination stations. This accounts to some extent for similarities between edges with similar origins and destinations in the network.

Each random coefficient is formed of a spatial and aspatial component. The spatial com-

ponent allows edges with origin and destination stations geographically close to each other to have similar coefficients. The spatial random effects are set to follow an intrinsic CAR distribution (Besag, York, et al. 1991). This uses a spatial weight matrix to specify the relationships between stations. The CAR distribution allows the random effects coefficients to covary based on these relationships – where two locations are near each other, the weight matrix should specify that the coefficients are more similar. A spatial weight matrix is a matrix with numbers of rows and columns equal to the number of locations of interest. Each row summarises the relationships of one location with all others (Dubin 2009). This might be represented as a binary presence or absence of a relationship, or by continuous weights indicating the strength of each relationship. A spatial weight matrix can be specified in different ways, with two major approaches being 'neighbour' or 'contiguous' relations, and distance-based relationships (Dubin 2009).

Neighbour or contiguous relationships are based on the sharing of boundaries between locations. They are easy to construct when the locations are areas, for example counties within a country, or if they are point locations forming a lattice. Two areas are deemed neighbours if they share a boundary (Dubin 2009). Contiguity relations between points in a lattice can be specified in several ways. For example, adjacent points within rows or columns might be considered neighbours, or diagonally adjacent points, or both (Dubin 2009). Often contiguity relations specify only the binary presence or absence of a relationship between locations.

If point locations are of interest and they are not spread in a lattice, a weighting scheme based on distance might be used (Dubin 2009). Normally, one expects the strength of correlations between locations to decrease with distance (Tobler 1970). Therefore, we must specify the weights as some function that decreases with distance. Two common functions are inverse distance ($\frac{1}{distance}$) and negative exponential distance ($e^{(-a \times distance)}$ where a is some constant) (Dubin 2009).

It is possible to specify relationships in this way for all pairs of locations, but sometimes we wish to include a cut-off point outside of which we consider two points to have zero relationship (Browne and Draper 2006) This may be because we have information about the underlying processes that suggests there will be no relationship beyond a certain distance. Two common ways to do this are by including a maximum set number of the

nearest neighbours for each location (k-nearest neighbours) or by including a distance cut off – if the distance between two points exceeds the cut off, the spatial weight is set to zero (Browne and Draper 2006). Unlike using a distance cut off, the k-nearest neighbours approach does not guarantee a symmetric weight matrix – for example, one point may be included as a nearest neighbour of another point that is not included in its own list of nearest neighbours. This makes the nearest neighbour approach unsuitable for the CAR distribution, as the weight matrix specifying relationships for this must be symmetric (A. Thomas et al. 2014).

The random effects associated with the intrinsic CAR distribution are restricted such that they must sum to zero (A. Thomas et al. 2014). If any station in the network is considered not spatially related to any others, no random coefficient is produced from the CAR distribution – this is why an additional aspatial component is included in this model, to account for any geographically isolated stations (A. Thomas et al. 2014).

4.2.4 Temporal autocorrelation

Temporal autocorrelation between successive measurements within each edge is specified using autocorrelated residual structures. The residuals in the models represent the departure of each observation from the estimated model function. In this model, a first order temporal autocorrelation structure is specified. This means that the model residual for a given observation is a function of the previous model residual value, plus some random variation (Steele 2008). In this case, the autoregressive function specifies that the effect of the previous residual value on the current one is inversely proportional to the time difference between them plus one (adding one allows simultaneous measurements).

4.2.5 Spline structure

Spline structures were previously discussed in Sections 3.2.4 and 3.6.1. Multilevel models use parametric representations of the relationships between variables, which can make capturing complex temporal patterns difficult. One way to capture complicated shapes parametrically is to use spline functions which split the time-axis into sections, each of which is modelled with a parametric spline function (De Boor 2001). When splines are included in a model, a basis function must be chosen. This includes the choice of the type of spline, the number of sections to divide the time-axis into and where the divisions,

which are called knots, occur. After this, a coefficient for each spline in the basis function is estimated as part of the model.

There are a range of different types of spline functions: for example, b-splines, natural cubic splines, smoothing splines. B-splines were described in section 3.6.1 in relation to functional data analysis. Natural cubic splines are a series of cubic polynomials joined at knots points. Both the splines and their first and second derivatives are constrained to be continuous over these knot points (De Boor 2001). Natural cubic splines also satisfy another condition – at the knots at the tail ends of the spline system, the second derivative of the cubic splines must be equal to zero (Perperoglou et al. 2019). This reduces erratic behaviour of splines at these boundaries which is common in b-splines (Perperoglou et al. 2019).

The choice of knot positions for splines is important as it partially determines the shape and fit of a spline function. Including more knots can lead to a more closely fitting spline function that is capable of capturing more complex or 'wiggly' patterns, but will make the model harder to estimate as more parameters are included (Ramsay and Silverman 1997). This is especially the case where random effects are included for the spline coefficients. There are automatic knot selection procedures available for single-level spline models, but so far no procedures for multilevel models have been identified (Yeh et al. 2020; Yuan et al. 2013; S. Zhou and Shen 2001).

Some spline functions additionally include a smoothing parameter to control their behaviour. For natural cubic and b- splines, the 'smoothness' of the function produced (whether it fits the data very closely and is very 'wiggly' or whether it is very smooth) depends on the number and placement of knots (Ramsay and Silverman 1997). Splines with a smoothing parameter reduce the importance of this choice by including an additional parameter to control smoothness (Perperoglou et al. 2019). Penalised splines (p-splines) are an example of this type of construction.

While smoothing splines such as p-splines reduce the importance of knot placement, for b-splines and natural cubic splines, the value of each spline at each observed time point can be calculated before model fitting, and then each spline can be added to the model as one would include any other variable – for these splines estimation of the model coefficients can then proceed as normal which is convenient. For penalised splines, the smoothing

parameter is integral to the estimation process, which can introduce extra complications when estimating models with penalised splines (Perperoglou et al. 2019).

In the following chapters, b-splines were chosen as the basis of the model specified in Section 4.2.2. As discussed later in Section 4.3, performing calculus on the resulting individual level models forms an important aspect of producing interpretable results from models with coefficients that do not necessarily hold a real-world meaning. While b-splines can behave in a more erratic way than natural cubic splines at the boundaries of the spline function, the R package 'splines2' provides a method for producing derivatives and integrals of different orders with b-splines, which made the process outlined in Section 4.3 more convenient to complete (W. Wang and Yan 2020).

4.2.6 Software and estimation procedure

Due to the complexity of the model, Bayesian estimation procedures were selected over maximum likelihood estimation which often struggles to converge on accurate estimates, if at all, in the presence of complex variance structures. The abilities of several software packages for modelling complex structures were investigated, including OpenBUGS, JAGS, STAN, MLWin and MPlus (Bürkner 2017; Carpenter et al. 2017; Lunn et al. 2009; L. Muthén and B. Muthén 2010; Plummer 2003; Rabash et al. 2020; Stan Development Team 2019). Of these, only OpenBUGS was capable of fitting temporally autocorrelated residuals *and* spatially correlated random effects (A. Thomas et al. 2014). Therefore, this package was chosen for model fitting.

OpenBUGS uses the BUGS engine to estimate the models – this uses a Gibbs sampling routine to sample parameter estimates. Gibbs sampling is a Markov chain Monte Carlo estimation procedure based on the conditional distributions of each parameter to be estimated. The conditional distribution of a parameter is the probability distribution of that parameter given (i.e. conditional on) the values of all other parameters (S. Geman and D. Geman 1984; Zyphur and Oswald 2015). Gibbs sampling produces a series of sets of parameter estimates. For each of these, the parameter values are estimated conditional on the values of all other parameters at the previous step. There are alternative Markov chain Monte Carlo estimation processes for fitting Bayesian models, some of which are faster than Gibbs sampling, meaning that they take fewer samples to converge on the posterior

distribution of each parameter (Hoffman and Gelman 2014). However, as OpenBUGS was the only software with the facility to fit spatially autocorrelated random effects, these could not be used for the model in Section 4.2.2.

The result of the BUGS estimation is a number of posterior parameter samples representing the posterior parameter distributions which can be summarised using summary statistics. If a point estimate of parameters is required, the mean value of the distributions can be taken (Zyphur and Oswald 2015). 95% credible intervals can be calculated using the 2.5^{th} and 97.5^{th} quantiles of the distributions. These give an indication of the spread of the posterior distribution for a parameter, and a range of values it is most likely to take (Zyphur and Oswald 2015).

4.3 Interpretable summaries of multilevel models for spatio-temporal networks

The model in section 4.2.2 can be used to capture temporal patterns of flow for individual edges in a spatio-temporal network. These patterns of edge properties are represented as a spline function. Interpreting the coefficients of spline models does not always provide a real-world or easily understood meaning. In addition, visualising the temporal patterns of flow for all edges in a potentially large network typically would produce cluttered and unreadable graphs, especially if trying to incorporate spatial dimensions.

4.3.1 Temporal pattern features

Colak et al. (2013) presented, on a map of a transport network, the timing of a specific event, defined as a specific feature extracted from a stochastic model of road congestion. This approach to choosing a specific aspect of interest of a temporal pattern lead to the production of a relatively uncluttered graph with a clearly interpretable real-world meaning. The uncluttered nature of this graph does rely on the structure (density, spatial configuration) of the network, but overall a graph containing one numeric piece of information per edge is less cluttered than one attempting to display a temporal pattern for each edge.

To simplify their interpretation, various pattern features can be extracted from the func-

tions specifying continuous temporal patterns of flow along network edges. These could then be mapped or further analysed. For example, similar to Colak et al. (2013), the timing of when flow reaches a certain value can be extracted. Alternatively, the timing of particular shapes in the function, regardless of the flow value at that time could be derived. For example, calculus could be used to extract maxima and minima or the slope of the function at any time point. These would be calculated by differentiating the flow function fitted for each edge. The derivative produced represents the slope of the flow function for a given edge over time. Where the slope is equal to zero, the flow function is either undergoing a change in direction – this represents either a minimum or maximum – or a 'saddle point' where the slope of the function momentarily reaches zero but the value of the function is not extreme (John 2009; Thompson 1919). The minimum and maximum would represent the times of day at which the fewest or most passengers are flowing through a particular edge on the network, respectively. The slope itself gives an indication of how quickly the number of passengers is changing. This information could be used to consistently define the start and end of a 'rush period' where the number of passengers begins increasing rapidly at the start of the rush period, or stops decreasing rapidly at the end of it. This could give a good indication of how long edges in different parts of the network are exposed to high travel demand.

4.3.2 Continuous whole-network properties

While extracting pattern features for each edge reduces the requirement to represent temporal patterns in a data visualisation, summarising whole-network properties would reduce the requirement to represent spatial variation. Many complex network properties are estimated using edge weight information (Newman 2003). Flow (or any edge property) might be thought of as an edge weight. Given a continuous temporal function of flow for each edge, we are able to calculate continuous temporal functions for a range of whole-network properties. For example, weighted edge density, or clustering coefficients (Horvath 2011; Saberi et al. 2017). These continuous functions could be used to clearly represent how network properties change over time, and to extract whole-network properties at any time point. If measurements of edge properties like flow are not made at regular intervals, and the same time for all edges, this would not be possible without modelling patterns of flow for each edge unless the data were aggregated.

4.3.3 Estimation of uncertainty

It is relatively easy to calculate point estimates for pattern features and a single function for continuous-time whole-network properties using mean parameter estimates from posterior parameter distributions. It is important to present these estimates with a representation of uncertainty, as we would with 95% confidence intervals for a frequentist analysis or with posterior parameter distribution summary statistics for Bayesian estimation.

For each sampled set of posterior parameters produced during the estimation process, there will be corresponding sample pattern features and a corresponding continuous whole-network property function. If all sample pattern features are extracted from the models using each set of parameter estimates, we produce a posterior pattern feature distribution. This can be summarised in much the same way as the parameters themselves, using quantiles of the posterior distribution to represent a 95% credible interval. If a continuous whole-network property function is calculated for each set of parameter estimates, this set of functions can be used to summarise the posterior distribution of the whole-network property at particular time points. This could be represented, for example, as a mean whole-network property function of time with a set of simultaneous 95% credible intervals. Note that if simultaneous credible intervals are calculated, the quantiles of the posterior distribution used to generate the intervals must be adjusted for multiple testing (Francisco-Fernández and Quintela-del-Río 2016).

4.4 Simulations to assess model performance

Ideally, we want models to estimate parameters, pattern features and/or continuous temporal whole-network properties with accuracy and precision. Accurate estimates are unbiased – that is, they do not systematically under- or over-estimate parameter values (Porta 2014). Precise estimates show a small amount of random variation around the parameter values (Porta 2014). For example, the posterior parameter distribution might be very narrow if an estimate is precise, but wide if it is not.

If we wish to assess accuracy, it is important to know the true parameter value or parameter distribution we are estimating, in order to compare estimates to it. One way to do this is to use simulated data which is generated with known underlying values. If a series

of simulated datasets with the same underlying values are generated and analysed, we can calculate the degree of bias for each dataset and use this to estimate the mean bias of our analysis method (Morris et al. 2019). We can also investigate the precision of our estimates in the context of known information about the dataset (Rasmussen et al. 1998). The use of simulations is a common method to assess performance of statistical models. However, it only assesses this performance in a specific set of circumstances. While this is important in investigating the performance of measures when assumptions required by certain models are broken, it does not produce general results regarding model performance and so performance of models in the face of many different circumstances should be evaluated (Morris et al. 2019).

Simulation studies also tend to view model assessment from a frequentist viewpoint – identifying the difference between estimated parameter values and the underlying 'true' parameter values used to simulate the data (Morris et al. 2019). This is not an ideal method for analysing the performance of truly Bayesian methods. However, while Bayesian estimation tools are being used in this situation, the analysis conducted is not following a true Bayesian philosophy, as such. The prior distributions specified (when the model is applied in 7) are loose and do not provide much information about prior knowledge or beliefs. It is common practice for this approach to be used when applying complex multilevel models because Bayesian estimation tools perform far better than traditional frequentist methods, for example likelihood-based methods, when estimating them (Browne and Draper 2006).

When evaluating methods through simulation, the simulated parameter value is compared to the estimated parameter value for each simulation. The mean and spread of the distribution of this measure across all simulations gives an indication of the amount of bias in parameter estimates (Bland and Altman 1999). The appropriateness of 95% confidence interval widths can be estimated by identifying what percentage of them cover the simulated value – one would expect this to be 95% (Rasmussen et al. 1998).

In the simulations in Chapter 8, the posterior distribution of each parameter is of interest as well as the point estimates of parameters calculated from them. We must compare the posterior distribution to point values for simulated parameters in each simulation. To do this, the location of the simulated parameter value in the posterior distribution was calculated as a z-score. Across all simulations we would expect these measures to

form a distribution with mean zero and standard deviation one. This process aims to assess whether the simulated parameter value is likely to be from the same distribution as the posterior parameter distribution and gives us an indication of whether the estimated posterior distributions for parameters are accurate in their location and appropriately precise in terms of their spread.

This approach is also useful for assessing individual-level parameters – in the process used to simulate data these parameters vary randomly (in a process analogous to a random effects model) and only population parameters relating to these random effects are specified as an input. This means the simulated parameter values for a given observational unit will vary between simulations considerably, and an estimate of accuracy and precision that is not affected by this is required. Details of the generation of simulated data for the purposes of analysing the models discussed in Section 4.2.2 is discussed in Chapter 5.

4.5 Summary

This chapter outlines a continuous-time, cross-classified multilevel model that can be used to model temporal patterns of edge properties in a sptio-temporal network. Also described in Section 4.3 is a process for simplifying the results from this model, by extracting temporal pattern features or by generating temporal patterns of whole-network properties. In Chapters 7 and 8 these methods will be applied to spatio-temporal network data and tested using simulations.

Chapter 5

Data

5.1 Introduction

This chapter introduces data used throughout this project to test and assess the performance of methods outlined in Chapter 4. First, the development of a longitudinal simulation tool to generate simulated data representing patterns of edge properties associated with a spatio-temporal network is outlined. Then, details of real data associated with the London Underground rail network from Transport for London Open Data are introduced.

5.2 Simulated data – development of a longitudinal simulation tool

To use Monte Carlo simulations to test the accuracy and precision of a method, we must simulate data with known properties that we aim to recover by modelling it (Morris et al. 2019). It is also important to be able to vary these properties and the data structure to test modelling methods in different circumstances – this will help identify situations in which the models do not perform well. A simulation function was developed to produce longitudinal data for multiple observational units with the option to include spatial locations or origin-destination data, along with appropriate covariance structures that reflect the data generating process assumed.

5.2.1 Requirements for a simulation function

Measurement schemes

Temporal data are available from a number of different sources: they may be collected specifically for an experimental or observational study (Albert 1999; Golding 2004), or secondary data such as electronic health records may be used (Herrett et al. 2015). It is normally not possible to continuously measure data, and so often repeated measurements of a longitudinal variable are taken. There are several different attributes which can describe the structure of these measurements, and these can govern which analysis methods are suitable for a given dataset. These include:

- Length of follow up – the length of time over which repeated measures of the variable are taken.
- Frequency of measurements – the amount of time between each measurement taken.
- Number of measurements – whether different individuals have the same number of measurements in the study or not (Fitzmaurice and Ravichandran 2008).
- Measurement simultaneity – whether all individuals are measured at the same time or not (Fitzmaurice and Ravichandran 2008).
- Between observational unit measurement interval heterogeneity – whether the intervals between measurements differ *between* individuals (Pullenayegum and Lim 2016).
- Within observational unit measurement interval heterogeneity – whether the intervals between measurements vary *within* individuals (Pullenayegum and Lim 2016).

Figure 5.1 shows examples of how the last three of these attributes – measurement simultaneity, between observational unit interval heterogeneity, and within observational unit interval heterogeneity – can interact to produce different measurement schemes. Some potential combinations are not possible: if there is between individual measurement interval heterogeneity, for instance, measurements cannot also be simultaneous. Variation in these three attributes is usually caused by the adoption of a particular measurement scheme. Measurement schemes may be categorised into three types: regular, wave and irregular.

Regular measurement schemes usually follow a strict measurement schedule with no vari-

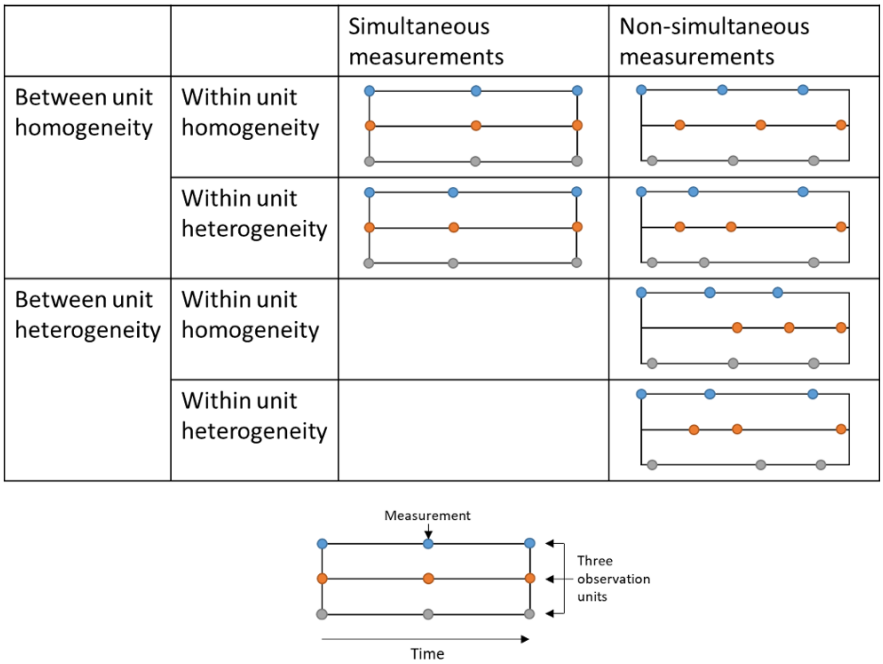


Figure 5.1: Visualisation of different properties describing measurement schemes

ation in measurement times between individuals (Hidas and Wagner 2004). These often result in between observational unit homogeneity in measurements and sometimes within observational unit homogeneity, especially if measurements are taken automatically by a sensor or similar device.

Wave measurement schemes often have a schedule, but there may be variation in the measurement times around this, such that the data collection takes place in a series of waves (Van Wissen and Meurs 1989). This may occur, for example, in a temporal healthcare study where participants are called back for examination every three months, but the dates of examinations may vary slightly due to their availability. This often results in between observational unit measurement interval heterogeneity.

Irregular measurement schemes do not follow a schedule, so measurements occur with no specific pattern throughout the time period (Pullenayegum and Lim 2016). This may occur if measurements are taken when an event occurs, for example, information about journeys taken on public transport collected when each journey takes place. Datasets with these measurement schemes may also be created by combining several datasets with differing measurement schemes. This often results in between and within observational unit measurement interval heterogeneity. The timing of measurements may be almost completely random, for example measurements of seismic events, but some event-driven

measurements may be distributed unevenly throughout time, for example if measurements are taken when passengers check into a rail system there will probably be more measurements during rush hour commuting periods. In some cases, the timing of the measurements may be related to their value.

It is important to be able to test the effect of different measurement schemes on performance of statistical models. Therefore, it is necessary for a simulation tool to be able to produce varied measurement schemes, including regular, irregular and wave data, and with control over the six attributes listed above.

Pattern flexibility

The ability of a model to recover a temporal pattern might be affected by the complexity of the pattern itself. Temporal patterns can vary greatly in complexity, from simple straight lines to complex curves (Cole et al. 2010) and patterns that might involve step changes (Matthews et al. 2016) or discontinuities (Howe et al. 2013). It is important for a simulation tool to be able to produce a large range of different patterns with varying degrees of flexibility in order to test models under varied circumstances. As discussed in Section 4.4, simulations only test the performance of methods in specific circumstances. Testing a large number of circumstances with abstract patterns provides a wide view of a model's capability, but it may be more important to be able to link simulated circumstances to real situations where the model might be used. Having a large degree of control over the temporal patterns is therefore key to making sure simulations reflect real situations as much as possible.

It is also important to have a degree of control over the variation in patterns between observational units in the data. For some processes, we would expect temporal patterns to follow a similar shape for all observational units. For example, in childhood growth, we expect the majority of children to grow relatively slowly before teenage years, go through a 'growth spurt' of fast growth, and then slow the speed of growth until they remain the same size (Cole et al. 2010). However, there will likely be some variation in the exact shape of these patterns between observational units – these can be thought of as transformations of a growth function as shown in Figure 5.2. A larger child will have a greater mean size, which can be represented by a y-axis translation of the growth curve. Some children may

have later growth spurts than others, resulting in the growth curve being translated along the x-axis. Some children may grow less in the same growth spurt period as others, which can be represented by a scale or stretch transformation in the y-axis. A longer growth spurt period might be described as a scale or stretch transformation in the x-axis. It would be useful for a simulation function to be able to produce variation in some of these pattern properties between individual observational units.

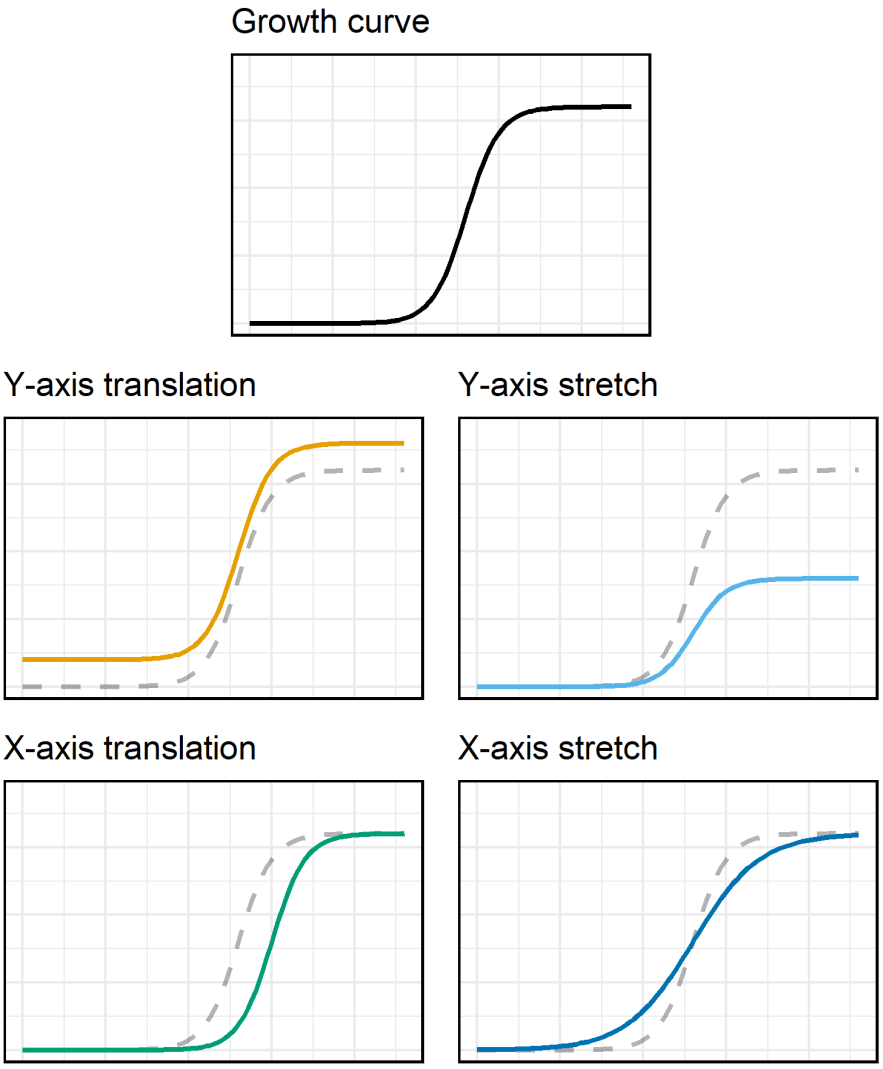


Figure 5.2: Different possible transformations of a growth function

Correlation structures

When repeated measurements are made to record temporal patterns for a number of different observational units, it is expected that measurements within observational units will not be completely independent (Fitzmaurice and Ravichandran 2008). That is, they will be more similar to each other, than to measurements for a different observational unit.

This phenomenon is called autocorrelation (or in some domains clustering or interference).

In addition to this, measurements closer to each other in time are also likely to be more similar than measurements farther apart in time. This is referred to as temporal autocorrelation (Fitzmaurice and Ravichandran 2008).

Correlation between individual observational units may also occur as a result of placement in space. According to Tobler's first law of Geography, "near things are more related than distant things" (Tobler 1970). The distance between two objects can be defined in a number of ways, however. For example, distance may be measured as Euclidian distance, travel time using a certain mode or transport, or as the number of edges traversed between two network nodes (Apparicio et al. 2008; Newman 2018). Spatial clusters or communities may also exist, for example, of people living in a certain district, or people with travel patterns between the same two areas (Gesler 1986; Zhu and D. Guo 2014).

It would be useful for a simulation tool to produce data with some of the types of temporal, spatial or network autocorrelation, though to capture the full range of possibilities would be extremely challenging. It would be useful for the strength of these correlations to be variable, to test how models cope with different data structures.

Individual level knowns

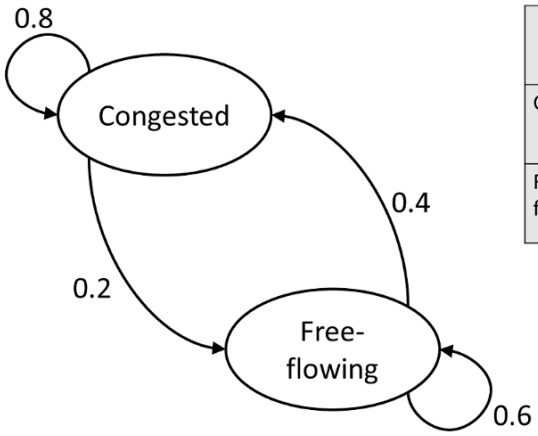
While many studies aim to capture population level information from a model, some studies may seek to examine the properties of subgroups or individual observational units. In Chapter 4, the model specified can be used to examine information at any of these three levels. It would be useful for a simulation tool to have known individual level parameters describing the observational unit level temporal patterns and subgroup level parameters (if subgroups are included), as well as known population level parameters, in order to examine the recovery of information at all levels of the data. However, it is important to note that attempting to predict or interpolate information at an individual level is not likely to yield precise results – statistical methods predict expected values for groups of individuals (either at the whole population level or in subgroups) and results drawn at the level of an individual unit of observation from a model such as that in Section 4.2.2 are based on this individual belonging to two groups (origin and destination stations), each of which have an average expected value (Wilkinson et al. 2020). This means the results

are likely to lack precision at this individual level, however, it is still of importance for a simulation tool to be able to assess this effect.

5.2.2 Potential approaches

Hidden Markov chains

Markov chains can be used to construct time series data. At each time point, an observational unit is assigned a state out of a finite set of states. The state of the observational unit at the next time point depends on the probabilities of transitioning from the current state to any other available states (P. Jones and P. Smith 2018). These transition properties depend only on the current state, not any previous states, which gives rise to the Markov property, or 'memorylessness' (P. Jones and P. Smith 2018). The structure of a Markov chain system can be visualised using a phase diagram, such as Figure 5.3. This represents the state of traffic on a road which can be congested or free-flowing. The diagram shows the states in ovals with transition probabilities attached to the arrows between them. When in a free-flowing state, there is a 0.6 probability of the state at the next time step also being free-flowing, and a smaller probability of the road becoming congested by this time. If congested, the road is much more likely (probability equal to 0.8) to be congested at the next time step. There is only a 0.2 probability of the road becoming free-flowing by the next observation time. Figure 4.3 also shows the transition matrix for the system shown – this matrix defines the Markov chain by detailing transition probabilities between all possible pairs of states (P. Jones and P. Smith 2018).



	Congested	Free-flowing
Congested	0.8	0.2
Free-flowing	0.4	0.6

Figure 5.3: Transition diagram and matrix for a Markov chain representing the state of traffic on a road

Markov chains normally advance forwards in equal, discrete time steps, but continuous-time Markov processes can be described. These normally include two random processes for each state transition – a decay function (often exponential) describing the probability distribution for when the next state transition will occur from the current state, and transition probabilities for what the next state will be (Liggett 2010); both depend on the current state.

Transition probabilities in Markov chains can be allowed to depend on time, as well as the current state. Where this is the case, the Markov chain is said to be non- or in-homogenous (P. Jones and P. Smith 2018). The transition matrix may vary discretely over time or each transition probability may be defined as a continuous function of time.

The states in Markov chains could be thought of as various values of a categorical variable; however, these processes can also be used to simulate continuous variables in time series using hidden Markov chains (McLachlan and Peel 2000). In a hidden Markov chain, each observational unit moves between different states as described above. For each measurement in time, however, the state an observational unit is in determines the distribution from which a continuous variable is sampled (McLachlan and Peel 2000). This produces a series of samples from continuous distributions at different points in time, with the added complexity of moving between different states/sample distributions at different time points. The sampled continuous variables could be treated as absolute values, but also as differences between the current value and next measurement.

An initial approach to simulation was developed based on a non-homogenous hidden Markov chain process. This was intended to represent growth of individual observational units (for example, growth in the weight of children) through three 'growth phases': initial slow growth, growth spurt, steady state. State transition probabilities depended on the current state and time, following a function of time with limits zero and one. At each observation time, the amount of growth since the last observation was sampled from a distribution that depended on the current state. The phase diagram for this process is shown in Figure 5.4. The code for this is contained in Appendix B.

Figure 5.5 shows data produced by this hidden Markov chain process for 15 observational units, with the phase of each observation indicated. This initial approach was able to produce complex patterns through the inclusion of three phases. It could be extended

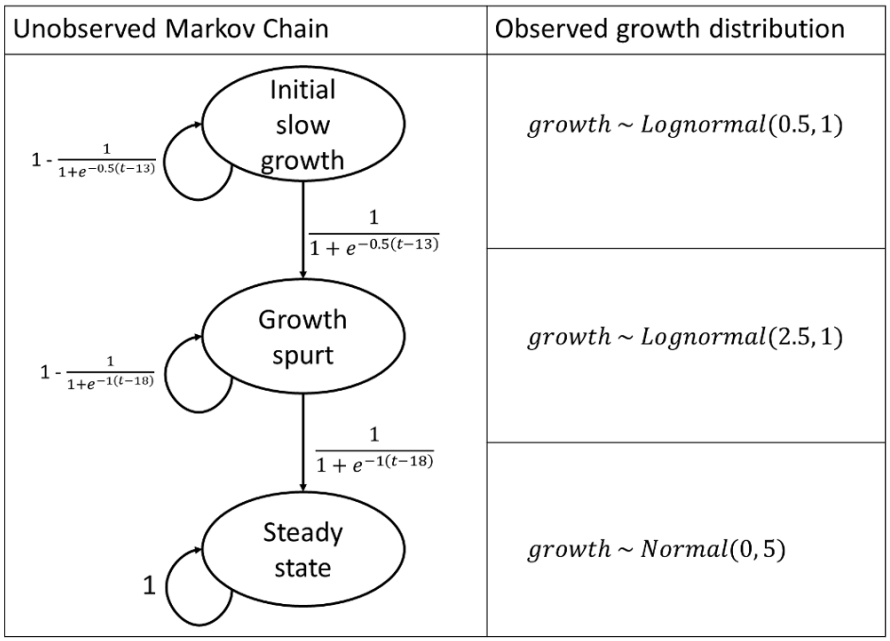


Figure 5.4: A hidden Markov chain process to simulate a three-phase growth curve

to produce measurement time variation and potentially extended to create clusters of similar chains (for example, with differing transition probabilities). The process produces autocorrelation within each observational unit, as change (growth) values are added to the previous value, meaning observed size is dependent on previous value. However, the degree of autocorrelation is not specified explicitly and depends on the distributions from which observations of growth are sampled. Additionally, while we know the average value of change for each growth phase (an average growth rate parameter) we do not have other known parameters, such as mean size. The known underlying parameters are also not at the individual observation level but known at the population level.

An attempt to develop a stochastic simulation process with a known individual level mean value, but still based on change to ensure autocorrelation, was developed using code in Appendix B.

In this data simulation process, each observational unit has a known average value selected from a normal distribution. Each is then assigned an initial value. Each successive change value is sampled from a different distribution depending on whether the previous value was above, below or equal to the desired average for each observational unit. If the previous value was above the average, the change value was sampled from a negative log normal distribution. If the previous value was below the average, the change value was sampled

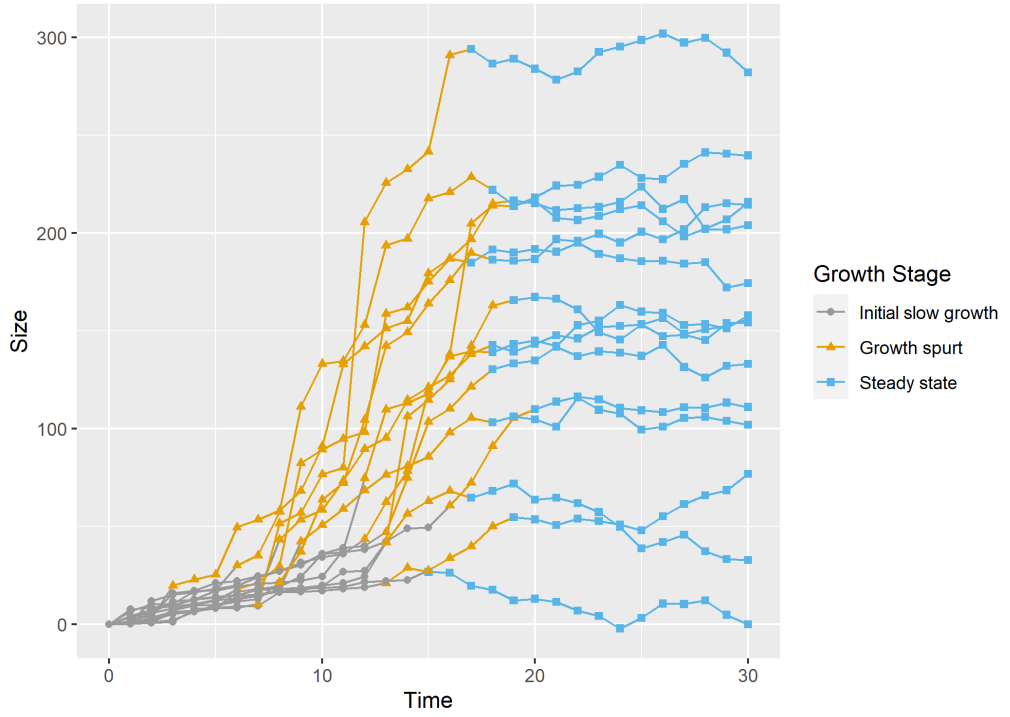


Figure 5.5: Three phase growth curves generated by the hidden Markov chain in Figure 5.4

from a positive log normal distribution. If the previous value was equal to the average, the change value was sampled from a normal distribution. This produced data such as that shown in Figure 5.6.

The simulation produced a series of random observations which generally had an average within one unit of the specified average for each observational unit. The performance of this process could potentially be enhanced by including information from more than one previous measurement to give an indication of the 'current' average value achieved. This would represent a departure from the hidden Markov chain process, which specifies that the current value only depends on the one prior. The simulation process used to generate these data means that only the mean and degree of random variation could be specified – the change for each observation (and therefore other parameters such as average slope) depends on this parameter and therefore cannot be specified separately.

A process such as this could be used to generate data within each phase of a hidden Markov process similar to that described earlier. One advantage of this approach is that is produces data based only on previous values in time and does not account for future values. The stochastic simulation approach reflects the nature of real-life data generation,

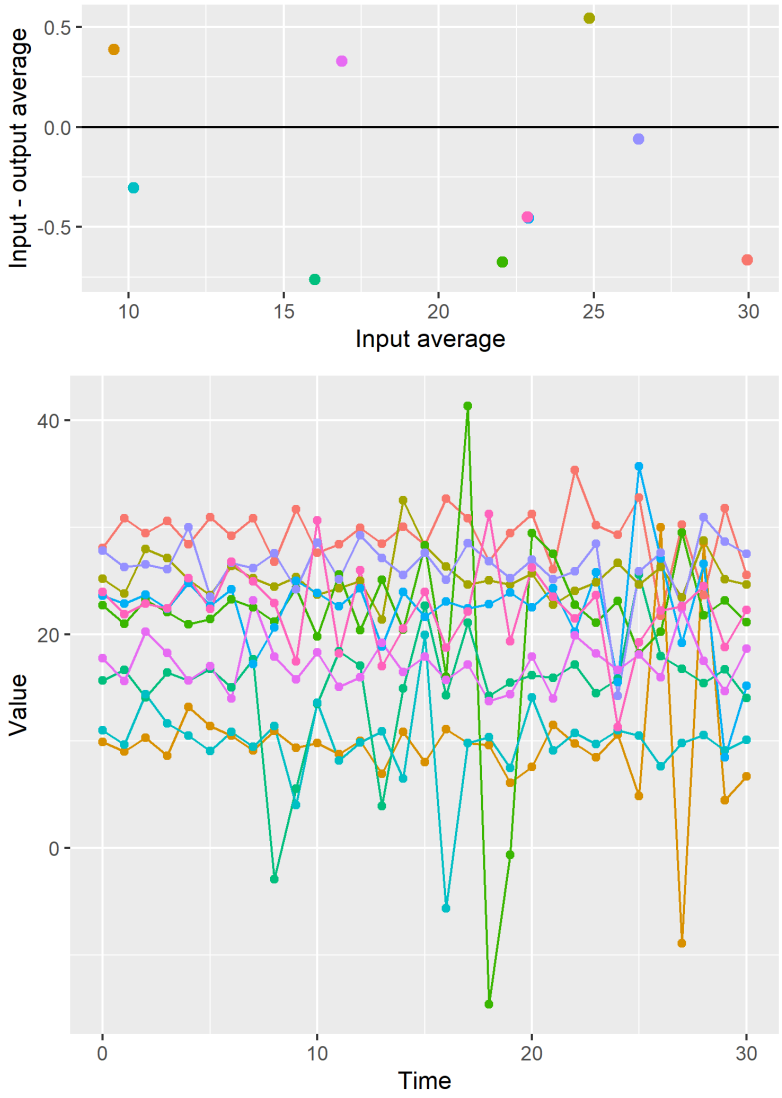


Figure 5.6: Data generated from a potential simulation process using a known average value for each observational unit

which is likely to truly be based on stochastic processes (Oud et al. 2012). However, there are some attributes listed in Section 5.2.1 that would be difficult to specify when simulating data using a hidden Markov chain approach. The strength of autocorrelation would be difficult to directly specify when constructing a hidden Markov chain simulation tool. It would also be hard to specify a range of other parameters not used in the construction of the process, such as individual level mean, slope or residual variation, particularly as many of these different parameters are highly related to each other. Additionally, it is hard to produce more complex patterns within a phase, for example, a curved temporal pattern. With these limitations, it is probable that a simulation tool based on a hidden Markov chain process would fall short of the requirements discussed earlier.

Hierarchical structure

An alternative approach to stochastic simulations was considered: one based on mathematical functions with random error added. This method of simulation would allow the user to specify a variety of different parameters of interest simultaneously, and in the form that they would likely be estimated by the statistical models under investigation.

The parameters would describe a known function underlying the simulated data, around which random error would be generated. A hierarchical structure would allow some variation in the 'true' function for individual observational units, which could be specified and recorded (Goldstein 2011). This variation could be structured in clusters, such that true functions for individual observational units are more similar within clusters than between them (Goldstein 2011).

To produce phases, similar to the first simulation in this chapter, a function was developed that included two continuous-time functions, joined by a step change. The basic structure of each function is specified, and the mean and amplitude are allowed to vary between observational units. The timing of the step change between functions is also allowed to vary between observational units. The distribution that the random variation follows can be specified, thereby tuning the autocorrelation structure. In addition, observational units can be grouped into specified or randomly selected clusters or subgroups with known average properties. Each cluster has a different average mean, amplitude and step change time for the functions, around which individual level means, amplitudes and step changes are sampled for relevant observational units.

Observed data can then be generated by adding random error to these known individual level functions at specified time points. Different measurement schemes can be generated by specifying or randomly sampling the times at which these observations occur. Temporal autocorrelation can be specified when sampling random error by using a multivariate normal distribution. Spatial autocorrelation can also be incorporated similarly if location information for each observational unit is specified or randomly generated. The layering of autocorrelation structures – temporal, spatial and cluster-based – allows fine tuning of the data generated which helps the user to tailor the complexity to match many applied contexts.

As well as including random variation between observational units, capacity to include a fixed effect of a third variable (with one value for each observational unit) on the amplitude of the first temporal function (before the step change) was included. The third variable must be generated separately before using the simulation function.

This function simulation approach allows for much greater specification of some parameters that a hidden Markov chain process would not (not easily, at least): known individual level continuous-time information, known spatial and temporal autocorrelation, fixed effects, and known subgroup level information. Having known individual level functions means that the value of the function at any time, or information about 'features' of it (e.g. maxima and minima) can be calculated easily.

However, the data generation process, with a function or relationship underlying observed data matches more closely how data are analysed than necessarily how they are generated – in reality, individual level data are generated through some sort of stochastic process like growth and only emulate following an underlying 'function' through some aspects of determinism operating within the stochastic processes driving data generation (Oud et al. 2012). This means that the hierarchical method of simulating data may be considered more of a departure from real data generation processes than the hidden Markov chain method, but its similarity to the models being used to analyse it is convenient for the purposes of testing them. For this reason, a hierarchical structure was chosen for the construction of a simulation tool.

5.3 Simulation tool function

5.3.1 Function process

The function `simulate.trajectory` simulates longitudinal data for one variable for a series of individual observational units. The data takes the form of a pattern or trajectory with two sections joined by a step change. The two sections can take a range of forms parameterised in terms of time. Individual parameters for trajectories are randomly sampled from a population distribution and recorded so they are known. A first order temporal autocorrelation structure can be specified for within-individual error terms. Spatial groups can be specified, either for a point location or origin-destination data, and the mean pa-

rameter values are free to differ randomly between groups, and randomly vary between observational units within them to form clustered data. Between individual error correlations can also be specified. Measurement times can be sampled to fall regularly, in waves or to be completely irregular.

The process the function uses to achieve these aims is outlined in the flow chart in Figure 5.7 which is duplicated in a larger size in Appendix C. Code, required libraries, arguments, function output and common warning messages for the simulation function are available in Appendix B.

5.3.2 Example of use

Figure 5.8 shows example data simulated for 50 individuals over 30 time points. Code for this simulation is shown in Appendix B. The first function section follows a $log(time)$ curve and the second follows a $cos(time) \times time$ curve. The coefficients for these are from a lognormal distribution with parameters 1 and 0.1 and normal distribution with parameters 1 and 0.5, respectively. Individual step change times are chosen from a uniform distribution spanning between times 10 and 20 and the step change value is chosen from a normal distribution with mean -100 and standard deviation 3. While error terms are normally distributed, individual level error standard deviations are chosen from a lognormal distribution. Adjacent measurements within individuals have an error correlation of 0.5. Simultaneous measurements for different individuals have an error correlation of 0.5. Measurements are taken irregularly, 10 times for each individual.

5.3.3 Limitations

Several limitations have been identified for this specific function approach to simulating longitudinal data. Firstly, a step change must be specified for the distribution to run. As the first and second section slope parameters (at the individual level) are randomly selected from orthogonal distributions, the function cannot be manipulated to follow exactly the same shape throughout the full measurement period, but the step change can be set to have a value of 0 for all individuals by specifying a distribution with no variation. It would also be possible to discard data on one side of the step change if desired, or to specify the location of the change after the sampling period. If the latter of these two options is used, some warnings are produced as plots of the distances from the step change and average 2

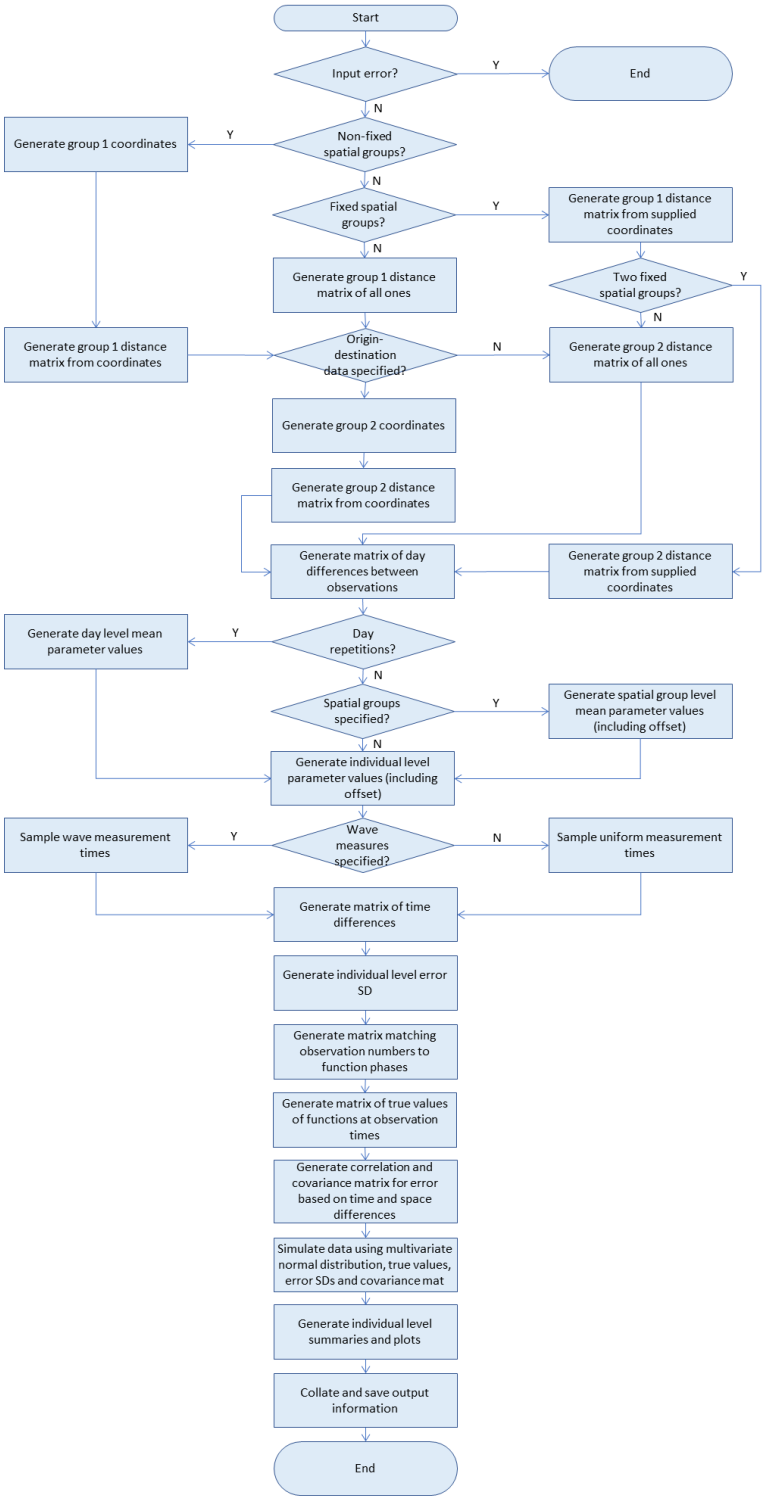


Figure 5.7: Flow diagram describing the processes underlying the simulation function. A large scale version is available in Appendix C

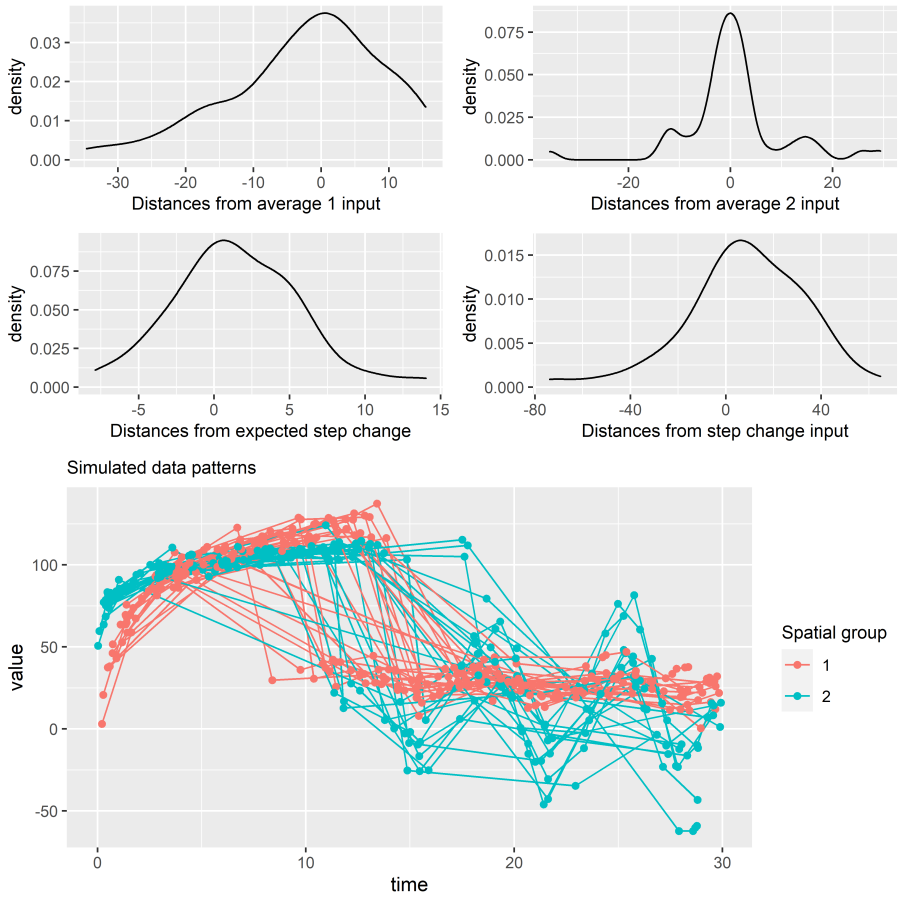


Figure 5.8: Example data generated by the simulation function

inputs cannot be produced. If the step change is required to occur at the same time for all individuals (for example, to represent an external event), a distribution with no variation may be specified for this time.

Secondly, two parameter values must be specified for each distribution involved, which excludes distributions with other numbers of parameters, such as the exponential distribution (R Core Team 2020b).

5.4 Real data – Oyster card sample, Transport for London open data

While simulated data can be constructed with a complex structure and random error, it still often represents an ideal situation, as the level of complexity underlying the generation of real data cannot feasibly be matched. While there is no way of using real data to assess the accuracy of methods, it is important to provide illustrative applications of new methods

methods to real data, on which they are ultimately intended to be used. This provides some information about how computationally intensive they are to apply and, for example, whether the estimation process performs adequately in the presence of the complexity of real data.

Common applications of network analysis are encountered in transport studies. For example, evaluation of road networks, rail networks or public bike-sharing systems (Bagchi and White 2005; Hidas and Wagner 2004). Transport systems with different structures will likely produce data with different structures. For example, in a public bike system where bikes are ridden between docking stations, or a rail system with a 'tap-in-tap-out' card system, it is easy to record the origin and destination of individual journeys made by users (Bagchi and White 2005). For car travel through a road network, there is no central system that automatically collects this information and it is more difficult to collect individual origin-destination trip data about a large volume of trips (Hidas and Wagner 2004). Information about traffic flow through car networks might therefore be collected in different ways, for example, a sensor system may count the number of cars on a particular road segment at a given time of day, or the speed of traffic flow (Hidas and Wagner 2004). Traffic flow counts, if automatic, are likely to occur at regular intervals, whereas origin-destination information about trips will be recorded at irregular intervals, whenever each trip occurs (Hidas and Wagner 2004).

Even though data from both public bike sharing systems and rail systems may be in the form of individual trips, the transport systems themselves have different structures. In a public bike sharing system, users are free to cycle along any route from their origin to destination whereas, in a rail system, a fixed range of movement along railway lines is in place (Bagchi and White 2005). This means that while, for both networks, a graph structure can be used to represent pairs of origin and destination stations for trips, rail systems contain additional network structure describing the railway line connections. This means that vertices and edges in a rail network have additional network-related properties beyond summarising trip information.

While there are a range of datasets available related to transport, one potential advantage of continuous-time methods is their inherent ability to deal with irregular measurement intervals. Trip data from public transport systems, that is measured whenever a trip

occurs, exhibit this structure and therefore would be ideal for testing continuous-time models. The existence of a fixed underlying network gives rail network data interesting properties to examine; for example, how the number or type of connections from a railway station relate to patterns of passenger flow to or from that station. These properties provided a useful basis for including fixed effects in some simulations. The data chosen concerned journeys in the Transport for London Oyster Card system. This included three separate datasets:

- A 5% sample of Oyster Card journeys for one week in November 2009 (Transport for London 2020b).
  - This includes all types of journeys using Oyster Cards (the tap-in-tap-out payment system for Transport for London) – bus, Underground rail, Overground rail, Docklands Light Railway and some National Rail services.
- The locations of London Underground Stations (Transport for London 2020a)
  - Locations were provided using longitude and latitude coordinates. Some stations were missing, these were added using information from Google maps.
- The rail network connections and zone information for each station in the London Underground
  - These data were collected by hand from a September 2009 London Underground Map (Transport for London 2009b) for rail connections and a December 2009 London Underground Map (Transport for London 2009a) for zone information (the September 2009 Map did not include Zone information).

5.4.1 Data Structure

Oyster Card Sample

The Oyster Card data consisted of a 5% random sample of all journeys recorded on Oyster Cards within one (unspecified) week in November 2009. 2623487 journeys were recorded in total. The information recorded about each journey is detailed in Table 5.1.

Variable	Variable meaning	Possible values
downo	Number referring to day of the week	Integer 1-7 (1 refers to Sunday)
daytype	Day of the week	Sun, Mon, Tue, Wed, Thu, Fri, Sat
SubSystem	Transport system used	LUL, NR, LRC, DLR, HEX, LTB, TRAM (London Underground, National Rail, London Overground, Docklands Light Railway, Heathrow Express (Train), London Transport Buses, Trams). Combinations of 2 or 3 subsystems are separated by '/'
StartStn	Origin Station Name or 'Unstarted'	Character
EndStation	Destination Station Name or 'Unfinished'	Character
EntTime	'Tap-in' time (journey start time)	4-digit numeric
EntTimeHHMM	'Tap-in' time (journey start time)	HH:MM time
ExTime	'Tap-out' time (journey end time)	4-digit numeric, for all bus journeys this is set to midnight as no tapping-out takes place
EXTimeHHMM	'Tap-out' time (journey end time)	HH:MM time, for all bus journeys this is set to midnight as no tapping-out takes place
ZVPPT	Zone restrictions for any loaded tickets (Oyster cards can be used on a pay-as-you-go system or by pre-loading travel cards)	Code consisting of a letter and four digits. Letter: B/Z – Bus or Zone First two digits: 01-04

Variable	Variable meaning	Possible values
		Second two digits: 02-10 If following the letter Z, digits indicate the two zones bounding the area in which journeys must take place '—' indicates no travelcard loaded '0' indicates no restriction, e.g. for the Freedom Pass ticket
JNYTYP	Type of payment (Oyster cards can be used on a pay-as-you-go system or by pre-loading travel cards)	TKT, PPY, MIX (Ticket i.e. travel card pre-loaded onto Oyster Card, pay per use, MIX appears to indicate a travelcard is loaded but a fare has been applied, for example, if travel has occurred outside zones covered by the travelcard)
DailyCapping	Whether the fare was capped (a daily cap, dependent on the zones used, is applied to journey fares)	Y/N (Yes or No)
FFare	Full fare applied in pence	Numeric
DFare	Discounted fare applied in pence, for example if daily cap is applied	Numeric
RouteID	Bus route ID (if a bus journey)	Numeric or 'XX' for non-bus journeys
FinalProduct	Method used to pay (Oyster cards can be used on a pay-as-you-go system or by pre-loading travel cards)	PAYG LUL Travelcard-7 Day LUL Travelcard-1 Month LUL Travelcard-Period LUL Travelcard-Annual

Variable	Variable meaning	Possible values
		LUL Travelcard-Annual
		LUL Travelcard-Time Not Captured
		Freedom Pass (Elderly)
		Freedom Pass (Disabled)
		Bus & Tram Pass-1 Month
		Bus & Tram Pass (Child Free)
		Bus & Tram Pass-7 Day
		Bus & Tram Pass – B&T Discount-7 Day
		Bus & Tram Pass – B&T Discount-1 Month
		Bus & Tram Pass-Annual
		Bus & Tram Pass-Period
		Bus & Tram Pass – B&T Discount-Period
		Staff Pass – Staff Nominee
		Staff Pass – Bus Operator
		Staff Pass – Bus Operator Nominee
		Staff Pass – Staff Retired including LCB
		Tfl Travel – Free

Table 5.1: Variables in Oyster Card data

Transport for London Station Locations

Information about the location of 302 Transport for London Stations was available as a KML file, the variables in which are detailed in Table 5.2. Not all stations were included – this is discussed below.

Variable	Variable meaning	Possible values
Name	Station Name	Character string
Description	Station Address (Name, London Underground Ltd., Street, London, Postcode)	Character String
Geometry	Longitude and latitude of station	Numeric vector of length three: longitude, latitude, 0

Table 5.2: Variables in Transport for London station location data

Rail network connections

Data about 305 Transport for London Stations were collected from the London Underground maps from September and December 2009. The September map was published at the most recent date before the Oyster Card sample dates (November 2009) and this was used to collect information about railway connections. The September map did not include zone information, so this was collected from the December 2009 map. The variable details are shown in Table 5.3. The full dataset is available in the Online Appendix and scanned copies of the maps are shown in Appendix D.

Variable	Variable meaning	Possible values
ID	ID number of each station	Integer
Name	Station Name	Character string
Zone	Fare zone of station according to 2010 Underground Map	Integer
Neighbours	ID of all stations directly linked by London Underground Rail line, according to 2009 London Underground Map, separated by underscore	Character formed of integers separated by underscores

Table 5.3: Variables collected describing London Underground network connections

Initially, stations were assigned a numeric ID in alphabetical order. These IDs were added as annotations for stations on scanned copies of the maps. For each station, the ID number of all stations directly connected by a London Underground line (without passing through any other stations and not including Docklands Light Railway or Overground lines, on the September 2009 map) were recorded, separated by an underscore character, in a string. This gave a compact version of the adjacency matrix describing the London Underground network. For each station, the fare zone in which it was located for the December 2009 map was recorded. Stations located in two fare zones (indicated on the maps) were recorded as belonging to the lowest numbered fare zone (e.g. a station in zones 2 and 3 was recorded as

zone 2). In this case, zone location was used only to select a subset of the data to analyse as an example application, so this variable does not form part of any models. Using R, the compact adjacency information was expanded into a full adjacency matrix. This should be symmetric as the connections are undirected (Newman 2018). Checking if the matrix was symmetric was used to check for errors in manual data collection.

Linking Datasets

Linkage of the three datasets was carried out by matching station names. There were three stations not included in the location data (Heathrow Terminal 5, London City Airport and Wood Lane) – these were added using longitude and latitude information collected from Google Maps. Shepherd's Bush Market was not explicitly included in location data, but two Shepherd's Bush stations ('Central' and 'Hammersmith and City') were included – these were taken to refer to the lines running through Shepherd's Bush Station (Central line) and Shepherd's Bush Market (Hammersmith and City line). Therefore, the location and name of 'Shepherd's Bush Hammersmith and City' were changed to represent Shepherd's Bush Market station.

References (Chapters 1-5)

Abdelghany, A., Mahmassani, H., & Chiu, Y. (2001). Spatial microassignment of travel demand with activity trip chains. *Transportation Research Record*, *1777*(1), 36–46.

Aber, M., & McArdle, J. (1991). Latent growth curve approaches to modelling the development of competence. In M. J. Chandler & M. Chapman (Eds.), *Criteria for competence: Controversies in the conceptualization and assessment of children's abilities*. L. Erlbaum Associates.

Albers, H., Hall, K., Lee, K., Taleghan, M., & Dietterich, T. (2018). The role of restoration and key ecological invasion mechanisms in optimal spatial-dynamic management of invasive species. *Ecological Economics*, *151*, 44–54.

Albert, P. (1999). Longitudinal data analysis (repeated measures) in clinical trials. *18*(13), 1707–1732.

Anderson, T., & Dragicevic, S. (2018). Network-agent based model for simulating the dynamic spatial network structure of complex ecological systems. *Ecological Modelling*, *389*, 19–32.

Anderson, T., & Dragićević, S. (2018). A geographic network automata approach for modeling dynamic ecological systems. *Geographical Analysis*, *52*, 3–27.

Anderson, T., & Dragićević, S. (2020a). NEAT approach for testing and validation of geospatial network agent-based model processes: Case study of influenza spread. *International Journal of Geographical Information Science*, *34*(9), 1792–1821.

Anderson, T., & Dragićević, S. (2020b). Representing complex evolving spatial networks: Geographic network automata. *ISPRS International Journal of Geo-Information*, *9*(4).

Apparicio, P., Abdelmajid, M., Riva, M., & Shearmur, R. (2008). Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics*, *7*(1), 7.

Aydin, B., Leite, W., & Algina, J. (2014). The consequences of ignoring variability in measurement occasions within data collection waves in latent growth models. *Multivariate Behavioral Research*, *49*(2), 149–160.

Bagchi, M., & White, P. (2005). The potential of public transport smart card data. *Transport Policy*, *12*(5), 464–474.

Barker, D., Osmond, C., Forsén, T., Kajantie, E., & Eriksson, J. (2005). Trajectories of growth among children who have coronary events as adults. *New England Journal of Medicine*, *353*(17), 1802–1809.

Basile, R., Parteka, A., & Pittiglio, R. (2018). Export diversification and economic development: A dynamic spatial data analysis. *Review of International Economics*, *26*(3), 634–650.

Baybeck, B., & Huckfeldt, R. (2002). Urban contexts, spatially dispersed networks, and the diffusion of political information. *Political Geography*, *21*(2), 195–220.

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), 1–20.

Bi, Q., Goodman, K., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American Journal of Epidemiology*, *188*(12), 2222–2239.

Bjork, T. (2004). *Arbitrage theory in continuous time*. Oxford University Press.

Bland, J., & Altman, D. (1999). Measuring agreement in method comparison studies. *Stat Methods Med Res*, *8*(2), 135–60.

Blonder, B., Wey, T., Dornhaus, A., James, R., & Sih, A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, *3*(6), 958–972.

Blozis, S., Conger, K., & Harring, J. (2007). Nonlinear latent curve models for multivariate longitudinal data. *International Journal of Behavioral Development*, *31*(4), 340–346.

Bollen, K., & Curran, P. (2004). Autoregressive latent trajectory (ALT) models a synthesis of two traditions. *Sociological Methods & Research*, *32*(3), 336–383.

Bollen, K., & Curran, P. (2005). *Latent curve models: A structural equation perspective*. John Wiley & Sons.

Browne, W., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *J Bayesian analysis*, *1*(3), 473–514.

Browne, W., Draper, D., Goldstein, H., & Rasbash, J. (2002). Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. *Computational Statistics & Data Analysis*, *39*(2), 203–225.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 28.

Burnett, J., & Zhao, X. [X.]. (2017). Spatially explicit prediction of wholesale electricity prices. *International Regional Science Review*, *40*(2), 99–140.

Byrne, B., & Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics*, *2*(3), 177–203.

Cai, H., Rahman, A., Su, X., & Zhang, H. (2014). A GIS-microscopic simulation approach for optimizing road barrier placement and configuration in university campus emergency evacuation. *International Journal of Disaster Resilience in the Built Environment*, *5*(4), 362–379.

Carlsson, A., & Kindvall, O. (2001). Spatial dynamics in a metapopulation network: Recovery of a rare grasshopper Stauroderus scalaris from population refuges. *Ecography*, *24*(4), 452–460.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *J Journal of Statistical Software*, *76*(1), 32.

Carter, N., Baeza, A., & Magliocca, N. (2020). Emergent conservation outcomes of shared risk perception in human-wildlife systems. *Conservation Biology*, *34*(4), 903–914.

Caruso, M. (2007). Disruptive dynamics: The spatial dimensions of the spanish networks in the spread of monitorial schooling (1815-1825). *Paedagogica Historica*, *43*(2), 271–282.

Chen, L., Han, K., Yin, Q., & Cao, Z. (2020). GDCRN: Global diffusion convolutional residual network for traffic flow prediction. https://doi.org/10.1007/978-3-030-55393-7_39

Chen, S., Ilany, A., White, B., Sanderson, M., & Lanzas, C. (2015). Spatial-temporal dynamics of high-resolution animal networks: What can we learn from domestic animals? *PLOS One*, *10*(6), e0129253.

Cheng, S., Xie, B., Bie, Y., Zhang, Y., & Zhang, S. (2018). Measure dynamic individual spatial-temporal accessibility by public transit: Integrating time-table and passenger departure time. *Journal of Transport Geography*, *66*, 235–247.

Cheng, T., Wang, J., Haworth, J., Heydecker, B., & Chow, A. (2014). A dynamic spatial weight matrix and localized space–time autoregressive integrated moving average for network modeling. *J Geographical Analysis*, *46*(1), 75–97.

Colak, S., Schneider, C., Wang, P., & Gonzalez, M. (2013). On the role of spatial dynamics and topology on network flows. *New Journal of Physics*, *15*, 113037.

Cole, T., Donaldson, M., & Ben-Shlomo, Y. (2010). SITAR - A useful instrument for growth curve analysis. *International Journal of Epidemiology*, *39*(6), 1558–1566.

Comber, A., & Wulder, M. (2019). Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS*, *23*(5), 879–891.

Crooks, A., & Heppenstall, A. (2011). Introduction to agent-based modelling. In A. J. Heppenstall, M. Batty, A. T. Crooks, & L. M. See (Eds.), *Agent-based models of geographical systems* (First edition, pp. 95–105). Springer Netherlands.

De Boor, C. (2001). *A practical guide to splines* (Revised edition). Springer.

de Haan-Rietdijk, S., Voelkle, M., Keijsers, L., & Hamaker, E. (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*, 1849–1849.

Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics*, *21*(3-4), 224–239.

Diao, Z., Wang, X., Zhang, D., Liu, Y., Xie, K., & He, S. (2019). Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 890–897.

Ditlevsen, S., & De Gaetano, A. (2005). Mixed effects in stochastic differential equation models. *REVSTAT-Statistical Journal*, *3*(2), 137–153.

Donfouet, H., Jeanty, P., & Malin, E. (2018). Analysing spatial spillovers in corruption: A dynamic spatial panel data approach. *Papers in Regional Science*, *97*, S63–S78.

Dubin, R. (2009). Spatial weights. In A. S. Fotheringham & P. A. Rogerson (Eds.), *The SAGE handbook of spatial analysis*. SAGE Publications, Ltd.

Duncan, T., Duncan, S., & Strycker, L. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (Second edition). Lawrence Erlbaum Associates.

Ermagun, A., & Levinson, D. (2018). An introduction to the network weight matrix. *Geographical Analysis*, *50*(1), 76–96.

Evans, L. (2013). *An introduction to stochastic differential equations* (Vol. 82). American Mathematical Society.

Færch, K., Witte, D., Tabák, A., Perreault, L., Herder, C., Brunner, E., Kivimäki, M., & Vistisen, D. (2013). Trajectories of cardiometabolic risk factors before diagnosis of three subtypes of type 2 diabetes: A post-hoc analysis of the longitudinal Whitehall II cohort study. *Lancet Diabetes Endocrinol*, *1*(1), 43–51.

Fan, X., Li, X., Yin, J., & Liang, J. (2019). Temporal characteristics and spatial homogeneity of virtual water trade: A complex network analysis. *Water Resources Management*, *33*(4), 1467–1480.

Fang, S., Zhang, Q., Meng, G., Xiang, S., & Pan, C. (2019). GSTnet: Global spatial-temporal network for traffic flow prediction. *2019-August*, 2286–2293.

Fang, Y., Wang, Z., Cheng, R., Li, X., Luo, S., Hu, J., & Chen, X. (2019). On spatial-aware community search. *IEEE Transactions on Knowledge and Data Engineering*, *31*(4), 783–798.

Feng, D., Wu, Z., Zhang, J., & Wu, Z. (2020). Dynamic global-local spatial-temporal network for traffic speed prediction. *IEEE Access*, *8*, 209296–209307.

Ferenc, S., Tamas, P., & Jozsef, B. (2016). Complex analysis of the dynamic effects of car population along the trajectories. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*.

Fielding, A., & Goldstein, H. (2006). Cross-classified and multiple membership structures in multilevel models : An introduction and review (Research Report No. 791).

Fitzmaurice, G., & Ravichandran, C. (2008). A primer in longitudinal data analysis. *Circulation*, *118*(19), 2005–2010.

Fotheringham, A., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, *47*(4), 431–452.

Francisco-Fernández, M., & Quintela-del-Río, A. (2016). Comparing simultaneous and pointwise confidence intervals for hydrological processes. *PLOS ONE*, *11*(2), e0147505.

Freeman, J. (1989). Systematic sampling, temporal aggregation, and the study of political relationships. *Political Analysis*, *1*, 61–98.

Freeman, L. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Garbin, S., Celegon, E., Fanton, P., & Botter, G. (2019). Hydrological controls on river network connectivity. *Royal Society Open Science*, *6*(2).

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721–741.

Gesler, W. (1986). The uses of spatial analysis in medical geography: A review. *Social Science & Medicine*, *23*(10), 963–973.

Gneiting, T., Genton, M., & Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability, and full symmetry. In B. Finkenstadt, L. Held, & V. Isham (Eds.), *Statistical methods for spatio-temporal systems*. CRC Press.

Golding, J. (2004). The Avon longitudinal study of parents and children (ALSPAC) - Study design and collaborative opportunities. *European Journal of Endocrinology*, *151*(Suppl 3), U119.

Goldstein, H. (1994). Multilevel cross-classified models. *J Sociological Methods Research*, *22*(3), 364–375.

Goldstein, H. (2011). *Multilevel statistical models* (Fourth edition). Wiley.

Goldstein, H., Healy, M., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *J Statistics in medicine*, *13*(16), 1643–1655.

Griffith, D. (2012). Space, time, and space-time eigenvector filter specifications that account for autocorrelation. *J Estadística española*, *54*(177), 7–34.

Grimm, K., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*(5), 1357–1371.

Guo, D., Zou, Y., Wu, Y., & Liu, M. (2013). Altruistic behavior in complex network based on reputation and future expectation. *2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics*, *2*, 35–38.

Guo, S., Lin, Y., Feng, N., Song, C., Wan, H., & Aaai. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Thirty-Third Aaai Conference on Artificial Intelligence / Thirty-First Innovative Applications of Artificial Intelligence Conference / Ninth Aaai Symposium on Educational Advances in Artificial Intelligence*, 922–929.

Guo, Y., Yang, L., & Gao, J. (2020). Coordinated perimeter control for multiregion heterogeneous networks based on optimized transfer flows. *Mathematical Problems in Engineering*, *2020*.

Hägerstraand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, *24*(1), 6–21.

Hall, K., Albers, H., Taleghan, M., & Dietterich, T. (2018). Optimal spatial-dynamic management of stochastic species invasions. *Environmental & Resource Economics*, *70*(2), 403–427.

Han, X., Hsieh, C., & Ko, S. (2019). Spatial modeling approach for dynamic network formation and interactions. *Journal of Business & Economic Statistics*.

Hawawini, G. (1978). A note on temporal aggregation and serial correlation. *Economics Letters*, *1*(3), 237–242.

He, B., Xu, Z., Xu, Y., Hu, J., & Ma, Z. (2020). Integrating semantic zoning information with the prediction of road link speed based on taxi GPS data. *Complexity*, *2020*.

Head, M., Holman, L., Lanfear, R., Kahn, A., & Jennions, M. (2015). The extent and consequences of P-hacking in science. *PLOS Biology*, *13*(3), e1002106.

Herrett, E., Gallagher, A., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., & Smeeth, L. (2015). Data resource profile: Clinical practice research datalink (CPRD). *International Journal of Epidemiology*, *44*(3), 827–836.

Hidas, P., & Wagner, P. (2004). Review of data collection methods for microscopic traffic simulation. *World Conference on Transport Research (WCTR)*.

Higham, D. J., Batty, M., Bettencourt, L. M. A., Greetham, D. V., & Grindrod, P. (2017). An overview of city analytics. *Royal Society Open Science*, *4*, 161063.

Hoffman, M., & 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.

Horvath, S. (2011). *Weighted network analysis: Applications in genomics and systems biology*. Springer.

Howe, L., Tilling, K., Matijasevich, A., Petherick, E., Santos, A., Fairley, L., Wright, J., Santos, I., Barros, A., Martin, R., Kramer, M., Bogdanovich, N., Matush, L., Barros, H., & Lawlor, D. (2013). Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research*, *25*(5), 1854–1874.

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, *24*(3), 383–401.

Hulsegge, G., Spijkerman, A., van der Schouw, Y., Bakker, S., Gansevoort, R., Smit, H., & Verschuren, W. (2017). Trajectories of metabolic risk factors and biochemical markers prior to the onset of type 2 diabetes: The population-based longitudinal Doetinchem study. *Nutr Diabetes*, *7*(5), e270.

Hwang, S. (2000). The effects of systematic sampling and temporal aggregation on discrete time long memory processes and their finite sample properties. *Econometric Theory*, *16*(3), 347–372.

Iotti, B., Antonioni, A., Bullock, S., Darabos, C., Tomassini, M., & Giacobini, M. (2017). Infection dynamics on spatial small-world network models. *Physical Review E*, *96*(5), 052316.

Jacoby, D., Brooks, E., Croft, D., & Sims, D. (2012). Developing a deeper understanding of animal movements and spatial dynamics through novel application of network analyses. *Methods in Ecology and Evolution*, *3*(3), 574–583.

Jeong, H., & Lee, L. (2020). Spatial dynamic models with intertemporal optimization: Specification and estimation. *Journal of Econometrics*, *218*(1), 82–104.

Jia, J., Tan, S., Ji, G., & Zhao, B. (2019). Discovering hotspots in dynamic spatial networks using mobility data, 102–107.

Jin, G., Cui, Y., Zeng, L., Tang, H., Feng, Y., & Huang, J. (2020). Urban ride-hailing demand prediction with multiple spatiotemporal information fusion network. *Transportation Research Part C-Emerging Technologies*, *117*.

Jin, Y., Peng, R., & Shi, J. (2019). Population dynamics in river networks. *Journal of Nonlinear Science*, *29*(6), 2501–2545.

John, F. (2009). An introduction to calculus. *A mathematical primer for social statistics*. SAGE Publications, Inc.

Johnson, W. (2014). Analytical strategies in human growth research. *American Journal of Human Biology*, *27*(1), 69–83.

Jones, P. (2017). Markov chains. *Stochastic processes: An introduction*. Chapman & Hall.

Jones, P., & Smith, P. (2018). Markov chains.

Jones, R., & Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics*, *47*(1), 161–175.

Jovanovic, T., Hale, R., Gironas, J., & Mejia, A. (2019). Hydrological functioning of an evolving urban stormwater network. *Water Resources Research*, *55*(8), 6517–6533.

Kraak, M.-J. (2003). The space-time cube revisited from a geovisualization perspective. *Proc. 21st International Cartographic Conference*, 1988–1996.

Kuersteiner, G., & Prucha, I. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, *88*(5), 2109–2146.

Kwan, M.-P. (2013). Beyond space (as we knew it): Toward temporally integrated geographies of segregation, health, and accessibility. *Annals of the Association of American Geographers*, *103*(5), 1078–1086.

Leenders, R. (2002). Modeling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks*, *24*(1), 21–47.

Lengyel, B., Bokanyi, E., Di Clemente, R., Kertesz, J., & Gonzalez, M. (2020). The role of geography in the complex diffusion of innovations. *Scientific Reports*, *10*(1).

Li, G., Knoop, V., & Lint, H. (2020). Dynamic graph filters networks: A gray-box model for multistep traffic forecasting.

Li, T., & Liao, Q. (2016). Dynamic networks analysis and visualization through spatiotemporal link segmentation. *IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 209–214.

Li, W., Wang, X., Zhang, Y., & Wu, Q. (2021). Traffic flow prediction over muti-sensor data correlation with graph convolution network. *Neurocomputing*, *427*, 50–63.

Li, X., & Griffin, W. (2013). Using ESDA with social weights to analyze spatial and social patterns of preschool children's behavior. *Applied Geography*, *43*, 67–80.

Li, Y., Bi, J., & Sun, H. (2008). Spatial price dynamics: From complex network perspective. *Physica a-Statistical Mechanics and Its Applications*, *387*(23), 5852–5856.

Li, Z., Li, L., Peng, Y., & Tao, X. (2020). A two-stream graph convolutional neural network for dynamic traffic flow forecasting. *2020-November*, 355–362.

Liggett, T. (2010). *Continuous time markov processes : An introduction*. American Mathematical Society.

Lin, H., Bai, R., Jia, W., Yang, X., & You, Y. (2020). Preserving dynamic attention for long-term spatial-temporal prediction, 36–46.

Liu, C., Gao, C., & Xin, Y. (2018). Measuring the diversity and dynamics of mobility patterns using smart card data. *Knowledge Science, Engineering and Management"*, 438–451.

Liu, S., Wang, L., Liu, C., & Destech Publicat, I. (2018). Visualized social network analysis on spatial dynamics of international trade between China and League of Arab States. *2018 3rd international conference on computational modeling, simulation and applied mathematics* (pp. 270–277).

Locantore, N., Marron, J., Simpson, D., Tripoli, N., Zhang, J., Cohen, K., Boente, G., Fraiman, R., Brumback, B., Croux, C., Fan, J., Kneip, A., Marden, J., Peña, D., Prieto, J., Ramsay, J., Valderrama, M., Aguilera, A., Locantore, N., ... Cohen, K. (1999). Robust principal component analysis for functional data. *Test*, *8*(1), 1–73.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine*, *28*, 3049–3067.

Luptáková, I., & Pospichal, J. (2018). Community detection based clustering. *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*, 000289–000294.

Ma, X., Li, Y., & Chen, P. (2020). Identifying spatiotemporal traffic patterns in large-scale urban road networks using a modified nonnegative matrix factorization algorithm. *Journal of Traffic and Transportation Engineering-English Edition*, *7*(4), 529–539.

Malaba, T., Phillips, T., Le Roux, S., Brittain, K., Zerbe, A., Petro, G., Ronan, A., McIntyre, J., Abrams, E., & Myer, L. (2017). Antiretroviral therapy use during pregnancy and adverse birth outcomes in South African women. *International Journal of Epidemiology*, *46*(5), 1678–1689.

Mateus-Anzola, J., Wiratsudakul, A., Rico-Chavez, O., & Ojeda-Flores, R. (2019). Simulation modeling of influenza transmission through backyard pig trade networks in a wildlife/livestock interface area. *Tropical Animal Health and Production*, *51*(7), 2019–2024.

Matthews, A., Herrett, E., Gasparrini, A., Van Staa, T., Goldacre, B., Smeeth, L., & Bhaskaran, K. (2016). Impact of statin related media coverage on use of statins: Interrupted time series analysis with uk primary care data. *BMJ*, *353*, i3283.

McLachlan, G., & Peel, D. (2000). Hidden markov models. *Finite mixture models* (pp. 326–341).

Mehta, P., & West, S. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*(1), 23–43.

Mendoza-Cota, J., & Torres-Preciado, V. (2019). The impact of regional remittances on economic growth in mexico: A dynamic space-time panel approach. *Papeles De Poblacion*, *25*(101), 113–144.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122.

Morris, T., White, I., & Crowther, M. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102.

Muthén, L., & Muthén, B. (2010). *Mplus user's guide* (Sixth edition).

Muthukrishnan, R., & Rohini, R. (2016). LASSO: A feature selection technique in predictive modeling for machine learning. *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, 18–20.

Neale, M., Boker, S., Xie, G., & Maes, H. (2003). *Mx: Statistical modeling.* (Sixth edition).

Neeson, T., Wiley, M., Adlerstein, S., & Riolo, R. (2012). How river network structure and habitat availability shape the spatial dynamics of larval sea lampreys. *Ecological Modelling*, *226*, 62–70.

Neto, P., Friesz, T., & Han, K. (2016). Electric power network oligopoly as a dynamic stackelberg game. *Networks and Spatial Economics*, *16*(4), 1211–1241.

Newman, M. (2002). The structure and function of networks. *Computer Physics Communications*, *147*(1), 40–45.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256.

Newman, M. (2018). Fundamentals of network theory. *Networks* (Second edition). Oxford University Press.

Niezink, N., Snijders, T., & van Duijn, M. (2019). No longer discrete: Modeling the dynamics of social networks and continuous behavior. *Sociological Methodology*, *49*(1), 295–340.

Oka, R., Shibata, K., Sakurai, M., Kometani, M., Yamagishi, M., Yoshimura, K., & Yoneda, T. (2017). Trajectories of postload plasma glucose in the development of type 2 diabetes in Japanese adults. *J Diabetes Res*, *2017*, 5307523.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251.

Ou, J., Sun, J., Zhu, Y., Jin, H., Liu, Y., Zhang, F., Huang, J., & Wang, X. (2020). STP-TrellisNets: Spatial-temporal parallel trellisnets for metro station passenger flow prediction, 1185–1194.

Oud, J., Folmer, H., Patuelli, R., & Nijkamp, P. (2012). Continuous-time modeling with spatial dependence. *Geographical Analysis*, *44*(1), 29–46.

Parent, O., & LeSage, J. (2010). A spatial dynamic panel model with random effects applied to commuting times. *Transportation Research Part B-Methodological*, *44*(5), 633–645.

Patuelli, R., Griffith, D., Tiefelsdorf, M., & Nijkamp, P. (2006). The use of spatial filtering techniques: The spatial and space-time structure of german unemployment data. *Tinbergen Institute Discussion Papers*.

Peng, H., Wang, H., Du, B., Bhuiyan, M., Ma, H., Liu, J., Wang, L., Yang, Z., Du, L., Wang, S., & Yu, P. (2020). Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. *Information Sciences*, *521*, 277–290.

Perez, L., & Dragicevic, S. (2009). An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics*, *8*(1).

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. *BMC Medical Research Methodology*, *19*(1), 46.

Perry, G., & Lee, F. (2019). How does temporal variation in habitat connectivity influence metapopulation dynamics? *Oikos*, *128*(9), 1277–1286.

Pfeifer, P., & Deutsch, S. (1980). A STARIMA model-building procedure with application to description and regional forecasting. *J Transactions of the Institute of British Geographers*, 330–349.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*, *124*.

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks, 284–293.

Porta, M. (2014). *A dictionary of epidemiology*. Oxford University Press, Incorporated.

Pullenayegum, E., & Lim, L. (2016). Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. *25*(6), 2992–3014.

Qu, Y., Zhu, Y., Zang, T., Xu, Y., & Yu, J. (2020). Modeling local and global flow aggregation for traffic flow forecasting. https://doi.org/10.1007/978-3-030-62005-9_30

R Core Team. (2020b). R: The exponential distribution. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/Exponential.html

Rabash, J., Steele, F., Browne, W., & Goldstein, H. (2020). *A user's guide to MLwiN, v3.05*. Centre for Multilevel Modelling, University of Bristol.

Ramachandran, K., Sikdar, B., & Ieee. (2007a). Modeling malware propagation in networks of smart cell phones with spatial dynamics. *IEEE infocom 2007* (pp. 2516–2520).

Ramachandran, K., Sikdar, B., & Ieee. (2007b). On the stability of the malware free equilibrium in cell phones networks with spatial dynamics. *2007 IEEE International Conference on Communications*, *1-14*, 6169–6174.

Ramsay, J., & Silverman, B. (1997). *Functional data analysis*. Springer.

Rasmussen, P., Staggs, M., Beard Jr., T., & Newman, S. (1998). Bias and confidence interval coverage of creel survey estimators evaluated by simulation. *127*(3), 469–480.

Rossana, R., & Seater, J. (1995). Temporal aggregation and economic time series. *Journal of Business & Economic Statistics*, *13*(4), 441–451.

Royston, P., & Altman, D. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *43*(3), 429–467.

Ryser, R., Haussler, J., Stark, M., Brose, U., Rall, B., & Guill, C. (2019). The biggest losers: Habitat isolation deconsructs complex food webs from top to bottom. *Proceedings of the Royal Society B-Biological Sciences*, *286*(1908).

Saberi, M., Mahmassani, H., Brockmann, D., & Hosseini, A. (2017). A complex network perspective for characterizing urban travel demand patterns: Graph theoretical analysis of large-scale origin–destination demand networks. *Transportation*, *44*(6), 1383–1402.

Scott, J. (1988). Social network analysis. *Sociology*, *22*(1), 109–127.

Semboloni, F. (2000). The growth of an urban cluster into a dynamic self-modifying spatial pattern. *Environment and Planning B: Planning and Design*, *27*(4), 549–564.

Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*, *22*(11), 1359–66.

Srivastava, A., & Salapaka, S. (2020). Simultaneous facility location and path optimization in static and dynamic networks. *IEEE Transactions on Control of Network Systems*, *7*(4), 1700–1711.

Stan Development Team. (2019). RStan: The R interface to stan. R package version 2.19.2.

Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*(1), 5–19.

Sterba, S. (2014). Fitting nonlinear latent growth curve models with individually varying time points. *Structural Equation Modeling-a Multidisciplinary Journal*, *21*(4), 630–647.

Sun, B., Zhao, D., Shi, X., & He, Y. (2021). Modeling global spatial-temporal graph attention network for traffic prediction. *IEEE Access*, *9*, 8581–8594.

Sun, G., Tang, T., Peng, T., Liang, R., & Wu, Y. (2017). SocialWave: Visual analysis of spatio-temporal diffusion of information on social media. *ACM Transactions on Intelligent Systems and Technology*, *9*(2), 2157–6904.

Tabák, A., Jokela, M., Akbaraly, T., Brunner, E., Kivimäki, M., & Witte, D. (2009). Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: An analysis from the Whitehall II study. *The Lancet*, *373*(9682), 2215–2221.

Thomas, A., Best, N., Lunn, D., Arnold, R., & Spiegelhalter, D. (2014). GeoBUGS user manual.

Thompson, S. (1919). *Calculus made easy: Being a very-simplest introduction to those beautiful methods of rekoning which are generally called by the terrifying names of the differential calculus and the integral calculus* (Second edition). Macmillan.

Tian, K., Guo, J., Ye, K., & Xu, C. (2020). ST-MGAT: Spatial-temporal multi-head graph attention networks for traffic forecasting. *2020-November*, 714–721.

Tobler, W. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, *46*, 234–240.

Transport for London. (2009a). London Underground map December 2009.

Transport for London. (2009b). London Underground map September 2009.

Transport for London. (2020a). Our open data. https://tfl.gov.uk/info-for/open-data-users/our-open-data

Transport for London. (2020b). Transport for London unified API. https://api.tfl.gov.uk/

Tu, Y.-K., D'Aiuto, F., Baelum, V., & Gilthorpe, M. (2009). An introduction to latent growth curve modelling for longitudinal continuous data in dental research. *European Journal of Oral Sciences*, *117*(4), 343–350.

Tu, Y.-K., & Gilthorpe, M. (2010). *Statistical thinking in epidemiology*. Chapman & Hall/CRC.

Tu, Y.-K., Tilling, K., Sterne, J., & Gilthorpe, M. (2013). A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. *International Journal of Epidemiology*, *42*(5), 1327–1339.

Ugrinowitsch, C., Fellingham, G., & Ricard, M. (2004). Limitations of ordinary least squares models in analyzing repeated measures data. *36*(12), 2144–2148.

Van Wissen, L., & Meurs, H. (1989). The Dutch mobility panel: Experiences and evaluation. *Transportation*, *16*(2), 99–119.

Voelkle, M., Oud, J., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*(2), 176–192.

Wang, J.-L., Chiou, J.-M., & Müller, H.-G. (2016). Functional data analysis. *3*(1), 257–295.

Wang, W., & Yan, J. (2020). splines2: Regression spline functions and classes.

Wang, X., Ge, J., Wei, W., Li, H., Wu, C., & Zhu, G. (2016). Spatial dynamics of the communities and the role of major countries in the international rare earths trade: A complex network analysis. *PLOS One*, *11*(5), e0154575.

Wanzenböck, I., & Piribauer, P. (2018). R&d networks and regional knowledge production in europe: Evidence from a space-time model. *Papers in Regional Science*, *97*, S1–S24.

Ward, J., Evans, A., & Malleson, N. (2016). Dynamic calibration of agent-based models using data assimilation. *Royal Society Open Science*, *3*(4), 150703.

Wei, V., Wong, R., & Long, C. (2020). Architecture-intact oracle for fastest path and time queries on dynamic spatial networks, 1841–1856.

Wei, W. (1981). Effect of systematic sampling on ARIMA models. *Communications in Statistics - Theory and Methods*, *10*(23), 2389–2398.

Weiss, A. (1984). Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics*, *26*(3), 271–281.

Wilkinson, J., Arnold, K., Murray, E., van Smeden, M., Carr, K., Sippy, R., de Kamps, M., Beam, A., Konigorski, S., Lippert, C., Gilthorpe, M., & Tennant, P. (2020). Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health*, *2*(12), e677–e680.

Wolf, C., & Best, H. (2014a). Linear regression. In C. Wolf & H. Best (Eds.), *The SAGE handbook of regression analysis and causal inference*. SAGE Publications Ltd.

Wolf, C., & Best, H. (2014b). Non-linear and non-additive effects in linear regression. In C. Wolf & H. Best (Eds.), *The SAGE handbook of regression analysis and causal inference*. SAGE Publications Ltd.

Wolf, C., & Best, H. (2014c). Regression analysis: Assumptions and diagnostics. In C. Wolf & H. Best (Eds.), *The SAGE handbook of regression analysis and causal inference*. SAGE Publications Ltd.

Wu, J., Roy, J., & Stewart, W. (2010). Prediction modeling using EHR data: Challenges, strategies, and a comparison of machine learning approaches. *Medical Care*, *48*(6), S106–S113.

Xu, M., & Gao, Z. (2009). Chaos in a dynamic model of urban transportation network flow based on user equilibrium states. *Chaos Solitons & Fractals*, *39*(2), 586–598.

Yang, B., Kang, Y., Li, H., Zhang, Y., Yang, Y., & Zhang, L. (2020). Spatio-temporal expand-and-squeeze networks for crowd flow prediction in metropolis. *IET Intelligent Transport Systems*, *14*(5), 313–322.

Yao, H., Tang, X., Wei, H., Zheng, G., Li, Z., & Aaai. (2019). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. *Thirty-Third Aaai Conference on Artificial Intelligence / Thirty-First Innovative Applications of Artificial Intelligence Conference / Ninth Aaai Symposium on Educational Advances in Artificial Intelligence*, 5668–5675.

Yeh, R., Nashed, Y., Peterka, T., & Tricoche, X. (2020). Fast automatic knot placement method for accurate b-spline curve fitting. *Computer-Aided Design*, *128*, 102905.

Yu, H., Fang, Z., Lu, F., Murray, A., Zhao, Z., Xu, Y., & Yang, X. (2019). Massive automatic identification system sensor trajectory data-based multi-layer linkage network dynamics of maritime transport along 21st-century maritime silk road. *Sensors*, *19*(19).

Yu, W., Guan, M., & Chen, Z. (2019). Analyzing spatial community pattern of network traffic flow and its variations across time based on taxi gps trajectories. *Applied Sciences-Basel*, *9*(10).

Yuan, Y., Chen, N., & Zhou, S. (2013). Adaptive b-spline knot selection using multi-resolution basis set. *IIE Transactions*, *45*(12), 1263–1277.

Zambrano-Monserrate, M., Ruano, M., Ormeño-Candelario, V., & Sanchez-Loor, D. (2020). Global ecological footprint and spatial dependence between countries. *Journal of Environmental Management*, *272*.

Zhang, C., Yu, J., & Liu, Y. (2019). Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting. *IEEE Access*, *7*, 166246–166256.

Zhang, H., Zhang, L., Che, F., Jia, J., & Shi, B. (2020). Revealing urban traffic demand by constructing dynamic networks with taxi trajectory data. *IEEE Access*, *8*, 147673–147681.

Zhang, W., Tian, Z., Zhang, G., & Dong, G. (2019). Spatial-temporal characteristics of green travel behavior based on vector perspective. *Journal of Cleaner Production*, *234*, 549–558.

Zhang, X., Chen, G., Wang, J., Li, M., & Cheng, L. (2019). A GIS-based spatial-temporal autoregressive model for forecasting marine traffic volume of a shipping network. *Scientific Programming*, *2019*.

Zhang, Z., Li, M., Lin, X., Wang, Y., & He, F. (2019). Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. *Transportation Research Part C-Emerging Technologies*, *105*, 297–322.

Zhao, N., Ye, Z., Pei, Y., Liang, Y., & Niyato, D. (2020). Spatial-temporal attention-convolution network for citywide cellular traffic prediction. *IEEE Communications Letters*, *24*(11), 2532–2536.

Zhao, X. [X.Y.], & Pu, Y. (2018). The complex system of interprovincial migration flows in China, 1985-2015: A spatial dynamic panel modeling approach. In S. Hu, X. Ye, K. Yang, & H. Fan (Eds.), *2018 26th international conference on geoinformatics*.

Zheng, L., Yang, J., Chen, L., Sun, D., & Liu, W. (2020). Dynamic spatial-temporal feature optimization with ERI big data for short-term traffic flow prediction. *Neurocomputing*, *412*, 339–350.

Zhou, F., Yang, Q., Zhang, K., Trajcevski, G., Zhong, T., & Khokhar, A. (2020). Reinforced spatiotemporal attentive graph neural networks for traffic forecasting. *IEEE Internet of Things Journal*, *7*(7), 6414–6428.

Zhou, S., & Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association*, *96*(453), 247–259.

Zhu, X., & Guo, D. (2014). Mapping large spatial flow data with hierarchical clustering. *18*(3), 421–435.

Zoppou, C. (2001). Review of urban storm water models. *Environmental Modelling & Software*, *16*(3), 195–231.

Zyphur, M., & Oswald, F. (2015). Bayesian estimation and inference: A user's guide. *Journal of management*, *41*(2), 390–420.

Chapter 6

Analysing trajectories of a longitudinal exposure: A causal perspective on common methods in lifecourse research

Citation: Gadd SC, Tennant PWG, Heppenstall AJ, Boehnke JR, Gilthorpe MS (2019) Analysing trajectories of a longitudinal exposure: A causal perspective on common methods in lifecourse research. PLoS ONE 14(12): e0225217. https://doi.org/10.1371/journal.pone.0225217

This chapter completes the second aim of the thesis - to identify continuous time methods suitable for analysis of spatio-temporal networks. In Chapter 3, methods that condition on the outcome when analysing temporal patterns of variables were discussed. There are a number of papers in the literature, particularly within lifecourse epidemiology, that use these methods. For example, Oka et al. (2017) investigate 'trajectories' of plasma glucose levels over nine years before type-2 diabetes diagnosis using a multilevel model with time-diabetes interaction term. Sabia et al. (2017) use a similar model when studying 'trajectories' of physical activity in the years before participants in a cohort study did or did not develop dementia. Various papers take the approach of plotting z-scores over time (Bhargava et al. 2004; J. Eriksson et al. 2000; J.G. Eriksson et al. 2003), including

Forsén et al. (2004) and Barker et al. (2005) which both plot z-scores of body mass index, weight and height throughout childhood in individuals who later experience coronary heart disease.

While methods that condition on the outcome have been mentioned elsewhere in the literature (Tu and Gilthorpe 2010; Tu, Tilling, et al. 2013), no evidence to show the biased inferences they lead to has been produced. This paper produces evidence to show this and therefore eliminates methods that condition on the outcome from those suitable for analysing spatio-temporal networks.

6.1 Abstract

Longitudinal data is commonly analysed to inform prevention policies for diseases that may develop throughout life. Commonly methods interpret the longitudinal data as a series of discrete measurements or as continuous patterns. Some of the latter methods condition on the outcome, aiming to capture 'average' patterns within outcome groups, while others capture individual-level pattern features before relating these to the outcome. Conditioning on the outcome may prevent meaningful interpretation. Repeated measurements of a longitudinal exposure (weight) and later outcome (glycated haemoglobin levels) were simulated to match three scenarios: one with no causal relationship between growth rate and glycated haemoglobin; two with a positive causal effect of growth rate on glycated haemoglobin. Two methods that condition on the outcome and one that did not were applied to the data in 1000 simulations. The interpretation of the two-step method matched the simulation in all causal scenarios, but that of the methods conditioning on the outcome did not. Methods that condition on the outcome do not accurately represent a causal relationship between a longitudinal pattern and outcome. Researchers considering longitudinal data should carefully determine if they wish to analyse longitudinal data as a series of discrete time points or by extracting pattern features.

6.2 Introduction

Lifecourse data comprise longitudinal data (repeated measurements) that span some or all of life. Analyses of lifecourse data are popular for informing preventative policies to improve population health and wellbeing (World Health Organisation 2015). For example,

temporal patterns of growth (recorded in repeated measures of weight) throughout childhood might be related to risk of type-2 diabetes by age 40 years to target preventative measures at those with certain 'high risk patterns'. To do this effectively, results from analyses must truly reflect a relationship between patterns of growth and diabetes. This may not be the case for some commonly used lifecourse methods.

Repeated measurements of a longitudinal exposure, such as weight throughout infancy, are usually correlated with each other, a phenomenon known as autocorrelation (Littell Ramon et al. 2000). Therefore, they do not satisfy the requirement for independence of observations needed for many common statistical analyses (Davidian and Giltinan 1995). Methods for capturing and analysing longitudinal exposures typically aim to describe how different patterns of the exposure (e.g. rate of adolescent weight growth) relate to the outcome. Alternatively, some methods aim to identify specific times or 'critical periods' during which the (causal) effect of the exposure is especially strong, or estimate the cumulative effect over multiple exposure times (A. Smith, Hardy, et al. 2016; A. Smith, Heron, et al. 2015).

Generalised methods (g-methods), such as marginal structural models, and methods that explicitly examine "lifecourse hypotheses" offer the most obvious solution to achieving these objectives, given their theoretical foundation within an explicit causal framework (A. Smith, Hardy, et al. 2016; A. Smith, Heron, et al. 2015). G-methods are however very rarely utilised in applied research, perhaps due to perceived complexity (Naimi et al. 2017). Simpler and more common methods are less likely to incorporate causal thinking; focusing instead on estimating non-causal associations that are consequently less useful for informing policies and interventions (Glass et al. 2013; Russo 2012).

With the aid of simulations, this paper explains how lack of causal thinking in analyses of longitudinal exposures in relation with later-life outcomes can lead to interpretational biases. Methods that can lead to these biases are compared to an alternative approach that avoids them. This alternative method, however, is not suitable for all situations; other methods, such as g-methods, would be necessary in the presence of time-varying confounding, which is not examined in this paper.

6.3 Methods

Data were simulated to represent the illustrative example of weight measured yearly from birth until age 2 years (the exposure) and diabetes diagnosed at age 40 years (the outcome) from percentage glycated haemoglobin (HbA1c) (Barker et al. 2005). This is analogous to routinely-collected health data or data from birth cohorts. Three illustrative scenarios with different causal structures were simulated matching the directed acyclic graphs in Fig 6.1. Each arrow specifies a direct causal relationship between variables. The absence of an arrow means there is no direct causal relationship, but there may still be a correlation. In Scenario A, birthweight causes HbA1c and there is no causal effect of growth rate on HbA1c. In Scenario B weight_1 directly causes HbA1c and growth rate indirectly causes HbA1c through weight_1. In Scenario C weight_2 directly causes HbA1c and growth rate indirectly causes HbA1c through weight_2. For ease of illustration, confounding (by e.g. genetics or in utero nutrition) was represented by a single unmeasured common cause of birthweight and growth (U).

1000 datasets comprising 1000 observations were simulated using R 3.4.3; exceeding the number required to achieve >99% accuracy for the parameters of interest (Burton et al. 2006). Simulation code is available in the S1 Appendix. Each directed acyclic graph was converted into a covariance matrix of the weight and HbA1c variables using the parameters in Table 6.1 and standardised path coefficients in Fig 6.1. Data were simulated with multivariate normal distributions. Linear growth was simulated for ease of interpretation. HbA1c was dichotomised into a binary variable at the National Institute for Health and Care Excellence threshold for diagnosing type-2 diabetes (HbA1c >6.5%) (National Institute for Health and Care Excellence 2016). The mean and standard deviation (SD) of simulated weight values, along with the correlation of each weight measure with HbA1c, were averaged across all simulations with 2.5^{th} and 97.5^{th} centiles depicting empirical 95% confidence intervals (CIs).

Variable	**Weight_0 (kg)**	**Weight_1 (kg)**	**Weight_2 (kg)**	**HbA1c (%)**
Mean	4	8	12	5.8
Standard deviation	2	2	2	1

Table 6.1: Parameters of latent variables and error terms used to simulate data in Section 6.3. The path diagram used to generate observed variables from these is shown in Fig 6.1.

Data were analysed using three methods: Z-score plots, multilevel models (outcome as

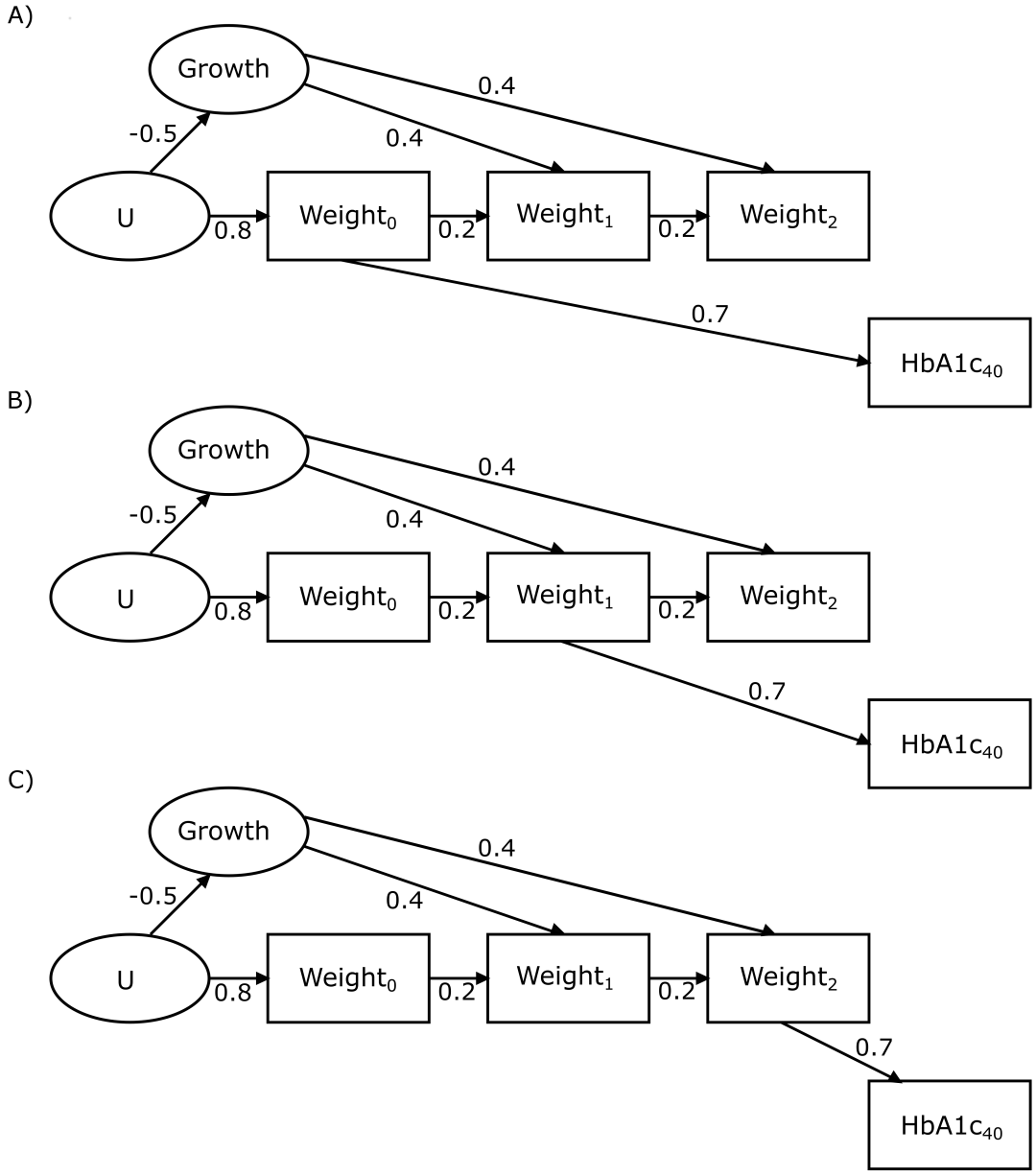


Figure 6.1: Directed acyclic graph showing the structure of causal relationships between variables in simulated scenarios A, B and C. Growth represents the growth rate of an individual and is not simulated or measured in the scenario. U is an unknown and unmeasured variable. The age in years at which known variables are measured is shown in subscript.

covariate), and multilevel models (two-step). Z-score plots are a simple, graphical approach that aims to identify exposure patterns that lead to an outcome (Barker et al. 2005; Tu, Tilling, et al. 2013). Weight at each age was standardised into z-scores using the time-specific sample mean and SD (National Institute for Health and Care Excellence 2016). The mean z-scores in those who did and did not develop diabetes were then plotted at each age and connected. Z-score plots are often viewed and interpreted as 'patterns' of weight that 'lead to' the outcome (Tu and Gilthorpe 2010). The plots presented show mean values from all simulations, with 2.5^{th} and 97.5^{th} centiles depicting empirical 95% CIs.

The multilevel model (outcome as covariate) analysis involved fitting multi-level models of weight over time, with covariates for age, diabetes status, and an age-diabetes interaction term, defined by the following equations (where i indexes observations and j individuals):

$$\begin{aligned} weight_{ij} &= \beta_{0j} + \beta_1 time_{ij} + \beta_2 diabetes_j + \beta_3 time_{ij} * diabetes_j + \epsilon_{ij} \\ \beta_0 &= \gamma_0 + u_{0j} \\ \beta_1 &= \gamma_1 + u_{1j} \\ var^{u_{0j}}_{u_{1j}} &= \begin{pmatrix} \theta^2_{u0} & \theta_{u01} \\ \theta_{u01} & \theta^2_{u1} \end{pmatrix} \end{aligned} \tag{6.1}$$

Intercept and age coefficients were free to vary randomly between individuals. Age was centred at one year (Blozis and Cho 2008; Mehta and West 2000). A first order autocorrelated error structure was specified to account for the effect of each weight measure on the subsequent. Multilevel models like this are typically interpreted from the coefficient of the interaction term; for example, a positive interaction between diabetes and time would be interpreted as meaning that an increased growth rate leads to diabetes. Coefficients for these interaction terms were recorded over the 1000 simulations to obtain a median and empirical 95% CIs.

The multilevel model (two-step) approach involved fitting two models, defined by the following equations (where i indexes observations and j individuals):

$$
\begin{aligned}
weight_{ij} &= \beta_{0j} + \beta_1 time_{ij} + \epsilon_{ij} \\
\beta_0 &= \gamma_0 + u_{0j} \\
\beta_1 &= \gamma_1 + u_{1j} \\
var \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} &= \begin{pmatrix} \theta^2_{u0} & \theta_{u01} \\ \theta_{u01} & \theta^2_{u1} \end{pmatrix}
\end{aligned}
\tag{6.2}
$$

$$
\begin{aligned}
log\left(\frac{P(diabetes)}{1 - P(diabetes)}\right) &= \beta_3\beta_1 + \epsilon \\
\epsilon &\sim N(0, \theta^2_\epsilon
\end{aligned}
\tag{6.3}
$$

The first (Eq 6.2) was a multilevel model of weight by age, with a first order autocorrelated error structure, to estimate growth as depicted in each directed acyclic graph in Fig 6.1. The intercept and age coefficients were permitted to vary randomly across individuals and age was centred. The individual-level age coefficients were recorded, representing individuals' growth rates. In the second step (Eq 6.3), a logistic regression model was fitted with diabetes as the outcome, the age coefficient (growth rate) as the exposure, and a birthweight covariate to condition for its confounding influence (Shrier and Platt 2008). The exponentiated model coefficients represent the change in odds of developing diabetes for each increase of 0.1kg/year (selected due to the small growth rate). Coefficients greater than one suggest that higher growth rates lead to diabetes. Coefficient point estimates for the growth rate exposure were recorded to obtain a median and empirical 95% CI from the 2.5^{th} and 97.5^{th} centiles over the 1000 simulations. All multilevel models were fitted using R package 'nlme' (Pinheiro et al. 2018).

Any errors from the multilevel model (outcome as covariate) and multilevel model (two-step), such as failure to converge, were recorded, and the estimates from these datasets were discarded.

6.4 Results

One dataset for each of scenarios A and B, and 20 datasets in scenario C were discarded due to models failing to converge. The mean and SD of weight at each time, averaged across all remaining simulations for each scenario are shown in Tables 6.2, 6.3 and 6.4, along with mean correlations of each weight measure with HbA1c. In scenario A, there was a large positive correlation at birth, decreasing to a small positive correlation at age 1, and a small negative correlation at age 2. In scenario B, there was a near-zero correlation at birth, increasing to a large positive correlation at age 1, and decreasing to a small positive correlation at age 2. In scenario C, there is a small negative correlation at birth, increasing to a small positive correlation at age 1, and a large positive correlation at age 2.

	Weight$_0$ (kg)		**Weight$_1$ (kg)**		**Weight$_2$ (kg)**		**HbA1c40 (%)**	
	Mean	**95%CI**	**Mean**	**95%CI**	**Mean**	**95%CI**	**Mean**	**95%CI**
Mean	4.003	3.879, 4.129	8.003	7.889, 8.133	11.997	11.863, 12.117	5.800	5.737, 5.862
SD	2.000	1.910, 2.093	1.998	1.910, 2.090	2.000	1.911, 2.092	0.999	0.955, 1.043
Correlation with HbA1c	0.699	0.664, 0.729	0.029	-0.033, 0.091	-0.105	-0.167, -0.044	1	

Table 6.2: Summary of simulated variables in Scenario A. 95%CI represents 95% empirical confidence intervals.

	Weight$_0$ (kg)		**Weight$_1$ (kg)**		**Weight$_2$ (kg)**		**HbA1c40 (%)**	
	Mean	**95%CI**	**Mean**	**95%CI**	**Mean**	**95%CI**	**Mean**	**95%CI**
Mean	4.001	3.870, 4.124	7.999	7.876, 8.127	11.997	11.885, 12.118	5.801	5.741, 5.859
SD	2.000	1.912, 2.088	2.000	1.917, 2.084	1.999	1.915, 2.09	1.000	0.956, 1.044
Correlation with HbA1c	0.027	-0.034, 0.088	0.699	0.666, 0.731	0.229	0.169, 0.283	1	

Table 6.3: Summary of simulated variables in Scenario B. 95%CI represents 95% empirical confidence intervals.

	Weight$_0$ (kg)		**Weight$_1$ (kg)**		**Weight$_2$ (kg)**		**HbA1c40 (%)**	
	Mean	**95%CI**	**Mean**	**95%CI**	**Mean**	**95%CI**	**Mean**	**95%CI**
Mean	3.997	3.873, 4.113	8.003	7.882, 8.122	12.001	11.875, 12.122	5.801	5.738, 5.861
SD	2.001	1.919, 2.087	2.000	1.911, 2.087	2.003	1.915, 2.101	1.001	0.959, 1.047
Correlation with HbA1c	-0.106	-0.166, -0.043	0.229	0.167, 0.288	0.700	0.666, 0.731	1	

Table 6.4: Summary of simulated variables in Scenario C. 95%CI represents 95% empirical confidence intervals.

The z-score plots for each scenario are shown in Fig 6.2. In scenario A, the diabetic group has a much higher weight at birth and the points on the graph are far apart, and far from the overall mean (zero). The points converge over time until they meet, cross and begin to diverge between age 1 and 2 years. By age 2, the diabetic group have a lower mean weight z-score than the non-diabetic group. In scenario B, the points are close to the overall mean at birth, diverging substantially at age 1, before converging back towards the mean at age 2; the diabetic group always has a higher mean weight z-score than the non-

diabetic group. In scenario C, the diabetic group starts with a slightly lower birthweight than the non-diabetic group, but the z-score increases over time, while the non-diabetic group decreases, leading to a large difference at age 2.

Results from the multilevel models (outcome as covariate) are in Table 6.5 and Fig 6.3, which show the model-fitted regression lines and true mean weight values for the diabetic and nondiabetic groups at each time point. The model values do not always fit well with the mean values (see especially scenario B in Fig 6.3B) because the models were constrained to linearity (because *growth* was simulated to be linear for simplicity), but the mean values in each outcome group change nonlinearly. In all scenarios, the coefficient of age is positive, confirming that weight increases from birth to age 2. For scenario A, the *negative* age-diabetes interaction term and shallower slope of increasing weight in the diabetes group (Fig 6.3A) suggests that those who developed diabetes grew slightly *slower* than those who did not develop diabetes. In scenario B, the small *positive* age-diabetes interaction term and slightly steeper slope (Fig 6.3B) suggests that those who developed diabetes grew slightly *faster* than those who did not develop diabetes. In scenario C, the large *positive* diabetes-age interaction term and steeper slope (Fig 6.3C) suggests that those who developed diabetes grew substantially faster than those who did not develop diabetes.

	Scenario A		**Scenario B**		**Scenario C**	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Diabetes	0.729	0.566, 0.900	1.070	0.899, 1.25	0.939	0.769, 1.111
Age	4.327	4.223, 4.430	3.914	3.807, 4.021	3.670	3.563, 3.767
Diabetes*Age	-1.372	-1.572, -1.166	0.340	0.133, 0.573	1.375	1.166, 1.573
Intercept	7.823	7.737, 7.911	7.740	7.655, 7.821	7.770	7.690, 7.862
Intercept variance	0.481	0.348, 0.601	0.503	0.127, 0.816	0.099	0.000, 0.255
Age Variance	0.680	0.547, 0.799	0.897	0.714, 1.08	0.567	0.000, 0.711
Residual Variance	1.770	1.710, 1.836	1.727	1.54, 1.849	1.840	1.778, 1.908
Constant-Age Covariance	0.973	0.894, 0.988	0.268	0.02, 0.845	-0.729	-0.957, 0.579
Autocorrelation parameter	0.120	0.069, 0.169	0.042	-0.118, 0.134	0.160	0.119, 0.196

Table 6.5: Average parameter estimates from multilevel models of weight (outcome as covariate).

Results from the multilevel models (two-step) are shown in Table 6.6. In scenario A, the odds ratio for growth rate was 1.000 (95% empirical CI: 0.943, 1.057), suggesting that the odds of diabetes were unaffected by growth rate. In scenario B, the odds ratio was 1.194 (95% empirical CI: 1.122, 1.316), suggesting that the odds of diabetes increased modestly with increasing growth rate. In scenario C, the odds ratio was 1.679 (95% empirical

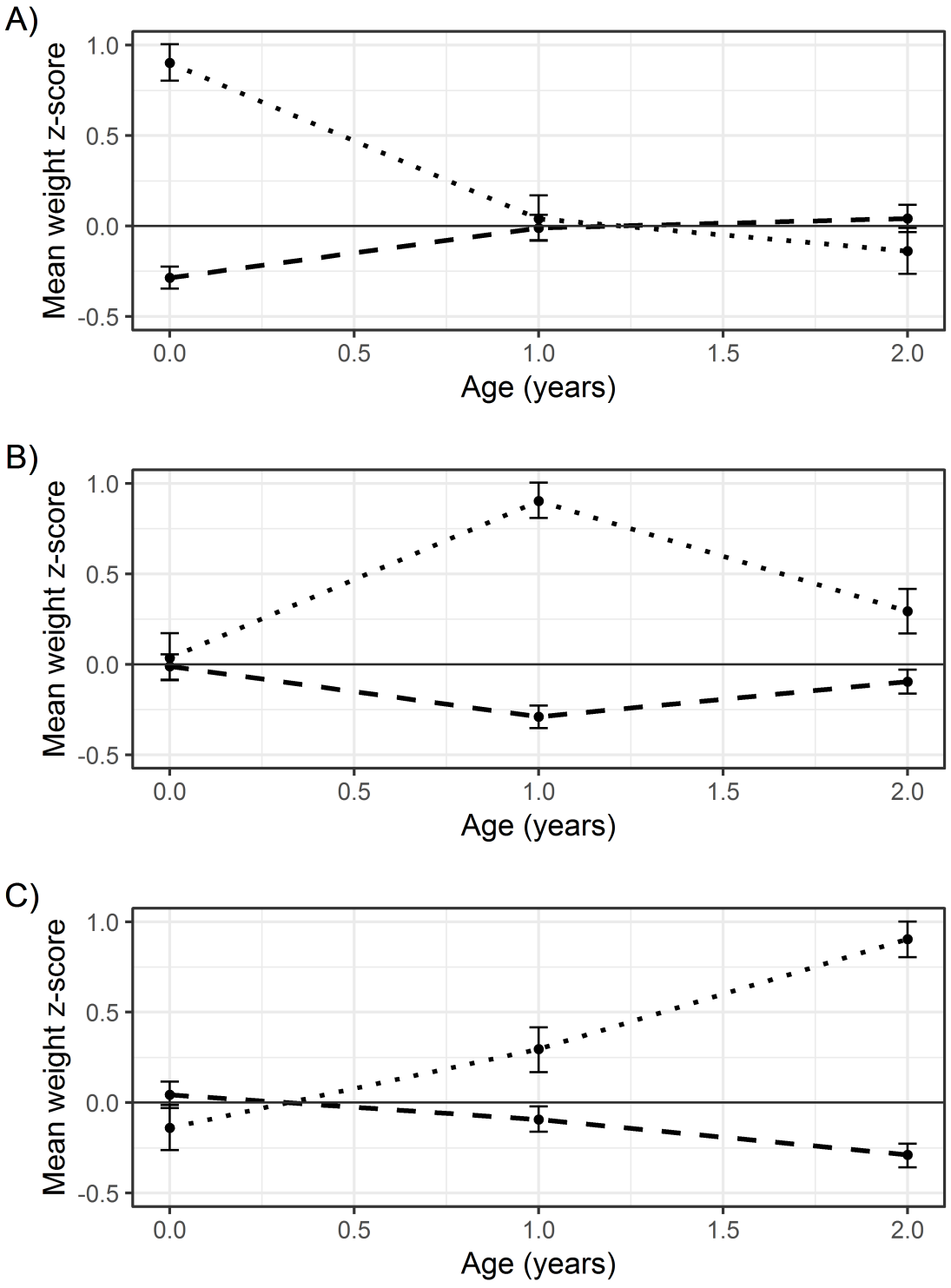


Figure 6.2: Z-score plots of weight from birth to age 2 years for scenarios A, B and C. Dotted lines show the group diagnosed with diabetes at age 40 and dashed those without a diagnosis. Error bars show empirical 95% confidence intervals.

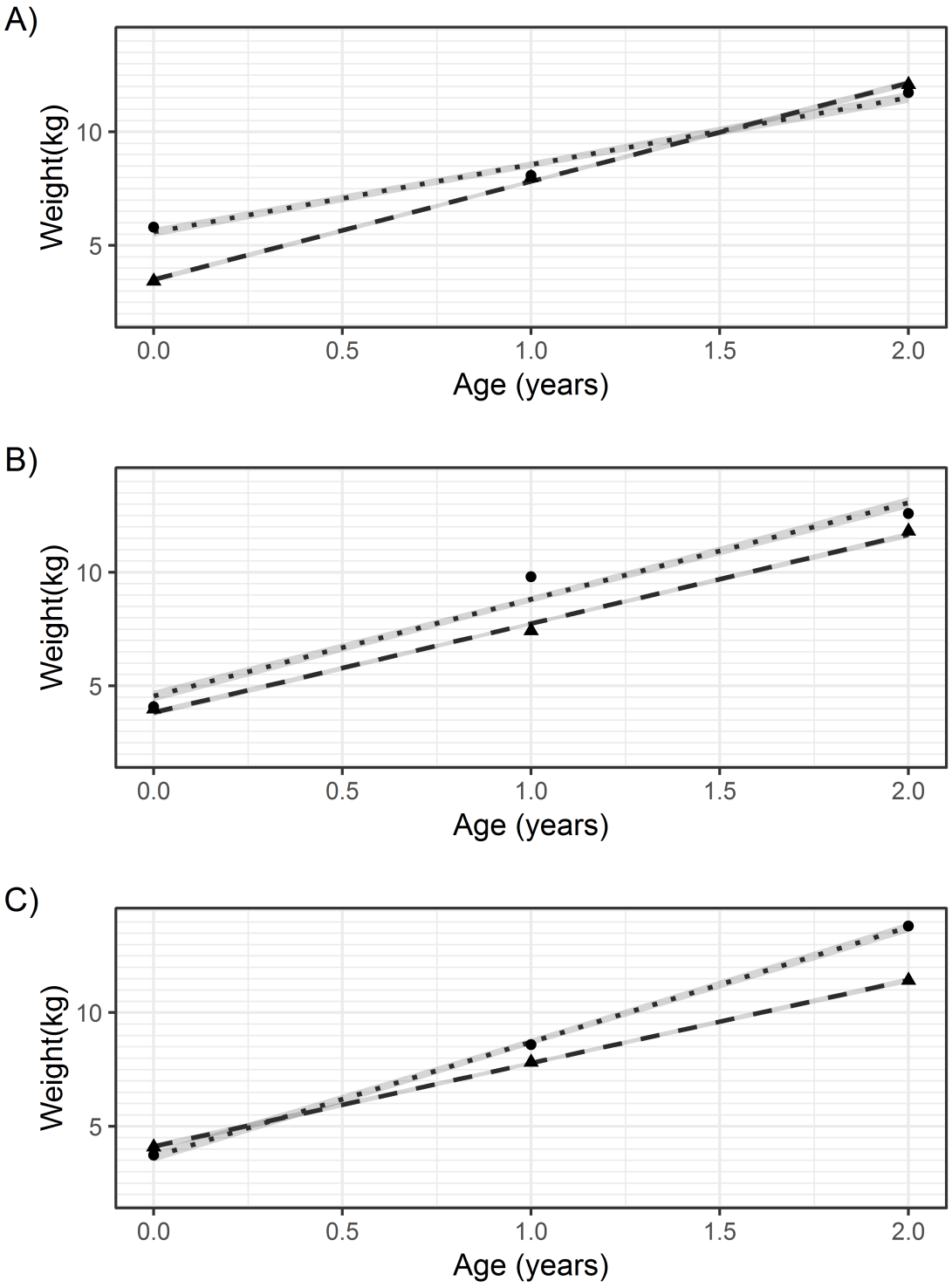


Figure 6.3: Fitted weight values from multilevel models (outcome as covariate) and average mean weight values for scenarios A, B and C. Dotted lines (fitted values) and circular points (average mean weight values) represent fitted values for the group with a diabetes diagnosis at age 40. Dashed lines (fitted values) and triangular points (average mean weight values) represent those without a diagnosis. The grey ribbon represents an empirical 95% confidence band around the fitted values.

CI: 1.477, 2.191), suggesting the odds of diabetes increased substantially with increasing growth rate.

	Scenario A		**Scenario B**		**Scenario C**	
Parameter	**Mean**	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
Growth rate	1.000	0.943, 1.057	1.194	1.122, 1.316	1.679	1.477, 2.191
Weight_0	2.000	2.060, 2.745	1.000	1.149, 1.429	2.000	1.339, 2.265
Constant	$6.030x10^{-03}$	$3.654x10^{-04}$, $9.185x10^{-02}$	$9.309x10^{-05}$	$1.659x10^{-06}$, $1.291x10^{-03}$	$2.221x10^{-11}$	$2.552x10^{-16}$, $7.591x10^{-09}$

Table 6.6: Average parameter estimates from the logistic regression model of diabetes status on weight growth rate. Growth rate was estimated using a multilevel model of weight over age (agnostic to the outcome, diabetes status)

6.5 Discussion

In Scenario A, we simulated no causal effect of growth rate on the risk of developing diabetes; Birthweight causes HbA1c, but any pattern of growth thereafter is irrelevant. Neither the z-score plot nor the multilevel model (with the outcome as a covariate) reflect this and would be erroneously 'interpreted' as showing that slower growth leads to diabetes. Conversely, the coefficient from the two-step multilevel model correctly implies no effect of growth rate on diabetes risk.

In scenario B, we simulated that weight_1 caused diabetes, which could be interpreted as growth causing HbA1c through weight_1. The z-score plot however suggests that faster growth up to age 1 *and* slower growth thereafter leads to diabetes. This does not reflect the causal relationship simulated, where higher growth only increased the risk of diabetes by increasing weight at age 1. Both the multilevel model (with the outcome as a covariate) and the two-step multilevel model reflect this more closely, suggesting that higher growth rates caused diabetes.

In scenario C, we simulated that weight_2 caused diabetes, which could again be interpreted as growth causing HbA1c through weight_2 (and indirectly through weight_1). Here, the results from all three methods correctly suggest that higher growth rate cause diabetes.

The z-score plots (and common interpretation thereof) only reflected the simulated truth in one of the three scenarios, revealing this is not a reliable approach for examining the causal effect of a longitudinal exposure on a distal outcome. This is because average weight z-scores at each time point are explicitly calculated and presented within groups of

the outcome. By inappropriately conditioning on the outcome in an attempt to examine 'average patterns' of weight associated with diabetes, the method actually examines cross-sectional associations between weight and the outcome at each time point. This problem remains even if only one group (e.g. those with diabetes) is considered. The value of each mean z-score (e.g. weight) has no obvious causal meaning; instead, it reflects the size of the cross-sectional correlation between the exposure and the outcome at each time point. Because the standardisation process fixes the scale, time points with the strongest cross-sectional correlations will always appear most different, and those with the weakest correlations will always appear most similar. For example, in Scenario A, there was a strong positive correlation between birthweight and diabetes due to the causal effect of birthweight, and weaker correlations at ages 1 and 2 years as the contribution of birthweight to weight decreased. This is reflected by the z-score plot in Fig 6.2; the mean weight z-scores values are farthest apart at birth and coverage over time. In scenarios B and C, the strongest correlations are at ages 1 and 2 years respectively; the corresponding z-scores plots are likewise farthest apart at these time points. The absolute value of each time point z-score should not therefore be joined or compared to the z-score values at other time points because the 'patterns' that appear have no causal meaning and do not represent individual growth *trajectories*.

Inappropriate conditioning on the outcome also affects multilevel models where the outcome is included as a model covariate. The consequences are not identical to the z-score plot because the scale has not been fixed by standardisation and the correlation pattern is assumed to follow a specific parametric shape. In our example, the linearity constraint introduced misfit between the modelled regression lines and the mean weight values (Fig 6.3). In Scenario B this meant the model failed to highlight that the largest cross-sectional correlation between weight and diabetes occurred at age 1 year, and this explains the difference in interpretation with the corresponding z-score plot. In scenarios A and C, models were similar enough to the average weight values to provide similar interpretations. Had we simulated nonlinear growth, however, the linearity constraint would likely have introduced further differences in interpretation compared with the z-score plot.

The multilevel model (two-step) approach is more robust than the other approaches because it does not involve conditioning on the outcome. Instead, exposure patterns are

modelled and only in the second step are these related to the outcome. This approach genuinely treats the exposure as a longitudinal variable and should therefore be strongly favoured over approaches that condition on the outcome whenever there is an interest in the causal interpretation of a *longitudinal* exposure pattern. This method is not, however, without limitations.

First, because the second step of the multilevel model (two step) approach treats the unobserved growth rate estimates as fully observed, it underestimates the standard errors (and confidence intervals), even when attempts are made to address this (Sayers et al. 2017). Alternative latent variable methods, like latent growth curve models, growth mixture models, and autoregressive latent trajectory models, which retain the latent, or unobserved, nature of the pattern features avoid this problem.

Second, two-step multilevel models and their constructed latent variable alternatives can still present some interpretational challenges from a causal inference perspective. By summarising the effect of multiple measurements that span a period into one or more average feature(s), such as growth rate, the causal contributions of each individual measurement occasion is lost, as too are any corresponding 'critical' period effects (Ben-Shlomo and Kuh 2002). This places such methods in contrast to g-methods, where the focus is explicitly on estimating the causal effect of the exposure as measured at each time point. Whether the underlying feature (e.g. growth) or the individual measure (e.g. weight at age 1) are the 'true' cause cannot be distinguished statistically because they are simply different ways of describing the same information. In scenarios B and C, for instance, both 'growth rate' and the individual measures of weight at ages 1 and 2 years respectively appear to cause diabetes. Changing either would therefore change the risk of diabetes, and neither can be described as more or less responsible. The choice of whether to analyse a longitudinal exposure as a series of discrete measures or a summary feature (e.g. growth rate) may therefore be down to philosophical and/or contextual preferences regarding the question(s) posed. That said, since many pattern features like 'growth rate' span several measurement intervals, they are susceptible to time-varying confounding by any variables that are simultaneously caused by earlier measures while causing later measures, i.e. so-called intermediate confounders. In such situations, there may be no alternative to g-methods, which are currently unique for their compatibility with intermediate confounding (Naimi

et al. 2017).

It is important to note that, for illustrative purposes, this paper presents a simplified scenario in which there are no competing events or loss to follow up, both of which would be present in reality. Any differential loss to follow up or occurrence of competing events would (further) bias the results from all three methods examined in this paper (Howe et al. 2013).

6.5.1 Recommendations

Methods that condition on the outcome are not appropriate for examining the causal relationship between patterns of a longitudinal exposure and a later outcome, as they only describe the cross-sectional correlations at each time point. The apparent 'patterns' that are observed have no causal interpretation and should not be interpreted as individual exposure trajectories that cause the outcome.

Alternative analytical strategies should seek to describe features of the exposure agnostic to the outcome, whether explicitly in two separate models or implicitly using latent variable methods. Researchers should however carefully consider whether pattern features or discrete measures are more appropriate, useful and/or interpretable ways to capture a specific 'exposure' in a specific context. If interested in the effect of exposures at specific 'critical' points in time then alternative methods are recommended (A. Smith, Hardy, et al. 2016; A. Smith, Heron, et al. 2015).

If a pattern feature is truly of interest, researchers should think very carefully about which pattern feature(s) are of interest before analysis. In the absence of a single, distinct and clearly identifiable causal feature it is tempting to consider summarising the 'average' of exposure 'trajectories' for individuals with different outcomes by conditioning on the outcome, but this risks highly misleading results. A longitudinal exposure – or pattern thereof – that spans a long period may be conflated with intermediate confounding and thus fail to describe the true causal process of interest. Features that occur at specific time periods that have a tangible real-world meaning may be best suited to the methods recommended, such as two-step multilevel models.

6.6 Conclusion

This paper explains how longitudinal data analyses that inappropriately condition on the outcome may lead to biased inferences about how exposure patterns affect later outcomes. Methods such as z-score plots and multilevel models with the outcome as a covariate do not create causally meaningful exposure 'patterns' and, as our simulations show, can be highly misleading.

In lifecourse research, or whenever interested in the causal relationship between a longitudinal exposure and later outcome, we recommend avoiding methods that inappropriately condition on the outcome in favour of methods that capture patterns *a priori*, although the potential influence of intermediate confounding should be carefully considered.

6.7 Supporting information

S1 Appendix. Code used to simulate and analyse data. (DOCX)

6.8 Author contributions

Conceptualization: Mark S. Gilthorpe.

Formal analysis: Sarah C. Gadd.

Investigation: Sarah C. Gadd.

Methodology: Sarah C. Gadd, Peter W. G. Tennant, Mark S. Gilthorpe.

Supervision: Peter W. G. Tennant, Alison J. Heppenstall, Mark S. Gilthorpe.

Writing – original draft: Sarah C. Gadd.

Writing – review & editing: Peter W. G. Tennant, Alison J. Heppenstall, Jan R. Boehnke, Mark S. Gilthorpe.

References

Barker, D., Osmond, C., Forsén, T., Kajantie, E., & Eriksson, J. (2005). Trajectories of growth among children who have coronary events as adults. *New England Journal of Medicine*, *353*(17), 1802–1809.

Ben-Shlomo, Y., & Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, *31*(2), 285–293.

Bhargava, S., Sachdev, H., Fall, C., Osmond, C., Lakshmy, R., Barker, D., Biswas, S., Ramji, S., Prabhakaran, D., & Reddy, K. (2004). Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood. *350*(9), 865–875.

Blozis, S., & Cho, Y. (2008). Coding and centering of time in latent curve models in the presence of interindividual time heterogeneity. *Structural Equation Modeling-a Multidisciplinary Journal*, *15*(3), 413–433.

Burton, A., Altman, D., Royston, P., & Holder, R. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, *25*(24), 4279–4292.

Davidian, M., & Giltinan, D. (1995). *Nonlinear models for repeated measurement data* (Vol. 62). Chapman & Hall.

Eriksson, J. [J.], Forsén, T., Tuomilehto, J., Osmond, C., & Barker, D. (2000). Fetal and childhood growth and hypertension in adult life. *Hypertension*, *36*(5), 790–794.

Eriksson, J. [J.G.], Forsen, T., Osmond, C., & Barker, D. (2003). Pathways of infant and childhood growth that lead to type 2 diabetes. *Diabetes Care*, *26*(11), 3006.

Forsén, T., Osmond, C., Eriksson, J., & Barker, D. (2004). Growth of girls who later develop coronary heart disease. *Heart*, *90*(1), 20.

Glass, T., Goodman, S., Hernán, M., & Samet, J. (2013). Causal inference in public health. *Annual review of public health*, *34*, 61–75.

Howe, L., Tilling, K., Matijasevich, A., Petherick, E., Santos, A., Fairley, L., Wright, J., Santos, I., Barros, A., Martin, R., Kramer, M., Bogdanovich, N., Matush, L., Barros, H., & Lawlor, D. (2013). Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research*, *25*(5), 1854–1874.

Littell Ramon, C., Pendergast, J., & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*, *19*(13), 1793–1819.

Mehta, P., & West, S. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*(1), 23–43.

Naimi, A., Cole, S., & Kennedy, E. (2017). An introduction to g methods. *International Journal of Epidemiology*, *46*(2), 756–762.

National Institute for Health and Care Excellence. (2016). Clinical knowledge summary: Diabetes - type 2. https://cks.nice.org.uk/diabetes-type-2#!diagnosissub

Oka, R., Shibata, K., Sakurai, M., Kometani, M., Yamagishi, M., Yoshimura, K., & Yoneda, T. (2017). Trajectories of postload plasma glucose in the development of type 2 diabetes in Japanese adults. *J Diabetes Res*, *2017*, 5307523.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2018). nlme: Linear and nonlinear mixed effects models.

Russo, F. (2012). Public health policy, evidence, and causation: Lessons from the studies on obesity. *Medicine, Health Care and Philosophy*, *15*(2), 141–151.

Sabia, S., Dugravot, A., Dartigues, J., Abell, J., Elbaz, A., Kivimaki, M., & Singh-Manoux, A. (2017). Physical activity, cognitive decline, and risk of dementia: 28 year follow-up of Whitehall II cohort study. *Bmj-British Medical Journal*, *357*.

Sayers, A., Heron, J., Smith, A., Macdonald-Wallis, C., Gilthorpe, M., Steele, F., & Tilling, K. (2017). Joint modelling compared with two stage methods for analysing longitudinal data and prospective outcomes: A simulation study of childhood growth and BP. *Stat Methods Med Res*, *26*(1), 437–452.

Shrier, I., & Platt, R. (2008). Reducing bias through directed acyclic graphs. *BMC Medical Research Methodology*, *8*(1), 70.

Smith, A., Hardy, R., Heron, J., Joinson, C., Lawlor, D., Macdonald-Wallis, C., & Tilling, K. (2016). A structured approach to hypotheses involving continuous exposures over the life course. *International Journal of Epidemiology*, *45*(4), 1271–1279.

Smith, A., Heron, J., Mishra, G., Gilthorpe, M., Ben-Shlomo, Y., & Tilling, K. (2015). Model selection of the effect of binary exposures over the life course. *Epidemiology*, *26*(5), 719–26.

Tu, Y.-K., & Gilthorpe, M. (2010). *Statistical thinking in epidemiology*. Chapman & Hall/CRC.

Tu, Y.-K., Tilling, K., Sterne, J., & Gilthorpe, M. (2013). A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. *International Journal of Epidemiology*, *42*(5), 1327–1339.

World Health Organisation. (2015). *The case for investing in public health* (Report).

Chapter 7

Simplifying the interpretation of continuous-time models for spatio-temporal networks

Citation: Gadd SC, Comber A, Gilthorpe MS, Suchak K, Heppenstall AJ (2021) Simplifying the interpretation of continuous time models for spatio-temporal networks. Journal of Geographical systems.

This chapter illustrates the application of the model specified in Chapter 4 and addresses the third aim of the thesis: to develop (and communicate) an analytical procedure to apply continuous-time models to a spatio-temporal network and simplify their interpretation. Here, the model is applied to both simulated and real data describing a rail network. Whole-network functions of average speed over time are produced and the relationship between origin station centrality and the maximum delay passengers experience during the day for each origin-destination pair in the network is examined. The simulated data in this chapter are generated using the simulation tool from Chapter 4.

7.1 Abstract

Autoregressive and moving average models for temporally dynamic networks treat time as a series of discrete steps which assumes even intervals between data measurements and can introduce bias if this assumption is not met. Using real and simulated data from the

London Underground network, this paper illustrates the use of continuous-time multilevel models to capture temporal trajectories of edge properties without the need for simultaneous measurements, along with two methods for producing interpretable summaries of model results. These included extracting 'features' of temporal patterns (e.g. maxima, time of maxima) which have utility in understanding the network properties of each connection and summarising whole-network properties as a continuous function of time which allows estimation of network properties at any time without temporal aggregation of non-simultaneous measurements.

Results for temporal pattern features in the response variable were captured with reasonable accuracy. Variation in the temporal pattern features for the exposure variable was underestimated by the models. The models showed some lack of precision. Both model summaries provided clear 'real-world' interpretations and could be applied to data from a range of spatio-temporal network structures (e.g. rivers, social networks). These models should be tested more extensively in a range of scenarios, with potential improvements such as random effects in the exposure variable dimension.

Keywords: Spatio-temporal data, hierarchical modelling, networks, multilevel modelling.

7.2 Introduction

Network or graph structures, like that shown in Figure 7.1, are commonly used to represent connections or relationships between objects, places or individuals. They are typically cast such that the objects are represented by vertices (or nodes) and connected by edges (or arcs). They have been shown to be useful for examining and simulating transport systems (Angeloudis and Fisk 2006; Austwick et al. 2013), river networks (Erős et al. 2011), social connections (Scott 1988) and many other systems. Properties can be assigned to edges and vertices to indicate, for example, the amount of traffic flowing through a transport route or the geographical location of vertices.

Many networks change dynamically over time and understanding the temporal patterns of edge and vertex properties could provide useful insight to researchers and planners. In public transport, such information could be used to help target activities and resources. For example, identifying the optimum times to close different parts of the network for

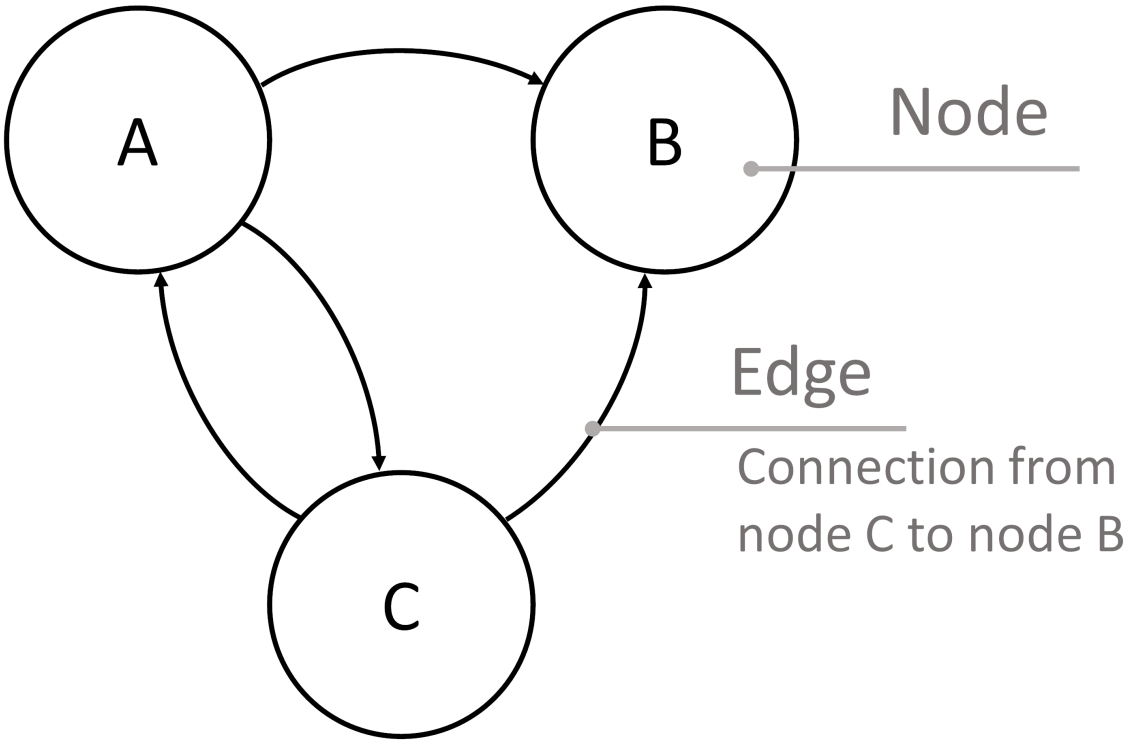


Figure 7.1: Example of a simple, directed network graph with three vertices (A, B and C) and four directed connections (A to B, A to C, C to A, C to B). Vertices could represent, for example, railway stations, and arrows the journeys of passengers between origin and destination stations.

maintenance while minimising disruption to passengers; or finding the most useful times and locations to add new services to reduce overcrowding within specific cost constraints. Several structures have been developed for such analyses including time-aggregated networks which represent temporally dynamic networks as a series of static graphs capturing temporal snapshots or windows (Blonder and Dornhaus 2011). These approaches summarise the temporal dynamics, but there can be aggregation problems when discretising events, for example when measurements of properties for different edges are not simultaneous. In contrast, time-ordered networks show connections between vertices as a continuous function of time, but generally focus on the binary presence or absence of a connection, rather than continuous network property values like traffic flows (Blonder and Dornhaus 2011). A range of machine learning methods have been successfully applied in this area, for example for traffic flow prediction (Ke et al. 2017). Although machine learning methods are often easier to implement than statistical methods and impose fewer restrictions (for example, on error distributions), this paper specifically examines statistical methods as they more easily allow us to use prior knowledge of a process to inform modelling, which

can be important when making inferences (Comber and Wulder 2019).

Models based on Space-Time Autoregressive Integrated Moving Average (STARIMA) have been developed for predicting traffic conditions in road networks (T. Cheng et al. 2011; T. Cheng et al. 2014; Pfeifer and Deutsch 1980). STARIMA models capture spatial and temporal autocorrelation using moving average and autoregressive terms. Spatial relationships are specified using a weight matrix that indicates the influence of other locations on any given location (Pfeifer and Deutsch 1980) which can be based on distances or other spatial relationships like network structures (Ermagun and Levinson 2018). Temporal autoregressive and moving average processes divide time into discrete temporal units (often reflecting the sampling intervals in the data). Measurements at set numbers of temporal units before a given observation are included as covariates to inform analysis. N-dimensional (N) STARIMA and localised (L) STARIMA were developed specifically for use in road networks where the nature of spatial autocorrelation within the network may vary over time, for example reflecting changes in the traffic flow through the network (T. Cheng et al. 2011; T. Cheng et al. 2014). NSTARIMA uses an alternative spatial weight matrix that accounts for the distance between two locations and the speed of traffic at both locations, with the result that the matrix changes over time in response to changes in traffic speeds (T. Cheng et al. 2011). LSTARIMA additionally accounts for spatial heterogeneity by allowing spatial weight matrices to differ across locations at a given time (T. Cheng et al. 2014).

The spatial weight matrices in NSTARIMA and LSTARIMA approaches are typically based on road network distances (as well as speed) but their formulation could potentially be modified to reflect different types of network structures. Research in stream networks has employed modelling strategies that do not frequently appear in the analysis of transport networks. For example, geostatistical models with covariance structures reflecting both Euclidian and hydrologic distances (distances along the river network, accounting for flow direction) have been constructed to predict fish populations (Peterson and Ver Hoef 2010; Ver Hoef and Peterson 2010). These models, based on moving average processes, allow error terms to covary with each other according to these metrics and could be extended to additionally account for a temporal dimension (Peterson and Ver Hoef 2010; Ver Hoef and Peterson 2010). Other research has employed multilevel models (also known

as mixed models or mixed effects models) with continuous-time autoregressive error structures to account for temporal autocorrelation to predict sediment concentrations at sites in stream networks (Leigh et al. 2019). Multilevel models are hierarchical models and are capable of modelling temporal patterns of variables for many individuals or objects (Goldstein 2011). Their flexible error structure allows them to capture complex data generation processes, such as those with spatial, temporal or network autocorrelation. These models treat time as a continuous variable, so can be used to interpolate or predict (with a level of uncertainty) at any given time point, rather than only at set intervals.

While STARIMA-based methods are useful for examining spatio-temporal data, treating time continuously is an advantage over treating it as a series of discrete lags, as the latter may not capture the temporal dynamics of the process of interest, in this case the continuous evolution of transport networks (Comber and Wulder 2019). Continuous-time models allow measurement intervals to vary between and within locations, in contrast to STARIMA and similar models that use discrete time lags which assume that measurement intervals do not vary across time or with location (i.e. between edges in the model). Data frequently do have varying measurement intervals, for example, when data collection is not an automated process (e.g. manual traffic counts), when measurements from several systems (e.g. different traffic camera networks, or traffic cameras and manual traffic counts) are combined, or when measurements are made at the timing of an irregular event (e.g. 'tap in' and 'tap out' train systems). In such cases, autoregressive models, like STARIMA, require aggregation of data into temporal units. This aggregation and information loss can bias the results of the models, meaning they may fail to capture the temporal dynamics of the process being examined (de Haan-Rietdijk et al. 2017; Huang et al. 2010).

If data are systematically sampled at regular time intervals, no aggregation is required. In this case, however, the size of temporal units in a discrete-time model is often determined by the sampling interval in the data (J. Freeman 1989). Parameters relating to periodicity and autocorrelation have been shown to change in response to the size of the temporal unit chosen (Hawawini 1978; Huang et al. 2010; W. Wei 1981). This means they may reflect the sampling intervals more than the processes underlying the data, which makes them less adept at investigating these processes (Comber and Wulder 2019). Both systematic

sampling and temporal aggregation have been shown to introduce bias to results from autoregressive models (J. Freeman 1989; Hawawini 1978; Rossana and Seater 1995; W. Wei 1981; Weiss 1984). High frequency variation is masked by aggregation due to averaging or summing multiple data measurements in one temporal unit (J. Freeman 1989). It has also been shown that low-frequency variation (on a much longer time scale than the temporal-unit) is masked when autoregressive integrated moving average (ARIMA) models are fitted to aggregated data (Rossana and Seater 1995). If time is treated continuously, rather than a series of discrete units, autocorrelation or periodic processes can be specified in the model independent of the sampling intervals, allowing them to reflect the processes underlying the data, and the masking of variation due to temporal aggregation is avoided.

As suggested by Leigh et al. (2019), multilevel models with spatial and temporal autocorrelation terms could be applied to network data to model continuous temporal patterns of edge or vertex properties, such as traffic flow patterns for different routes in a road network. The results from such models, however, can be difficult to interpret. This is because coefficient estimates from complex models aiming to capture complicated temporal patterns (such as those including polynomial terms of spline functions) do not have clear 'real-world' interpretations (Stimson et al. 1978). In addition, many visual representations of temporal patterns within complex networks can be overcrowded and hence present similar difficulties in extracting anything but very general patterns (Aigner et al. 2007). Important information from the models is thus hard to identify and communicate.

This study proposes methods for analysing continuous temporal patterns of edge properties in networks that can summarise and present the results in a more interpretable way, with appropriate estimates of uncertainty. Multilevel models are applied to capture continuous temporal patterns of each edge property and two methods for simplifying the interpretation of multilevel models are then illustrated. The first involves extracting information about temporal pattern features for each edge in the network. These are specific parts of temporal patterns that are of end-user interest. The second method involves constructing a model of the continuous temporal patterns of whole-network properties using representations of flow patterns for each edge. As both methods are based on interpolation, processes to calculate 95% credible intervals representing uncertainty are also described. The calculation and extraction of continuous temporal network patterns and properties, along with uncertainty

parameters are, to the best of our knowledge, new to this field and potentially have a wide range of applications in transport geography. These methods are applied to real and simulated case studies. The former is based on 'tap in' and 'tap out' data from the London Underground System. The latter includes similar data simulated for a subsection of the London Underground System to provide a comparison of method performance without the difficulties associated with real data, such as regions of sparse measurements. The edge property modelled is the time taken for London Underground users to complete their journeys. This example extracts the time and extent of the longest delay for each origin-destination pair and calculates a single function representing the average speed of journeys across the whole network.

7.3 Methods

7.3.1 Data

Oyster card data sample

An example analysis was carried out on a dataset of Oyster Card journeys available from Transport for London Open Data. Oyster Cards are used as a payment system for the Transport for London Network, which includes bus and rail journeys throughout the city. There are three interlinked rail systems in London: the London Underground, Docklands Light Railway and London Overground, with the former covering the largest area and number of stations. The rail systems are divided into zones to denote fare differences. These range from zone 1 being the most central, to zone 9 being the farthest from the city centre.

The dataset includes origin-destination information for a 5% random sample of all journeys taken for one week in November 2009. The origin-destination data were combined with station locations, also available from Transport for London Open Data, data detailing the structure of the rail network, compiled from the September 2009 London Underground Map, and data detailing the zones of the Underground stations, compiled from the December 2009 London Underground Map (the September 2009 map did not include zone information). For this analysis, journeys taking place on Wednesday with an origin and destination station on the London Underground in zones 1 or 2 were considered. Journeys

taking place between 14:00 and 21:00 were considered to investigate network properties associated with the afternoon rush hour. Restrictions on the zones were to limit the computational intensity of this illustrative analysis. For each journey, the time taken to travel from origin to destination was recorded – this was chosen over journey counts as an outcome for the analysis because journey counts in a sample of the original data will be very different to those occurring in reality, but the records of time taken to travel will not be changed.

Simulated data

A simulated dataset was used in addition to the real dataset. This was included to provide an example free of the difficulties associated with real data, such as time periods with sparse measurements. Comparing this to results from the real data may give some indication of why and how method performance changes in real circumstances. In addition, while a single simulation cannot fully assess the accuracy and precision of results, a scenario with a known 'truth' does give some indication of them. A smaller network was used in the simulation to aid visualisation of the results for illustrative purposes. Twelve stations from North West Central London were selected to form the simulated data. This area was chosen to include large variation in the degree of stations, based on the London Underground infrastructure, as this was included as a fixed effect in the simulation. Data representing the time taken for journeys made at different times during a 24-hour period were simulated for connections between each origin and destination pair in the network (132 edges total). Each edge had 50 simulated journeys – this is a much larger number than in the real data and was intended to provide an 'ideal' scenario with many measurements recorded throughout the range of the data.

Each edge had a known function representing its temporal pattern of flow. The patterns were based on well-known daily commuting patterns in which peaks of public transport use occur in the morning and evening. The largest delays were set to correspond to these times. Stations with a larger number of connections in this subsection of the London Underground Network had a larger difference between their shortest and longest journey times, on average. The data had a hierarchical structure, allowing random variation in flow along an edge according to its origin and destination station. Journey times recorded for each edge were allowed to covary according to the distance between their origin and

destination stations.

7.3.2 Data analysis

Both real and simulated data were analysed using the same approach. The first aim of the analyses was to identify for each connection recorded in the network the time of day at which the journey from origin to destination takes the longest (time of maximum delay). These features were chosen for this example as they are simple to extract and have an unambiguous interpretation, but similar processes could be used to extract a wide range of features. In the real data, this aimed to estimate the time of maximum delay during the afternoon, and whether this aligned with the evening peak in transport use. This might be considered the time at which there is most congestion, or the most delay for the journey. The difference between the expected time to complete a journey at this time of maximum delay and the minimum expected journey length in the observation period were calculated to represent the size of maximum delay. The second aim of the analysis was to identify the time of day at which journey speed is the slowest, on average, across the network. The information from both these aims could be used, for example, to plan the location and timing of crowd control measures in tube stations, to ease congestion, or the timing of rail services to reduce delays.

Analyses were carried out using R version 3.6.3 (R Core Team 2020a) and OpenBUGS (Lunn et al. 2009). This work was undertaken on ARC3, part of the High-Performance Computing facilities at the University of Leeds, UK. The code for analysis of real data and simulations are available as supplementary material.

Modelling temporal flow patterns

Multilevel models were chosen to model temporal patterns of flow because their hierarchical structure accounts for the similarity between origin-destination pairs sharing stations. The model was fitted using a Markov chain Monte Carlo (MCMC) estimation procedure with Gibbs sampling algorithm (S. Geman and D. Geman 1984), since MCMC procedures are typically more effective than maximum likelihood methods at estimating models with complex correlation structures (Browne and Draper 2006). This is a Bayesian estimation procedure which, instead of aiming to estimate a single set of 'true' parameter values, produces a series of samples of model parameters from a joint posterior distribution that

represents how likely different parameter values are believed to be. This distribution is based on the data provided, model specification and prior distributions representing beliefs about parameters prior to data collection (Heck and S. Thomas 2015).

A cross-classified multilevel structure (Goldstein 1994) grouped observations by both their origin and destination stations (Bürkner 2017). Error terms within origin-destination pairs followed a temporal autoregressive structure. The degree to which an error term was affected by its predecessor depended on the time difference between measurements. Time was represented on the minute scale and offset by one minute to allow for simultaneous journeys. To capture spatial autocorrelation, random effects related to origin and destination stations were divided into two parts – a aspatial component and a component with an intrinsic conditional autoregressive (CAR) prior. The CAR prior allows random effects (random intercepts and slopes) for different stations to covary based on a weight matrix specifying their spatial relations (Besag, York, et al. 1991). In this case, weights were set to zero for distances greater than or equal to two kilometres. For distances less than this, the weight was set to 2000 metres minus distance (in metres). Weights were scaled to have a standard deviation of one. An inverse distance weight matrix was considered but this led to difficulties with model convergence, leading to the choice of the linear matrix. A cut-off was chosen as spatial correlation between stations is likely to be driven in part by their serving passengers arriving from a similar area. Most passengers arrive at Underground stations on foot, therefore the cut-off of two kilometres for spatial correlation was chosen to represent a reasonably short walk. Distances were calculated using the haversine formula (Xiao 2016).

To capture more complex temporal patterns, the model was based on a cubic b-spline basis function. B-splines were chosen as a basis function due to their flexibility, but also their mathematical convenience: unlike penalised splines which require specific optimisation algorithms, the value of each b-spline basis for each observation can be calculated before model estimation and incorporated as a covariate (Perperoglou et al. 2019). Functions to calculate b-spline derivatives in the 'splines2' R package also allowed easy calculation of pattern features (such as the time of longest journey) when using b-splines, as detailed later (W. Wang and Yan 2020). The data included were only from a short time span and did not have any periodic patterns. Therefore, splines that are intended for capturing periodicity,

such as Fourier basis systems, were not used in this analysis (Ramsay and Silverman 1997). Some epidemiological papers that fit spline models to extract pattern features use linear splines, instead of cubic splines (Howe et al. 2013). This simplifies a smooth continuous temporal pattern to a series of joined straight lines (a 'broken stick') which may be useful in some circumstances. However, the extraction of the pattern features chosen in this example relied upon using calculus to obtain the rate of change of flow (at a maximum or minimum point, the rate of change is zero). A simplified 'broken stick' model does not capture temporal variation in rate of change, as the splines are straight between each knot point, and was therefore unsuitable for this application.

In the absence of penalisation, the number of knots and their placement for b-spline bases can substantially affect model results (Perperoglou et al. 2019). Automatic knot placement procedures are available, but these are mostly aimed at single-level regression models (Dung and Tjahjowidodo 2017; Yeh et al. 2020). Results from these may not be applicable to multilevel models, as they consider all the data as one group, rather than in individual clusters. Knot placement may be chosen visually based on the location of very changeable areas in the data – placing more knots in these locations will capture these areas with more accuracy (Holmes and Mallick 2003). However, the changeable areas for these data varied considerably between different origin-destination flows – placing many knots in one time period may mean that origin-destination flows with their longest journey times outside of this period will be captured with less accuracy than within. Instead of visual or automatic knot placement, forty models with different knot placements were tested. The knot placements considered were: a) even placement of knots along the time-axis and b) placement of knots at quantiles of the time distribution in the data (to account for uneven distribution of measurements throughout time). Each of these was tested with 1 to 20 internal knots. The model with the lowest Deviance Information Criterion was chosen, as this criterion balances measures of both model fit and parsimony (Spiegelhalter et al. 2002). This resulted in the use of three knots at quantiles for the real data and 19 evenly spaced knots for the simulation.

Time was centred on the grand mean, as this can improve the precision of parameter estimates (Paccagnella 2006). The model structure is described in Equation 7.1, in which $from$ and to index the origin and destination stations, respectively, for edges in the net-

work; t refers to the time at which each measurement was made; $basis_n(t)$ refers to the value of basis function n at time t; $Nbasis$ refers to the number of basis functions specified; $e_{i,from,to}$ represents the error term for the i^{th} observation for a particular edge; u represents aspatial random effects, indexed by origin or destination station (and basis function for u_n); v represents spatial random effects following an intrinsic CAR distribution, indexed by origin and destination station (and basis function for v_n); and a represents the coefficient for the autoregressive error structure estimated by the model.

$$\begin{aligned} journeytime_{i,from,to} &= \beta_{0,from,to} + \sum_{n=1}^{Nbasis} \beta_{n,from,to} basis_n(t_i) + e_{i,from,to} \\ \beta_{0,from,to} &= \beta_0 + u_{0,from} + u_{0,to} + v_{0,from} + v_{0,to} \\ \beta_{n,from,to} &= \beta_n + u_{n,from} + u_{n,to} + v_{n,from} + v_{n,to} \\ e_{i,from,to} &= a e_{i-1,from,to}/(t_i - t_{i-1} + 1) \end{aligned} \tag{7.1}$$

As previously mentioned, an intrinsic CAR prior distribution was specified for spatial random effects in the intercept (β_0) and basis (β_n) coefficients. The CAR normal distribution was used which constrains random effects to follow a Gaussian distribution. The intrinsic CAR distribution sets a random effect of zero for any stations with no neighbours (in this case, no other stations within two kilometres) (A. Thomas et al. 2014). For this reason, an additional random effect with no spatial structure, following a normal distribution with mean zero, was specified for the intercept and basis coefficients. Otherwise these stations would have no random effects, and just the fixed coefficients β_0 and β_n. The prior distribution for the precision (inverse variance) of the spatial and aspatial random effects were also set as a gamma distribution with shape and rate both equal to two. When converted to a prior distribution for the standard deviation of these random effects, this gives the most probable standard deviation as 1.41 with a low probability (<0.01) of a standard deviation greater than five. As a positive autoregressive structure was expected for these data, the prior distribution for the autocorrelation parameter a followed a truncated gamma distribution, allowing only positive values between zero and one. The initial value was set to 0.5, in this middle of this range.

Initial values for all other parameters and priors for fixed coefficients were taken from

posterior distributions of a model run without the temporal autoregressive structure in the error terms included. This model ran with uninformative priors for fixed coefficients (Besag and Kooperberg 1995) and output a normally distributed posterior. CAR distributions can be sensitive to the initial values set for precision parameters (A. Thomas et al. 2014), so a sensitivity analysis was carried out for the simulated data in which different initial values were set for the first (non-temporally autoregressive) model. The effect of these changes on results was minimal and can be seen in the supplementary material. The first model was run with four chains of 4000 iterations with 3000 burn-in. The second was run with four chains of 20000 iterations with 19000 burn-in, leaving 4000 posterior samples to describe posterior distributions.

Extracting pattern features

To carry out the set aims, two pattern features were extracted: the time of maximum delay and the maximum delay. Similar methods could be used to capture alternative features. For example, the length of periods during which journey time is extended beyond a certain amount, or the time at which journey times are increasing the fastest.

Maxima and minima can be defined as the point at which the slope of a function is equal to zero. These can be identified by calculating the derivative of the model function, which represents the slope, and finding the times at which it is equal to zero (Soetaert and Herman 2009). The maximum point with the highest value of the un-differentiated function was selected as the time of maximum delay and maximum journey length. If there were multiple points where the derivative function was equal to zero, the minimum point with the lowest value was selected as the minimum journey length. Otherwise, an optimisation function may be used to estimate the minimum journey length, and the difference between this and the maximum journey length recorded as the maximum delay (R Core Team 2020a). This would be the case if the function only changed direction once during the observation period.

The mean value of sampled parameters was used to specify a model function for each origin-destination pair. This was used to extract point estimates for the pattern features. To estimate uncertainty in pattern features, maxima and minima were also extracted (for each origin-destination pair) for all 4000 sampled parameter sets produced by the MCMC

estimation procedure. This results in corresponding sets of maxima for each edge for these posterior samples. Quantiles of the time and journey length of these maxima were used to represent 95% credible intervals for time of maximum delay and extent of delay.

The relationship between the extracted pattern features and the vertex closeness (for real data) or vertex degree (for simulated data) of their origin and destination stations was visualised (L. Freeman 1978). This aimed to investigate the relationship between the centrality of a station in the London Underground infrastructure network and the maximum delay for journeys beginning or ending at it.

To compare simulated and real values for pattern features, the differences between simulated and modelled values were calculated. To assess agreement between modelled and simulated values, the Bland-Altman method was used to calculate the mean bias and limits of agreement, along with 95% confidence intervals, from these differences (Bland and Altman 1999). In addition, the Spearman's rank correlation coefficient was calculated to identify if the ranking of origin-destination pairs by simulated and modelled pattern features was similar.

Estimating continuous temporal network properties

As well as using calculus to extract pattern features from model equations, the equations can also be converted into continuous temporal functions describing properties of the network. In this example, it might be useful to look at the average speed of passengers moving through the network over time to identify times of day when journeys are much slower than usual. To calculate the average speed, s, from the time taken to complete a number of journeys over distance d, at a single point in time, T, we used Equation 7.2, where n indexes each origin-destination connection and t the time.

$$s = \frac{\sum_{i=1}^{n}(d_n/T_{n,t})}{n} \tag{7.2}$$

When evaluating this equation, journey times for journey i at timepoint t could be substituted with temporal functions of journey time for each origin-destination pair. This would generate a continuous temporal function representing the average speed in the network.

Simultaneous 95% credible intervals were generated at five time points along the function by calculating quantiles of the values of all the functions specified by posterior parameter samples at these time points. The quantiles were adjusted for multiple testing using the Bonferroni correction.

Although the temporal function for average speed calculated in this illustration has a very specific meaning, this principle concept of temporally dynamic network properties could be extended to other summary information, if it is usefully interpretable. For example, if functions representing the amount of passenger flow were calculated, a range of more conventional network properties could be calculated: weighted edge density (Horvath 2011) could be calculated to represent the passenger load on a network, relative to the number of possible connections, over time; vertex centrality could be calculated to identify when certain stations act as transport hubs (by having high flow connections to many other stations) (Opsahl, Agneessens, et al. 2010); or weighted clustering coefficients could be used to identify the extent to which the transport network is divided into communities of stations that people tend *not* to travel between, and if this varies at different times (Opsahl and Panzarasa 2009).

7.4 Results

Figure 7.2 shows a map of the London Underground network. The subsection used in simulations is shown in Figure 7.3.

7.4.1 Real data

There were 27085 journeys from the sample data that took place in the study period chosen. Table 7.1 summarises network properties for both the London Underground infrastructure (that is shown in Figure 7.2) and the network formed by connections between origin-destination pairs recorded during the observation period in the real data. For the London Underground Infrastructure, edge density is low, suggesting that the network is sparsely connected. Closeness, the centrality measure used in Table 7.1, represents how close (in terms of geodesic distance) each vertex is to others in the network (L. Freeman 1978). The whole-network centrality indicates whether centrality is concentrated in just a small number of vertices, or reasonably evenly distributed throughout the network. In

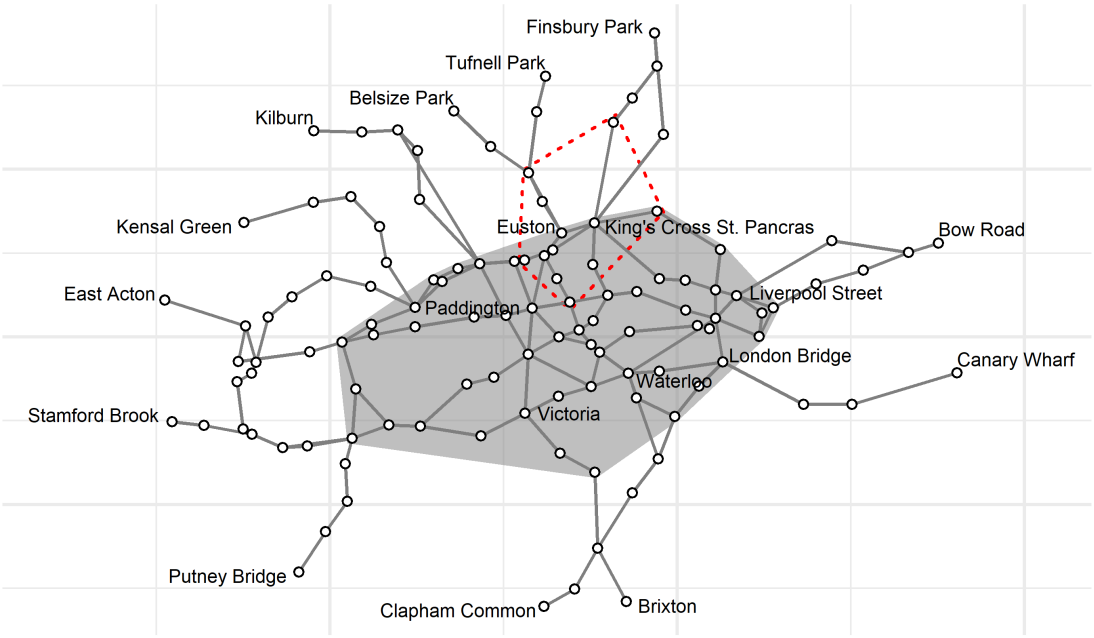


Figure 7.2: Map of zones 1 and 2 of the London Underground network. Zone 1 is shaded in grey. The area used to generate simulated data is outlined with a red dashed line.

this example, the value is low, suggesting most vertices have a similar centrality in this network, without large variation. This corresponds to the degree distribution in Figure 7.4, where most of the vertices in the network have the same degree (two).

Property	**Value**
London Underground Infrastructure	
Number of vertices	118
Network centrality (closeness)	0.14
Edge density	0.02
Mean shortest path length	7.11
Journey connections recorded	
Number of vertices	118
Network centrality (closeness)	0.47
Edge density	0.48

Table 7.1: Table showing the network properties of the section of London Underground infrastructure in zones 1 and 2 and the (directed, unweighted) network formed of connections made between these stations by passengers during the study period.

For the journey connections, the network centrality and edge density are higher. This suggests a more (but not highly) saturated network, with more of the possible connections between stations made; and more variation in closeness between vertices, with a small

Figure 7.3: Map of the section of the London Underground network used to generate simulated data.

number of stations being closely connected to many other stations, but most being more distantly connected. The degree distribution for this network, which is directed, is divided into in and out degrees, for connections ending at and starting at each station, respectively. This is shown in Figure 7.5. The range of these values is much larger than that for the infrastructure network, which is expected – people are likely to use the London Underground network to make journeys beyond immediately connected stations. Both degree distributions show a slight right skew, with slightly more low than high values; however, the average in and out degree are both high values. This suggests that for each station, passengers travel from/to a diverse range of origins/destinations.

Summary of pattern features

Table 7.2 summarises the pattern features extracted from the real data. The results here suggest that the distribution of times of maximum delay is centred around 17:47. This is close to what was expected from these data – many commuters will be travelling home from work at this time in the evening rush hour, so delays to travel might be expected due to congestion. The size of the delays themselves are not very large, with a mean maximum

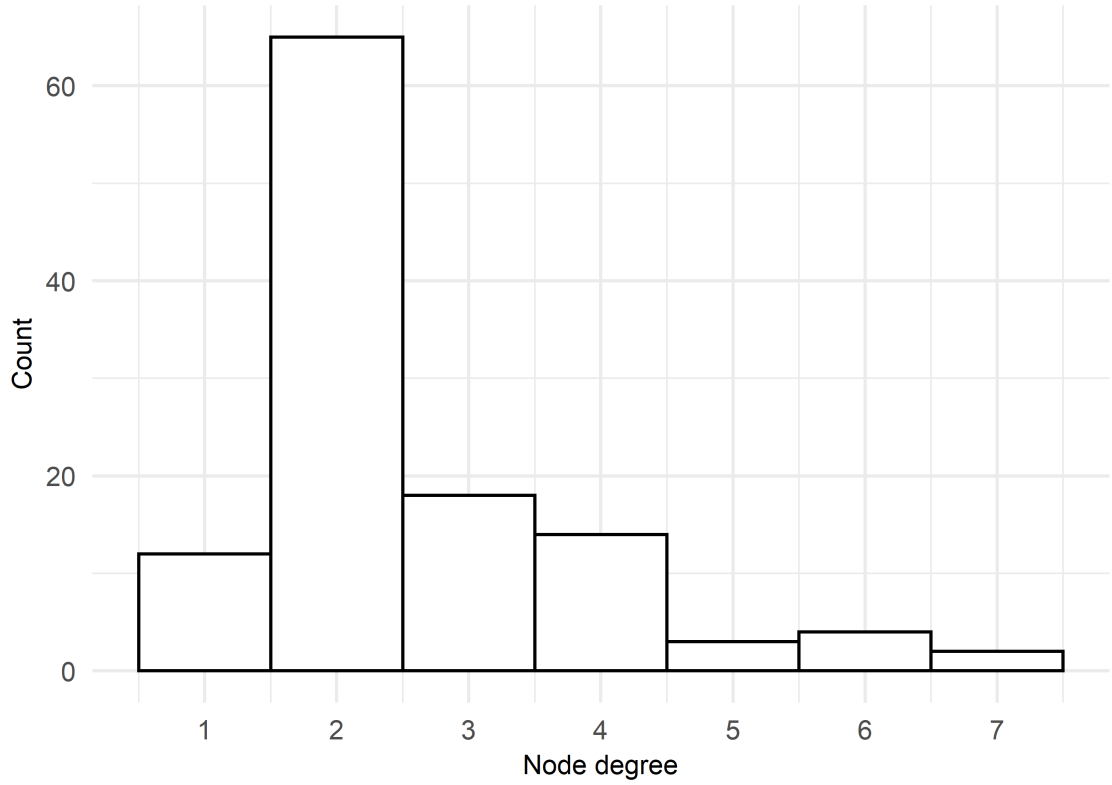


Figure 7.4: Vertex degree distribution for the London Underground infrastructure network.

delay value of just under seven minutes. The credible intervals for both these results are very wide on average. This suggests that the estimates here are not very precise, and there is a lot of uncertainty in this model.

Relationship with vertex centrality

Figure 7.6 shows the relationship of maximum delay with vertex centrality (based on the Underground network infrastructure) of origin and destination stations. There appears to be a negative relationship between centrality and the maximum delay value, particularly for lower values of centrality. This suggests that journeys from stations with fewer close connections to other stations, for example, those further from the city centre and serving fewer different lines, are likely to experience longer delays to their journeys at the time of maximum delay.

Continuous function of average speed

Figure 7.7 shows the graph of average speed throughout the network during the measurement period, along with 95% CIs. The credible intervals for this function are extremely

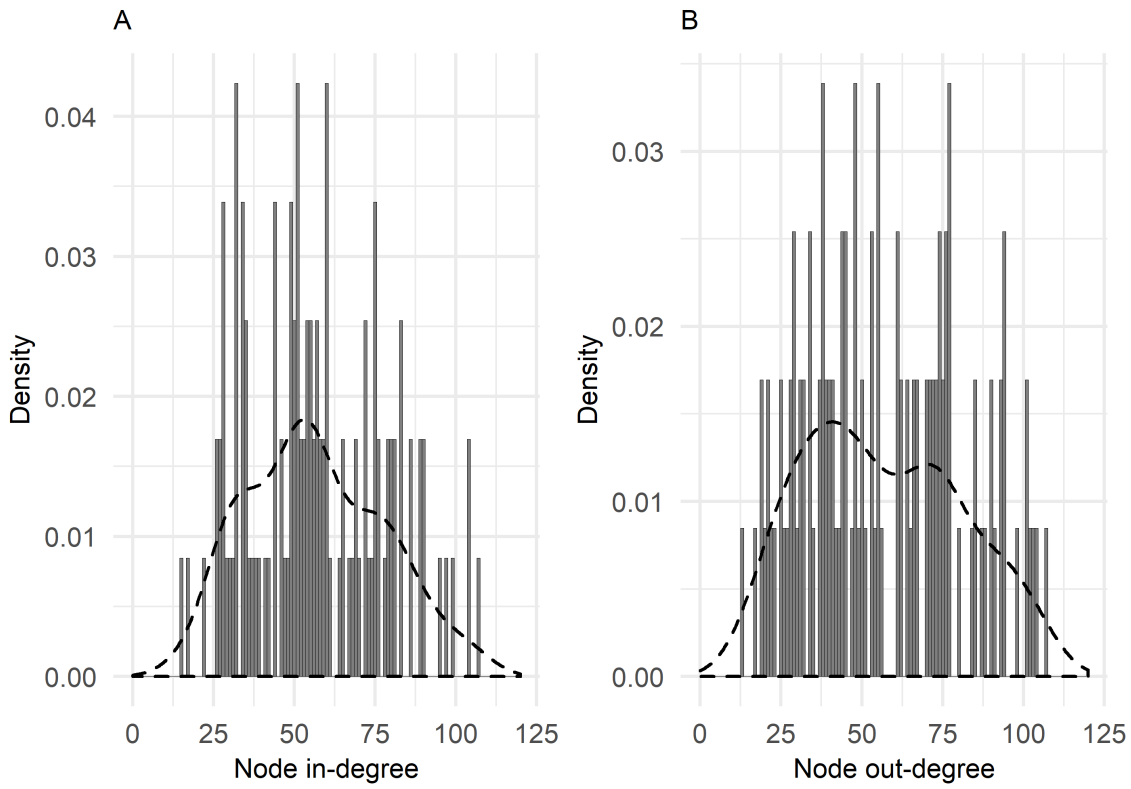


Figure 7.5: Vertex degree distribution for the network formed of connections made between London Underground stations in the Oyster card data. Panel A shows in-degree and panel B shows out-degree.

wide at the end of the time period. Throughout the time period, average journey speeds range from approximately 12.5 and 19 kilometres per hour, with the slowest speeds between 17:30 and 18:30. This timing corresponds to the mean time of maximum delay identified in Table 7.2.

Name	**Mean**	**SD**
Time of maximum delay (hours)		
Modelled	17:47	1.588
95% CI width	4.542	1.473
Maximum delay (minutes)		
Modelled	6.790	3.999
95% CI width	11.947	4.465

Table 7.2: Table showing the mean and standard deviation of modelled time of maximum delay and maximum delay and their 95% credible intervals (CIs) for sample Oyster card data.

7.4.2 Simulated data

Table 7.3 summarises network properties for the London Underground infrastructure used in this simulation. The edge density for this network is small, suggesting that the graph

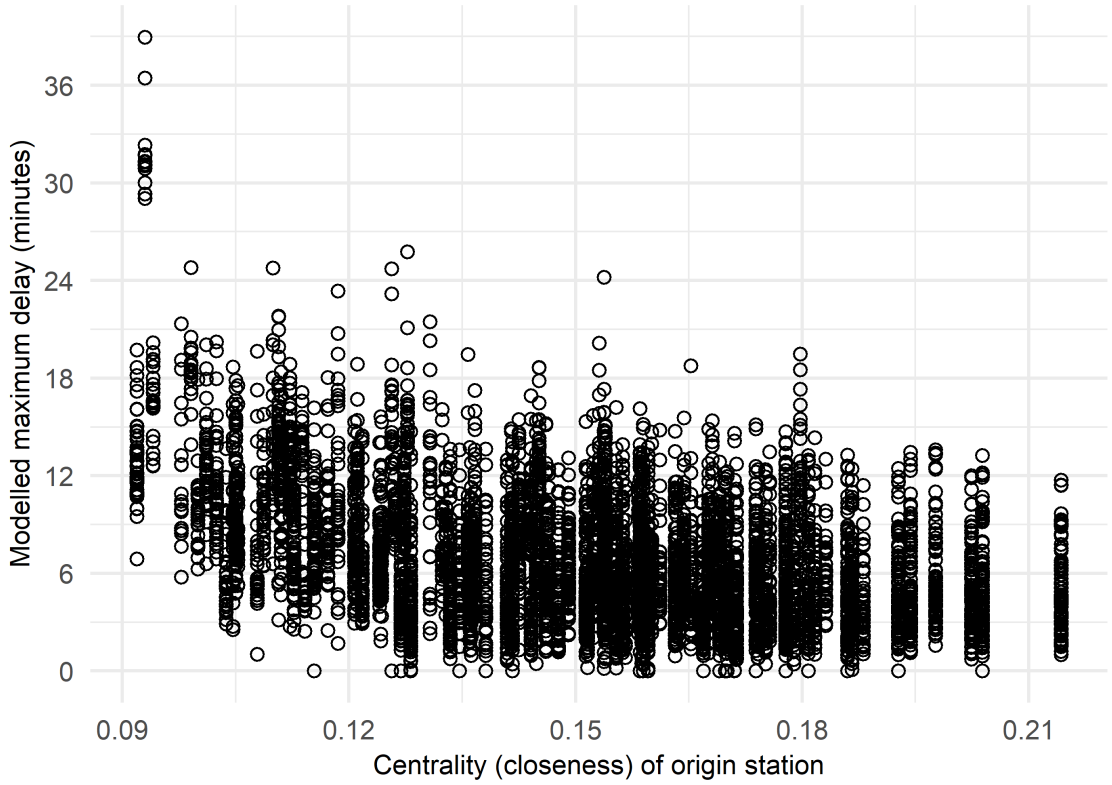


Figure 7.6: Relationship between origin station centrality and maximum delay values for origin-destination pairs in the Oyster card data sample.

is sparsely connected. Whole-network centrality is also relatively low, suggesting most vertices have a similar centrality in this network, without large variation. This is in line with the degree distribution in Figure 7.8 which shows that all stations in this network have a degree between one and five.

Property	**Value**
Number of vertices	12
Network centrality (closeness)	0.37
Edge density	0.18
Mean shortest path length	2.71

Table 7.3: Table showing the network properties of the section of London Underground infrastructure used to generate simulated data.

In this simulation, connections were generated between all possible pairs of origin and destination stations, so the graph representing these connections is saturated.

Summary of pattern features

Table 7.4 summarises simulated pattern features and those extracted from models of simulated data. For both time of maximum delay and maximum delay, the models recovered

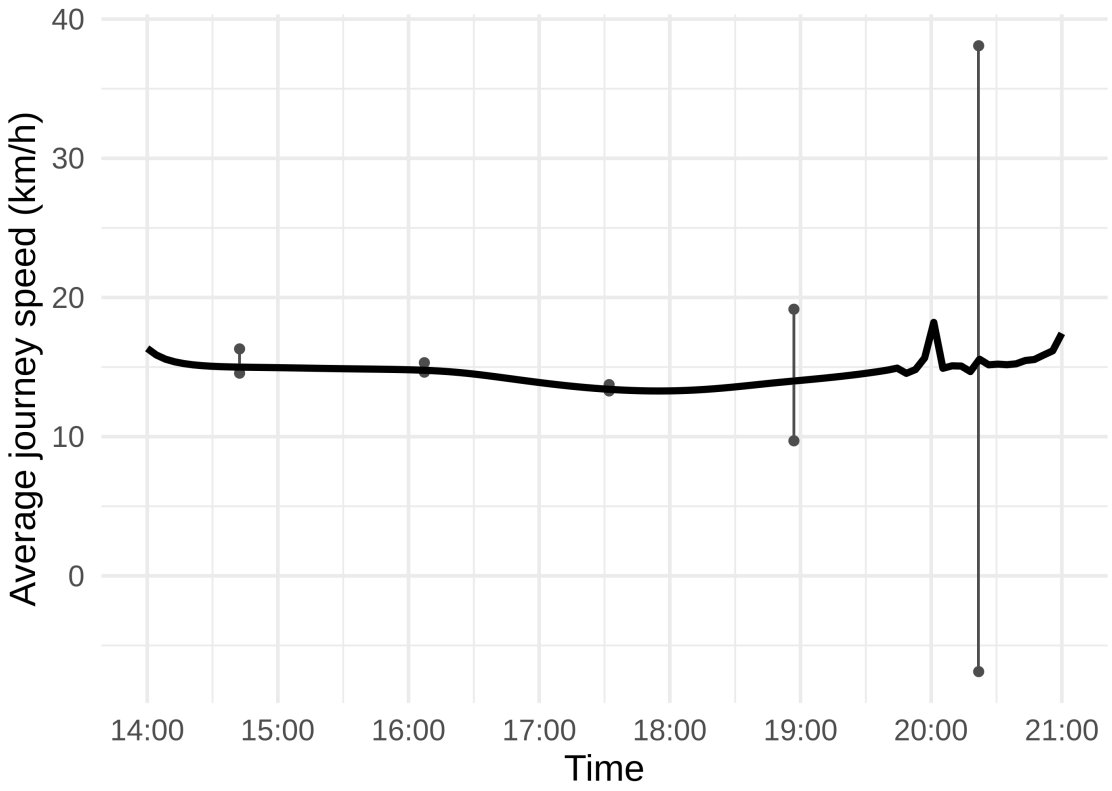


Figure 7.7: Continuous temporal function of average journey speed across the whole network with 95% credible intervals.

mean values very close to those simulated, with a difference of approximately four minutes in time of maximum delay and under one second in maximum delay.

For both features, the standard deviation in pattern feature values was smaller in the modelled than simulated values, indicating the full variation in these values between individuals was not being captured by the models. The difference between modelled and simulated standard deviations is much larger for the time of maximum delay (22.02 minutes versus 1.3 minutes for maximum delay).

The precision of estimates for maximum delay is also much greater than that for time of maximum delay. The mean credible interval width of maximum delay is just over six minutes, with a standard deviation of under 30 seconds; however, the mean credible interval for time of maximum delay has a width of approximately two hours eight minutes, with larger variation between individuals (standard deviation of one hour 46 minutes).

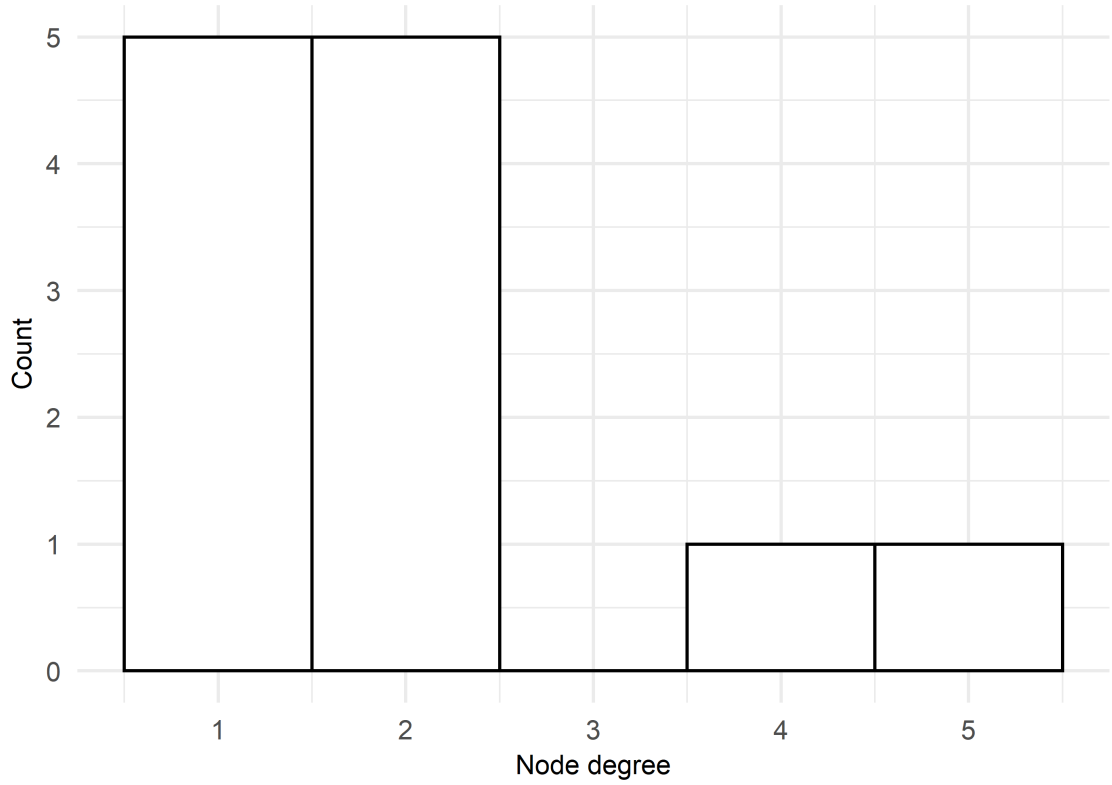


Figure 7.8: Vertex degree distribution for the section of the London Underground infrastructure network used to generate simulated data.

Relationship with simulated values

Figures 7.9 and 7.10 the relationship between simulated pattern features and the difference between modelled and simulated values along with Bland-Altman estimates of mean bias and 95% limits of agreement. These plots indicate whether the simulated value of maximum delay or time of maximum delay is related to the amount which the model over- or under-estimates its value. For both pattern features, there is a negative correlation, indicating that lower values are overestimated and higher underestimated. This relationship appears to be much stronger and linear for time of maximum delay – this is likely to be the cause of the underestimation of standard deviation for this feature. For maximum delay, the simulated values are split into four groups – this is a result of the fixed effects of origin vertex degree on this feature. The relationship here is not clearly linear, with the largest delay values being estimated more accurately than the second largest group.

The mean bias in estimates of maximum delay was 0.372 minutes (95% confidence interval -0.019, 0.764) with upper and lower 95% limits of agreement 4.862 minutes (95% confidence interval 4.156, 5.497) and -4.081 minutes (95% confidence interval -4.752, -3.411). This

Name	Mean	SD
Time of maximum delay (hours)		
Simulated	15:31	1.063
Modelled	15:35	0.696
Modelled - Simulated	0.067	1.177
95% CI width	2.134	1.767
Maximum delay (minutes)		
Simulated	18.282	11.216
Modelled	18.682	9.911
Modelled - Simulated	0.340	2.270
95% CI width	6.289	0.447

Table 7.4: Table showing the mean and standard deviation of simulated and modelled time of maximum delay and maximum delay. The differences between simulated and modelled values are also summarised, along with the width of the 95% credible intervals (CIs) associated with estimates.

suggests that there is a small positive bias in the modelled estimates of maximum delay (less than 24 seconds). The limits of agreement suggest that 95% of differences between simulated and modelled values would fall between -4 minutes 5 seconds and +4 minutes 52 seconds.

The mean bias in estimates of time of maximum delay was 0.065 hours (95% confidence interval -0.138, 0.268) This suggests that on average, the time of maximum delay is estimated 3.9 minutes late by the models. The upper and lower 95% limits of agreement were 2.375 (95% confidence interval 2.027-2.723) and -2.244 (95% confidence interval -2.592, -1.897). This suggests that 95% of differences between simulated and modelled values would fall between -2 hours 15 minutes and +2 hours 23 minutes.

The Spearman's rank correlation coefficient for modelled and simulated values for maximum delay was 0.872 (95% CI 0.706 to 0.947). This is a relatively high correlation coefficient showing a high, but not perfect, level of agreement between the ranking of different journeys by their modelled and simulated journey times. For the time of maximum delay, the degree of agreement is much lower, with a weak Spearman's rank correlation coefficient of 0.168 (95% CI -0.284 to 0.559). This suggests that modelled estimates of time of maximum delay do not correspond well to the simulated values.

Relationship with vertex degree

Figure 7.11 shows the relationship of simulated and modelled maximum delay with vertex degree. The overall pattern of the relationship is similar between the two graphs, which

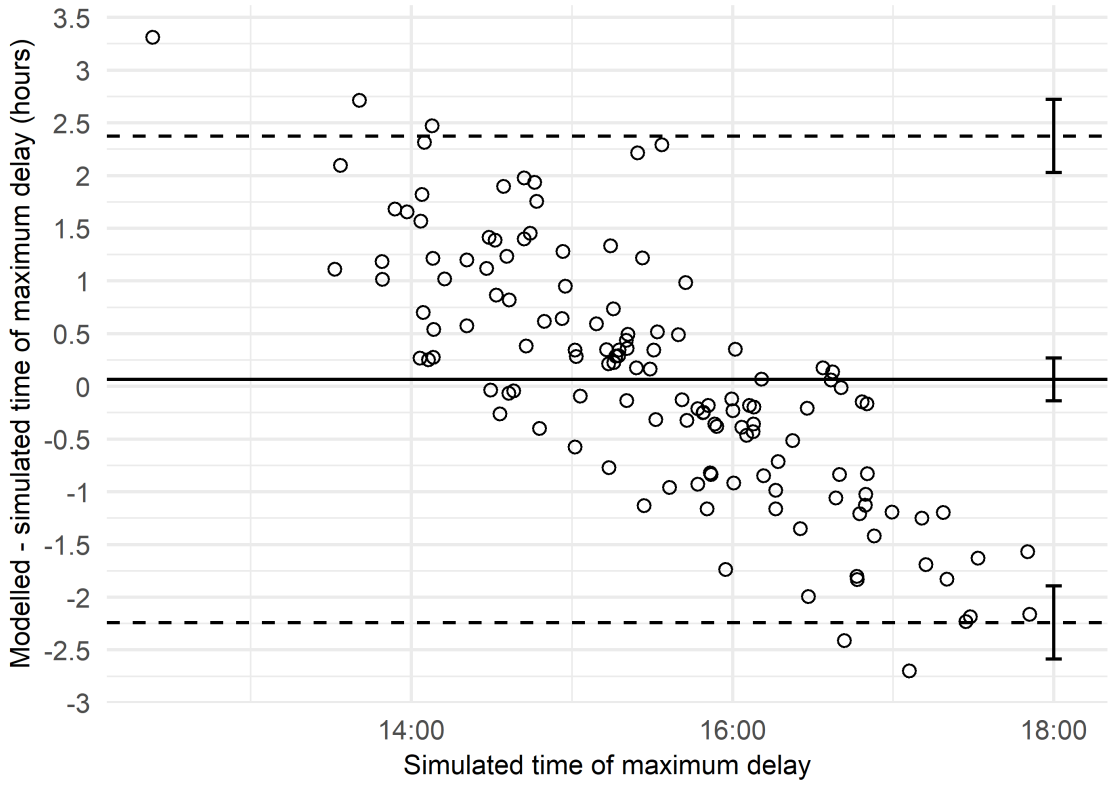


Figure 7.9: Relationship between over or underestimation of time of maximum delay by models and the simulated time of maximum delay. Horizontal lines show mean bias (solid) and 95% limits of agreement (dashed) with 95% confidence intervals.

suggests that the models capture the relationship between origin station degree and maximum delay accurately. There are some small differences in the distributions of simulated and modelled maximum delay values for each degree group, particularly for the lowest degree stations, where the distribution of modelled maximum delay has a higher median value and larger spread than the simulated values.

Continuous function of average speed

Figure 7.12 shows temporal functions of average speed throughout the network, along with 95% CIs, including both the simulated function and that extracted from models. The credible interval widths for this function are relatively narrow across most of the range of the data, with wider credible intervals at each end of the data, and before 5:00. The shape of the average speed function follows the simulated function reasonably well, but there are some discrepancies between the two. In most of the range of the model the speed is underestimated by a small amount, with the simulated function sitting outside the credible interval. The two areas with the largest bias are before 5:00, where the model

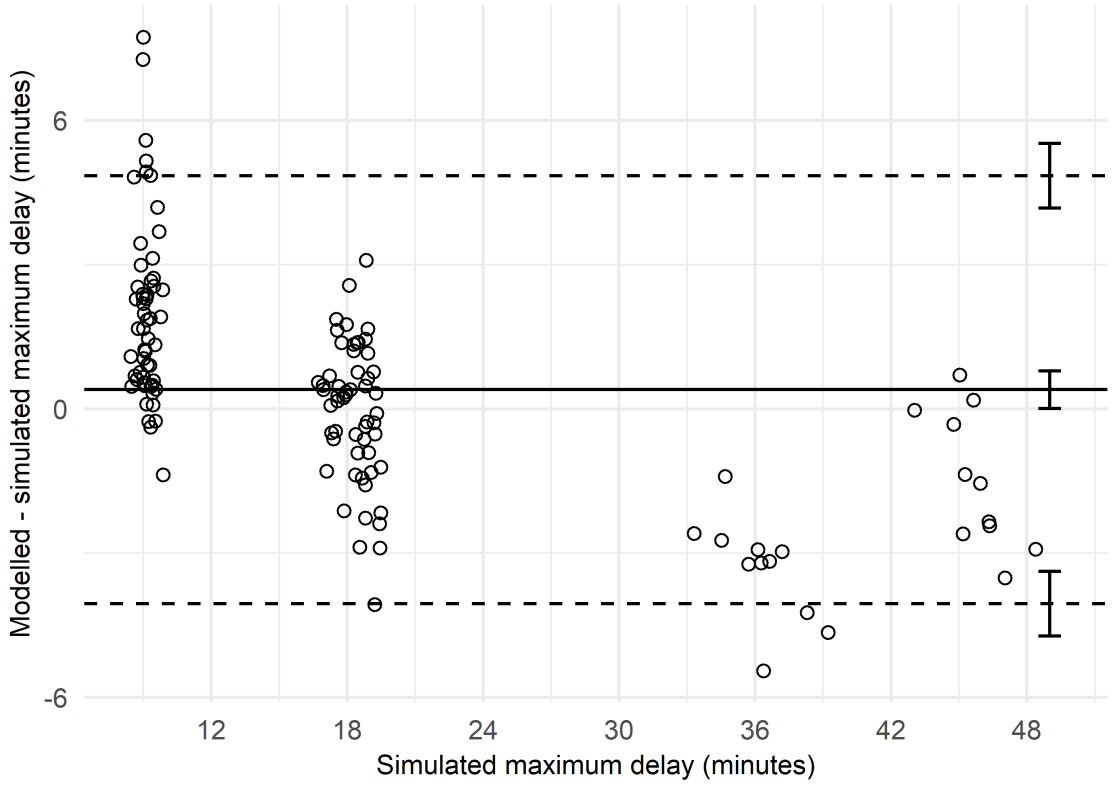


Figure 7.10: Relationship between over or underestimation of maximum delay by models and the simulated maximum delay. Horizontal lines show mean bias (solid) and 95% limits of agreement (dashed) with 95% confidence intervals.

underestimates average speed substantially and fails to capture the peak in the simulated average speed function, and after 21:00, where the underestimation of average speed is more pronounced than in the rest of the function. B-splines can be unstable at the limits of their range, which likely explains the large credible intervals at the limits of the graph and may be a factor contributing to the biased estimates near these areas (Perperoglou et al. 2019).

7.5 Discussion

This paper illustrates a method to examine spatio-temporal variation in network properties in an easily interpretable way. The properties were modelled using continuous-time models, in this case multilevel models, including both spatial and temporal autocorrelation.

Coefficients from models like those in this example can be difficult to interpret, so the results were simplified in two ways: through extracting pattern features and estimating continuous temporal network properties. Information about pattern features and contin-

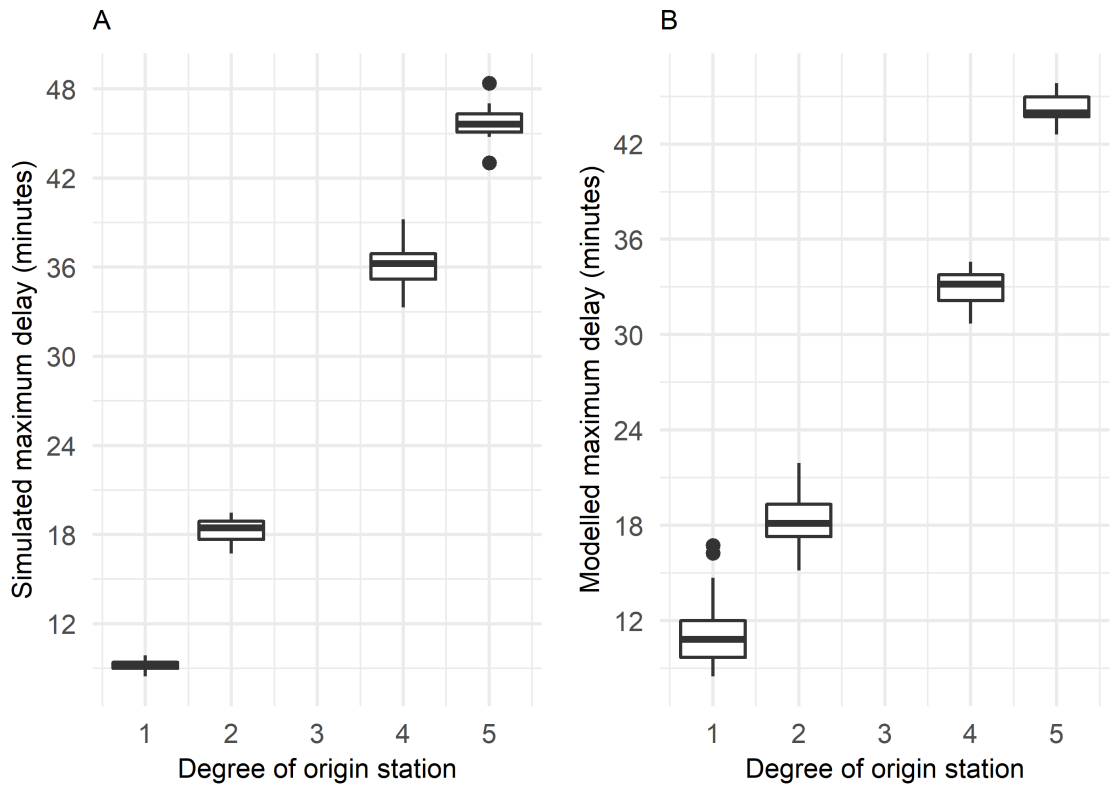


Figure 7.11: Relationship between degree of origin station and the simulated (panel A) and modelled (panel B) maximum delay for origin-destination pairs in the simulated data.

uous temporal network properties was easy to interpret with a clear 'real-world' meaning. In this case, these were the times of day at which users experienced the longest delays to journeys for each origin-destination pair and the length of these delays, along with a function representing the average journey speed over time for the whole network. The results were easily translated into interpretable summaries that address the aims of the illustrative analyses. The relationship between the pattern features and network properties was easy to display. It would also be possible for the relationship between pattern features and network or spatial information to be modelled or mapped.

The major challenge that arose when fitting these models was parameterising them in a form that allowed efficient estimation by the Bayesian sampling algorithm. Cross-classified models are often difficult to estimate, requiring long burn-in periods and often producing poor mixing. An alternative parameterisation of these models that involves including fixed effects (β_0 and β_n in equation 7.1) as the mean value of the aspatial random effects (u_o and u_n in equation 7.1), rather than fixing their mean to zero has been suggested to improve mixing (Browne and Draper 2006). However, results from models with this parameteri-

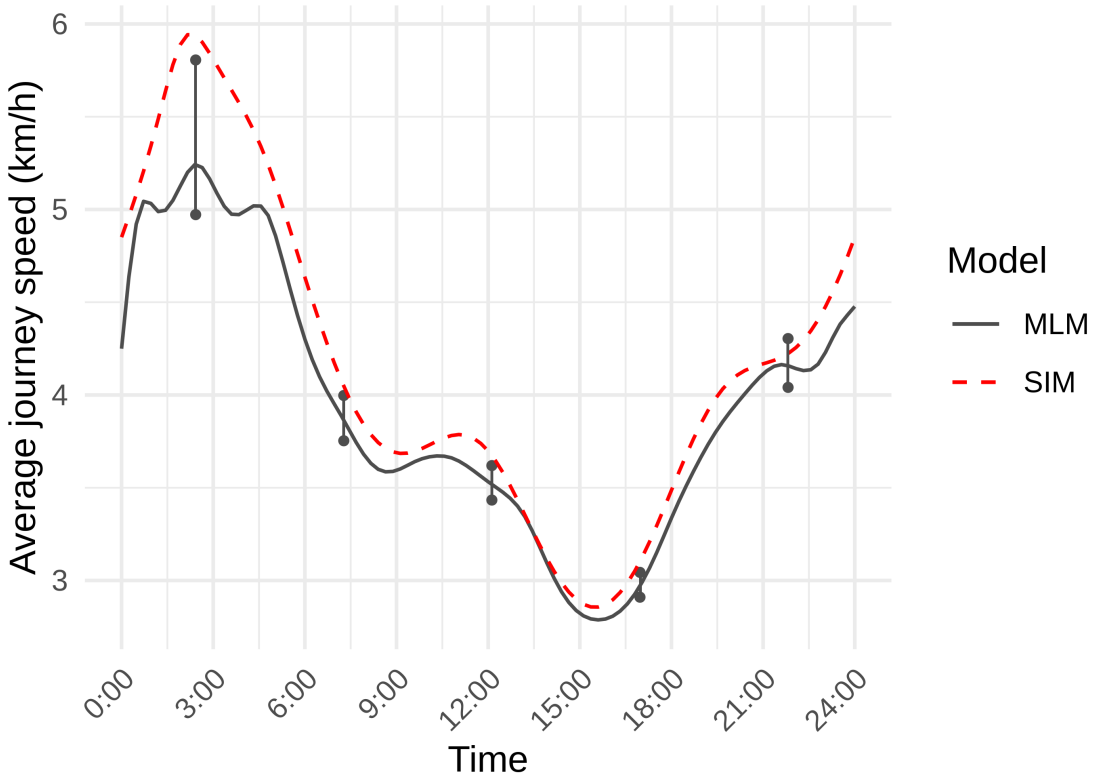


Figure 7.12: Simulated (dashed red line) and modelled (solid black line) temporal functions of average journey speed across the whole simulated network, with 95% credible intervals.

sation differ substantially from those with separate fixed and random effects, resulting in large underestimation of the temporal function of average speed for the simulated data. Incorporating temporal autocorrelation in error terms also proved difficult. The initial approach used was to specify the outcome variable as a deterministic function of two models – one for the mean value and one for the error term. This meant that model fit criteria could not be estimated. An alternative, equivalent parameterisation that expressed the previous error term as a function of the fixed effects was used as an alternative (Congdon 2014). This is outlined in the supplementary material code. The long burn-in period and complexity of this model meant that estimation with OpenBUGS was slow. Alternative Bayesian estimation software, such as JAGS (Plummer 2003) or Stan (Carpenter et al. 2017), might provide more efficient and faster estimation; however, these do not currently have the option to specify CAR spatial distributions for random effects, as in the present example. The model used in this example did not include network autocorrelation, but it would be possible to include this using a similar method to the spatial autocorrelation terms (Freni-Sterrantino et al. 2018).

The methods for simplifying and displaying these results rely on interpolation and calculus and, as such, could not easily be applied to discrete-time models, such as STARIMA, which rely on autoregressive or moving average structures (T. Cheng et al. 2011; T. Cheng et al. 2014). Treating time as a continuous variable means that the models are inherently equipped to deal with variation in measurement times and intervals between measurements, whereas methods using discrete time lags assume measurements take place at these evenly spaced intervals. This means that the applicability of the methods in this example extends to spatio-temporal network data with uneven measurement intervals (e.g. those not collected automatically or those from combined datasets). Continuous-time models also avoid aggregation of data into discrete temporal units, and do not require systematic sampling. Both processes have been shown to introduce bias in autoregressive models and to reduce the flexibility with which temporal autocorrelation parameters and periodicity can be specified (J. Freeman 1989; Hawawini 1978; Rossana and Seater 1995; W. Wei 1981; Weiss 1984).

The use of discrete time lags does have some advantages: it easily allows for the use of time-varying spatial weight matrices, as in LSTARIMA and NSTARIMA (T. Cheng et al. 2011; T. Cheng et al. 2014). In the model used for this example, spatial correlations remain constant across time and temporal autocorrelations are constant across space.

The MCMC estimation procedure used to fit models in this paper provided an ideal solution to derive estimates of uncertainty for pattern features and continuous temporal edge density. In addition, estimation of complex hierarchical models, such as the one in this example, is often difficult when using maximum likelihood estimation or other frequentist procedures. However, Bayesian results can be more difficult to explain and communicate than those from frequentist, which is somewhat at odds with the use of pattern features to summarise temporal variation in an interpretable way. An alternative method of representing uncertainty in multilevel models could be confidence bands; however, these were not used as they only represent uncertainty in the response variable (journey time), so could not capture uncertainty in time of maximum delay.

Bayesian estimation is also affected by the choice of prior distributions used when estimating model parameters. These represent prior beliefs and information the researcher has about the process underlying the data, translated into distributions that parameters

are expected to follow. This can make estimation of these models extremely difficult, compared to frequentist models or machine learning techniques which require less consideration of the underlying processes. However, having more input into the model and parameter specification could be seen as an advantage – a further opportunity to ensure models reflect known information about the process they are examining, which can help to make more informed inferences (Comber and Wulder 2019).

While spatio-temporal networks in different settings come with their own specific contextual nuances, the principle of the methods illustrated in this paper (using a transport network) could be applied to a range of settings by modifying model specification as appropriate. This could include, but is not limited to, ecological or social networks.

7.5.1 Accuracy and precision of results

In capturing the maximum delay pattern feature, the model showed reasonable accuracy; for the simulated data, the mean maximum delay was very close to the simulated value, and the standard deviation only underestimated by under 1.5 minutes. The mean bias and range values of the 95% limits of agreement were small for this feature, suggesting good correspondence between the simulated and modelled values. The rank correlation between simulated and modelled values was also relatively high, suggesting the ordering of simulated and modelled maximum delay for origin and destination stations is similar.

Time of maximum delay was not captured with the same degree of accuracy. The mean value extracted from models of simulated data was close to the known mean, and the mean value recovered from the real data reflected expected commuting patterns. The mean bias for time of maximum delay was also low – around four minutes. However, for the simulated data, the standard deviation in time of maximum delay was underestimated, suggesting that the full variation in these values is not being captured. The limits of agreement for this feature spanned a large range (over two hours either side of zero. This suggests that for many origin-destination pairs, the modelled time of maximum delay does not correspond well with the simulated value. Figure 7.8 revealed that for origin-destination pairs with earlier times of maximum delay, the model overestimated values, and vice versa for later times.

These results suggest that, while this model is effective at capturing pattern features in

terms of the outcome variable (e.g. maximum delay), pattern features in the exposure variable were not sufficiently captured. This may be because no random effects in this axis were explicitly included in the model, and it relied on the flexibility of spline functions to capture different times of maximum delay. Including an exposure variable random effect would be possible and is an approach sometimes used in models for childhood growth (Cole et al. 2010). It is also possible that including prior distributions with much higher values for the standard deviation of random effects would capture more of the variation in the time of maximum delay.

The precision of model estimates for the real data analysis was low, with credible intervals for both maximum delay and time of maximum delay spanning large ranges. For the simulated data, the model was also imprecise in capturing time of maximum delay, but the credible intervals for maximum delay were much smaller than in the real data model. The continuous temporal function of average speed also had wide credible intervals towards the ends of the range of the data for both simulated and real data analyses. Some of the uncertainty in the average speed functions may be due to instability of b-spline functions towards the end of their range. This uncertainty could possibly be reduced by using natural splines as an alternative – these have additional constraints and are less erratic at the boundaries of the data (Perperoglou et al. 2019).

The lack of precision for pattern features and the average speed function in the real data may be related to small numbers of observations for some origin-destination pairs, particularly given the complexity of the model used. The data are also unevenly distributed throughout time, with most observations concentrated at commuting times. This may leave gaps in the data, or time periods with sparse information to inform the model, increasing uncertainty in these areas (Kim et al. 2011). In this example, only a small sample of the real journey data was available, which may exacerbate this problem by making areas with very sparse data more likely as there are fewer observations overall. Using the full dataset could improve precision but would also increase the running time of the model (models for the real data ran in 17.5 hours with four parallel chains, models for simulated data ran in 20.7 hours with no parallelisation). Imbalance of data in the exposure variable is likely to present a problem for many different analyses that use data relating to individual journeys, as they do not occur randomly throughout time (Sperrin et

al. 2017). More extensive simulations could be carried out to investigate the computational cost of fitting the specified models to data of different sizes and with different degrees of imbalance.

For the simulated data, the function of average speed underestimated the simulated value along its range. Mostly this was by a small amount, but before 5:00 the difference reached almost one kilometre per hour. This larger difference may indicate poor model fit in this region, perhaps suggesting that the spline knot specification is not ideal. Automatic knot selection procedures for splines focus on single level, rather than multilevel, models (Dung and Tjahjowidodo 2017; Yeh et al. 2020). Therefore, for this analysis, a range of spline knot points were tested, and the one with the lowest Deviance Information Criterion was chosen. Visual examination was not used, as the large number of individual origin-destination trajectories to be fitted made this infeasible. As a result, the knot point specifications included evenly spaced knots or knots spaced at the quantiles of the data; specifications did not reflect areas of changeability in the data where it may be advisable to include more knots (Holmes and Mallick 2003). In the area where average journey speed is substantially underestimated, there is a relatively sharp peak in the data. It is possible that including more knot points in this area would improve the accuracy with which this peak was captured. In practice, it is advisable to consider both model fit information (such as Deviance Information Criterion) and prior knowledge about the underlying processes in the data to inform knot point selection, for example, by positioning more knots at times where more changeability in trajectories is expected (Holmes and Mallick 2003).

For the real data, the function of average speed is particularly jagged towards the end of the time period (after 20:00). The average function of journey time is smooth, as well as the sum of all individual journey time functions. However, individual level functions of journey times are more erratic in after 20:00 due to the low density of data in this period. When these functions are used as denominators in ratios (as in the equation for average speed), their peaks and troughs are exaggerated, producing much more jagged individual functions. This means that the sum and average of speeds no longer necessarily follow a smooth function. This is particularly noticeable in the period after 20:00 due to the erratic behaviour of individual origin-destination functions in this region and is likely to contribute to the wide credible intervals in this period.

For both the real and simulated data, a clear relationship between vertex closeness or degree and maximum delay was captured. In the simulated data, the results matched the simulated relationship closely. In the real data, the relationship between closeness and maximum delay is the opposite to the simulated data. The negative relationship between origin vertex closeness and maximum delay suggests that more 'well-connected' stations experience smaller delays to journeys. This could be due to a number of reasons: journeys starting at less well-connected stations may have fewer train services running, resulting in more waiting time between scanning an Oyster Card and leaving the origin station; more changes may be needed for journeys from these stations, as they are less likely to be connected to as many London Underground lines as stations with high closeness; the journeys from these stations, which are likely to lie towards the edges of the network, may be longer, and so small delays may compound over a larger time period; or systems to reduce delays may already be in place for stations with higher closeness, because they serve many London Underground lines or are in the city centre.

7.6 Conclusion

This paper illustrates the use of multilevel models to capture continuous temporal patterns of delays in a transport network. Two methods were demonstrated for simplifying the reporting and interpretation of results: extraction of temporal pattern features in order to examine variation in temporal patterns in relation to local network properties; and calculation of functions representing continuous temporal patterns of whole-network properties to examine changes in the use of the whole-network over time. Both the methods rely on interpolation, of which uncertainty is a key feature. Methods for estimating this uncertainty, based on MCMC estimation procedures, were also described.

Models recovered some pattern features accurately, but precision was low for the real data example, possibly because of regions of data sparsity or model specification. Further work that examines the effect of more complex data structure and model specification on the accuracy and precision of methods is therefore warranted. While the application of this model to real data was a fairly complex approach and presented challenges in terms of parameterisation and computational intensity, the results are ultimately simple to interpret and to analyse further, demonstrating the potential utility of this method for

examining spatio-temporal network data with inconsistent measurement intervals.

7.7 Acknowledgements

Sarah Gadd is funded by the Economic and Social Research Council [ES/P000746/1]. Alison Heppenstall is funded by an Economic and Social Research Council-Alan Turing fellowship [ES/R007918/1]. Mark Gilthorpe is funded by the Alan Turing Institute [EP/N510129/1]. Alexis Comber and Keiran Suchak received no specific funding for this work. This work uses data from TFL Open Data.

References

Aigner, W., Miksch, S., Müller, W., Schumann, H., & Tominski, C. (2007). Visualizing time-oriented data - A systematic view. *Computers & Graphics*, *31*(3), 401–409.

Angeloudis, P., & Fisk, D. (2006). Large subway systems as complex networks. *Physica A: Statistical Mechanics and its Applications*, *367*, 553–558.

Austwick, M., O'Brien, O., Strano, E., & Viana, M. (2013). The structure of spatial networks and communities in bicycle sharing systems. *PLOS One*, *8*(9).

Besag, J., & Kooperberg, C. (1995). On conditional and intrinsic autoregression. *Biometrika*, *82*(4), 733–746.

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), 1–20.

Bland, J., & Altman, D. (1999). Measuring agreement in method comparison studies. *Stat Methods Med Res*, *8*(2), 135–60.

Blonder, B., & Dornhaus, A. (2011). Time-ordered networks reveal limitations to information flow in ant colonies. *PLOS one*, *6*(5), e20298.

Browne, W., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *J Bayesian analysis*, *1*(3), 473–514.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 28.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *J Journal of Statistical Software*, *76*(1), 32.

Cheng, T., Wang, J., Haworth, J., Heydecker, B., & Chow, A. (2011). Modelling dynamic space-time autocorrelations of urban transport network. *Proceedings of the 11th international conference on Geocomputation 2011*, 215–210.

Cheng, T., Wang, J., Haworth, J., Heydecker, B., & Chow, A. (2014). A dynamic spatial weight matrix and localized space–time autoregressive integrated moving average for network modeling. *J Geographical Analysis*, *46*(1), 75–97.

Cole, T., Donaldson, M., & Ben-Shlomo, Y. (2010). SITAR - A useful instrument for growth curve analysis. *International Journal of Epidemiology*, *39*(6), 1558–1566.

Comber, A., & Wulder, M. (2019). Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS*, *23*(5), 879–891.

Congdon, P. (2014). *Applied Bayesian modelling*. Wiley.

de Haan-Rietdijk, S., Voelkle, M., Keijsers, L., & Hamaker, E. (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*, 1849–1849.

Dung, V., & Tjahjowidodo, T. (2017). A direct method to solve optimal knots of b-spline curves: An application for non-uniform b-spline curves fitting. *PLOS ONE*, *12*(3), e0173857.

Ermagun, A., & Levinson, D. (2018). An introduction to the network weight matrix. *Geographical Analysis*, *50*(1), 76–96.

Erős, T., Schmera, D., & Schick, R. (2011). Network thinking in riverscape conservation – A graph-based approach. *Biological Conservation*, *144*(1), 184–192.

Freeman, J. (1989). Systematic sampling, temporal aggregation, and the study of political relationships. *Political Analysis*, *1*, 61–98.

Freeman, L. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Freni-Sterrantino, A., Ventrucci, M., & Rue, H. (2018). A note on intrinsic conditional autoregressive models for disconnected graphs. *Spatial and Spatio-temporal Epidemiology*, *26*, 25–34.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721–741.

Goldstein, H. (1994). Multilevel cross-classified models. *J Sociological Methods Research*, *22*(3), 364–375.

Goldstein, H. (2011). *Multilevel statistical models* (Fourth edition). Wiley.

Hawawini, G. (1978). A note on temporal aggregation and serial correlation. *Economics Letters*, *1*(3), 237–242.

Heck, R., & Thomas, S. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus* (Third edition). Taylor & Francis.

Holmes, C., & Mallick, B. (2003). Generalized nonlinear modeling with multivariate free-knot regression splines. *Journal of the American Statistical Association*, *98*(462), 352–368.

Horvath, S. (2011). *Weighted network analysis: Applications in genomics and systems biology*. Springer.

Howe, L., Tilling, K., Matijasevich, A., Petherick, E., Santos, A., Fairley, L., Wright, J., Santos, I., Barros, A., Martin, R., Kramer, M., Bogdanovich, N., Matush, L., Barros, H., & Lawlor, D. (2013). Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research*, *25*(5), 1854–1874.

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, *24*(3), 383–401.

Ke, J., Zheng, H., Yang, H., & Chen, X. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*, *85*, 591–608.

Kim, K., Sentürk, D., & Li, R. (2011). Recent history functional linear models for sparse longitudinal data. *Journal of statistical planning and inference*, *141*(4), 1554–1566.

Leigh, C., Kandanaarachchi, S., McGree, J., Hyndman, R., Alsibai, O., Mengersen, K., & Peterson, E. (2019). Predicting sediment and nutrient concentrations from high-frequency water-quality data. *J BioRxiv*, 599712.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine*, *28*, 3049–3067.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251.

Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks*, *31*(2), 155–163.

Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *J Evaluation review*, *30*(1), 66–85.

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. *BMC Medical Research Methodology*, *19*(1), 46.

Peterson, E., & Ver Hoef, J. (2010). A mixed-model moving-average approach to geostatistical modeling in stream networks. *J Ecology*, *91*(3), 644–651.

Pfeifer, P., & Deutsch, S. (1980). A STARIMA model-building procedure with application to description and regional forecasting. *J Transactions of the Institute of British Geographers*, 330–349.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*, *124*.

R Core Team. (2020a). R: A language and environment for statistical computing.

Ramsay, J., & Silverman, B. (1997). *Functional data analysis*. Springer.

Rossana, R., & Seater, J. (1995). Temporal aggregation and economic time series. *Journal of Business & Economic Statistics*, *13*(4), 441–451.

Scott, J. (1988). Social network analysis. *Sociology*, *22*(1), 109–127.

Soetaert, K., & Herman, P. (2009). rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations.

Sperrin, M., Petherick, E., & Badrick, E. (2017). Informative observation in health data: Association of past level and trend with time to next measurement. *Stud Health Technol Inform*, *235*, 261–265.

Spiegelhalter, D., Best, N., Carlin, B., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.

Stimson, J., Carmines, E., & Zeller, R. (1978). Interpreting polynomial regression. *Sociological Methods & Research*, *6*(4), 515–524.

Thomas, A., Best, N., Lunn, D., Arnold, R., & Spiegelhalter, D. (2014). GeoBUGS user manual.

Ver Hoef, J., & Peterson, E. (2010). A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association*, *105*(489), 6–18.

Wang, W., & Yan, J. (2020). splines2: Regression spline functions and classes.

Wei, W. (1981). Effect of systematic sampling on ARIMA models. *Communications in Statistics - Theory and Methods*, *10*(23), 2389–2398.

Weiss, A. (1984). Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics*, *26*(3), 271–281.

Xiao, N. (2016). Basic geometric operations. *GIS algorithms: Theory and applications for geographic information science & technology*. SAGE Publications Ltd.

Yeh, R., Nashed, Y., Peterka, T., & Tricoche, X. (2020). Fast automatic knot placement method for accurate b-spline curve fitting. *Computer-Aided Design*, *128*, 102905.

Chapter 8

The utility of multilevel models for continuous-time feature selection of spatio-temporal networks

This chapter addresses the final aim of the thesis: to evaluate the accuracy and precision of the analytical procedure proposed in Chapters 4 and 7. This is achieved through a simulation experiment where the maximum delays in journeys between origin-destination pairs in the rail network are recovered. The simulations in this chapter are generated using the simulation tool from Chapter 4.

8.1 Abstract

Many models for the analysis of spatio-temporal networks specify time as a series of discrete steps. This either requires evenly spaced measurement times or the aggregation of data into measurement windows. This can lead to the introduction of bias. An alternative is to use continuous-time models, for example, multilevel models. Models capturing complex spatio-temporal variation are often difficult to visualise and interpret. This can be addressed by simplifying the results, for example by extracting 'features' of interest (such as maxima or minima) of temporal patterns associated with different network connections.

This paper uses simulation to evaluate the accuracy and precision with which b-spline-based multilevel models (a flexible form of continuous-time model that can easily capture

complex variation associated with a spatio-temporal network structure) capture the timing and extent of maximum delays to journeys made between pairs of stations in a small railway network.

On average models captured the timing and extent of maximum delay with small bias, but there was evidence of overestimation and underestimation of low and high values of these features, respectively. This systematic bias may have partially caused the undercoverage of credible intervals for the pattern features. Alternative model specifications – specifically to capture x-axis random variation, for example – should be considered in future work.

8.2 Introduction

Geographers are often interested in investigating the movement of people between or around cities. This typically involves public transport or other transport networks. Transport systems can be described by a series of connections, made by travel, between a set of locations. These connections between locations can be represented in the form of a network or graph (Newman 2018). For example, Y. Yang et al. (2019) used spatio-temporal networks to represent a public bicycle sharing system; T. Cheng et al. (2011) represented a road system as a spatio-temporal network when predicting traffic counts; and S. Chen, Claramunt, et al. (2014) represented a metropolitan rail system as a spatio-temporal network when studying accessibility.

Networks involve objects or locations represented as vertices connected by edges. For example, a rail network may include stations as vertices and edges where the stations are connected by railway lines. Each vertex and edge might be assigned properties, for example, a station might have a set capacity for people awaiting trains, or a railway line might have a speed limit (Newman 2003). Edges can be bidirectional or unidirectional. In the latter case, an edge between two vertices only specifies a connection in one direction; for example, trains may be able to travel from station A to station B, but not in reverse (Newman 2003).

Networks are used across a range of disciplines, for example: in studies of the social connections between people (X. Li and Griffin 2013); ecological networks, including animal interactions and river network settings (Anderson and Dragićević 2018; Neeson et al. 2012);

transport networks (Abdelghany et al. 2001); or economic trade networks (S. Liu et al. 2018). Some networks (e.g. transport, trade and ecological) have a spatial aspect, where vertices are in specific locations. Many networks also have a temporal aspect, where the network connections or the properties of edges and vertices change over time (Blonder, Wey, et al. 2012). When such properties change over time it may be of interest to model them, for prediction or causal inference purposes.

When analysing such spatio-temporal networks, it is important to account for the influence of temporal and spatial correlations on the processes of interest (Dubin 1998). Many models that aim to do this incorporate time as a series of discrete steps, rather than a continuous variable (T. Cheng et al. 2014). This limits the models to evaluating temporal properties of network objects at discrete time points, making it more challenging to use them for predictions or inferences at any time between these discrete points. As such, the length of these discrete time steps is of great importance and must be chosen carefully. Steps of an inappropriately large length can mask intermediate temporal variations in the data (J. Freeman 1989; Weiss 1984). However, the choice of the length of time steps is often predetermined by the frequency of measurements in the data generating process; if measurements are taken every ten minutes, for example, time steps shorter than this cannot be used (J. Freeman 1989). If data are not measured simultaneously for different network objects, or at regular intervals, data must be aggregated into common, regular time intervals for analysis (J. Freeman 1989; Hwang 2000). While this allows for some degree of choice in the length of the time steps (though still limited by the frequency of measurements), aggregation can also bias inference as it introduces error in the measurement times of observations, which many statistical models assume to be error-free (J. Freeman 1989; Hwang 2000). While discrete-time models have been used to great advantage in the context of spatio-temporal network analysis, it may be possible to avoid some of the issues outlined here by using continuous-time models (Niezink et al. 2019; Oud et al. 2012). Multilevel (or mixed effects) spline models are capable of modelling complex temporal patterns of properties for multiple observational units (Goldstein 2011). These models are able to account for spatial and temporal correlations among objects and could potentially be used to model temporal patterns of properties for a set of objects in a network (Besag, York, et al. 1991).

Interpreting information regarding spatio-temporal networks is also difficult – the inclusion of multiple dimensions (i.e. two-dimensional space and time) makes visualisation challenging. T. Li and Liao (2016) developed an interesting solution to the visualisation of temporal patterns of edge properties – each edge was divided into a number of time segments which were coloured to indicate the value of some property during each time window. For multiple properties on one edge, each time segment was further subdivided into smaller coloured sections. This method has some shortcomings – first, time is still treated as a discrete variable made of a series of time steps; second, the length of each edge in the network can be expected to differ, so time is not represented on a consistent scale across the network. This may confuse interpretations from the visualisation when comparing temporal patterns of properties between edges of different lengths. Subdivision of time segments to represent multiple properties occurs in the same dimension as the division into time segments. This may make it appear as if each property is measured at a different time. Furthermore, if edges overlap, as they may for dense networks, some time periods of interest will be obscured for certain edges.

Colak et al. (2013) represented changes in temporal properties of network edges in a different manner, in the context of a road traffic network. This involved focusing on a feature of the temporal pattern that constituted an event of interest. In this case, this was the time at which each edge (road segment) in the network reached 80% of capacity, thought to be the threshold at which 'congestion' exists. The authors represented this feature of the temporal pattern of edge properties in a map of the network with each edge coloured to indicate the time at which it reached 80% capacity. Including only one feature of the patterns of interest helps simplify the visualisation and reduces issues associated with temporal scales and overlapping edges. While less information is included than in the proposal by T. Li and Liao (2016), the visualisation answers a specific research question – i.e. when congestion occurs for each edge in the network. Having a specific question of interest is important in developing an appropriate analysis based on prior knowledge of the processes concerned and for avoiding some problematic practices such as 'p-hacking' (Head et al. 2015).

The specification of time in Colak et al. (2013) was discrete, yet pattern features for properties of individual edges or vertices can still be recovered from continuous-time mul-

tilevel models. This process often involves the use of calculus to extract features – such as minima, maxima, and slopes of a model – once fitted to the data (Gadd, Comber, et al. 2021). However, the accuracy and precision with which multilevel models can capture these features is unknown.

Assessing the performance of statistical methods can be carried out using algebraic results or repeated simulation and analysis of data (Morris et al. 2019). Unlike finding algebraic results regarding model performance, simulations only represent model performance in a specific situation, but they are often easier to conduct where algebraic solutions are difficult (or impossible) to produce and can be used to explore how changes in data structure, for example, can affect models (Morris et al. 2019). In addition, framing a simulation in a tangible example can aid understanding of the results describing model performance.

This paper aims to use simulations to evaluate the accuracy and precision with which continuous-time multilevel models recover features of temporal patterns of edge properties in a spatio-temporal network context. In particular, this paper focuses on the extraction of maxima from temporal patterns of journey lengths (the time taken to travel between a pair of stations) in a rail network as this measure has potential practical uses in identifying times and locations with very large delays in a rail network.

8.3 Methods

This section first outlines the process of generating simulated data associated with a spatio-temporal rail network, based on a real dataset and with known properties. This is followed by specification of the continuous-time multilevel model and process used to extract pattern features (maximum journey length, time of maximum and maximum delay). Finally, the methods used to compare the estimated temporal pattern features to the simulated values in order to assess accuracy and precision are outlined.

The ability of models to recover maxima was tested on simulated data representing journey lengths between pairs of rail stations over the course of a day (Gadd, Comber, et al. 2021). A series of simulated datasets with known underlying values were generated and analysed. The accuracy with which models recover these underlying values can then be assessed by investigating the difference between simulated and modelled values (Bland and Altman

1999). The relationship between estimated uncertainty and simulated values was examined to assess precision. Simulations are a simple way of assessing model performance in specific circumstances, for example, with specific data structures and were chosen here as a way of assessing one such circumstance (Morris et al. 2019).

8.3.1 Data

Simulated data were based on a small section of the London Underground network comprising 12 stations, shown in Figure 8.1 (Transport for London 2009a; Transport for London 2009b). This area was chosen as it included stations with a range of numbers of connections – some being very well connected, like King's Cross and Euston, and some with few connections, like Mornington Crescent and Goodge Street.

For each of the 132 origin-destination pairs in the network, data representing the length of time taken to complete 50 journeys at random times in one 24-hour period were generated. Each origin-destination pair had a 'true' underlying function of journey lengths over the 24-hour period, all based on the function shown in Figure 8.2. This function is intended to represent increased congestion, resulting in longer journey lengths due to waiting for trains and making slower changes, during a morning and afternoon rush hour.

There was random variation in the intercept and amplitude of this function between each origin-destination pair. Two pairs sharing an origin or destination were more similar than pairs with no shared stations.

The observed journey lengths were allowed to vary randomly around the 'true' function for each origin-destination pair. This random variation was allowed to covary based on temporal and spatial proximity (of origin and destination stations) of the observations.

Some variation in the amplitude of temporal patterns of journey lengths was introduced in relation to the properties of origin stations for each origin-destination pair. Journeys from stations with more connections (a higher degree) had greater changes between the shortest and longest journey lengths (a larger amplitude), resulting in larger maximum delays for these stations. This is intended to reflect that these stations may become busier during rush hours than smaller stations, leading to more congestion and delay. This 'fixed' variation was included in the simulation process, but information was not retained for analysis – this is intended to represent latent or underlying subgroups in data that we

Figure 8.1: Map of the area of the London Underground network used to simulate data

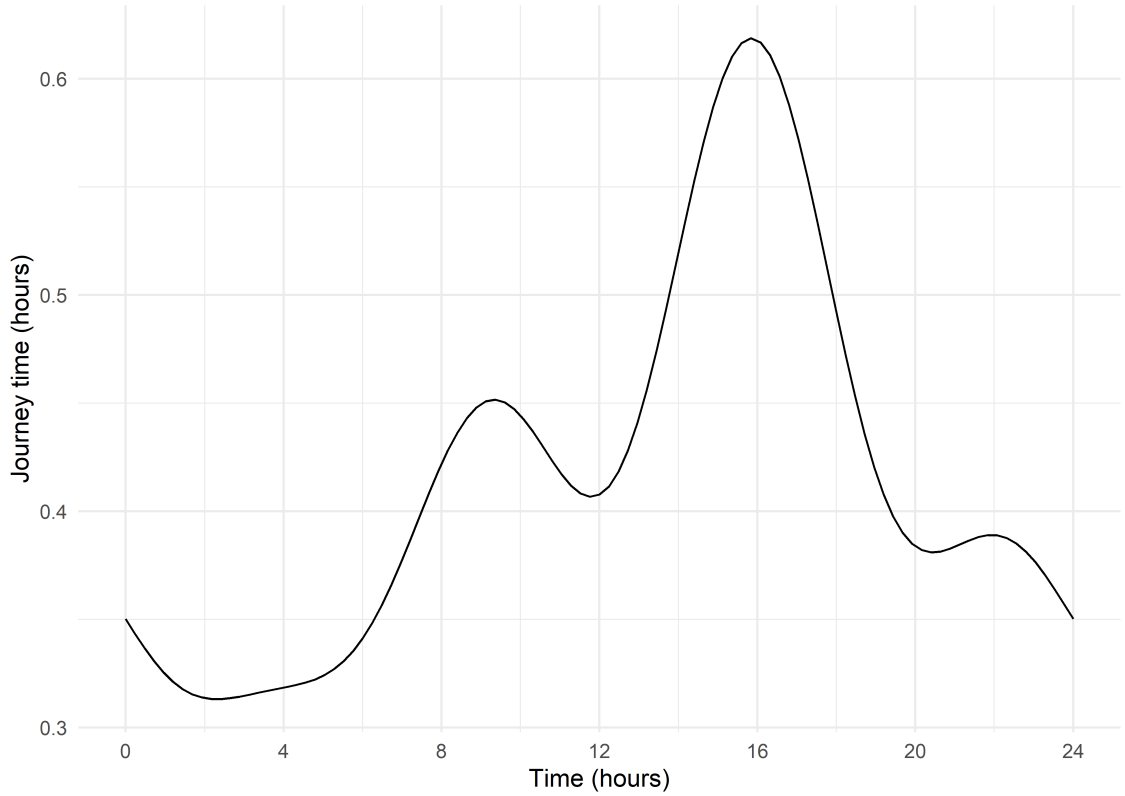


Figure 8.2: Basic function used to simulate temporal patterns of journey lengths

do not have information about and to test the model's ability to cope with this situation.

An example set of simulated data is shown in Figure 8.3. Data were simulated using R (R Core Team 2020a), code for simulation and analysis is available in the Supplementary Material.

8.3.2 Model

The journey lengths for each origin-destination pair were modelled over time using a multilevel model with a b-spline basis. Multilevel models were chosen because of their ability to model continuous temporal functions of properties for multiple observational units (origin-destination pairs), and their ability to incorporate complex random structures to account for spatial, temporal and network covariance of observations (Besag, York, et al. 1991; Goldstein et al. 1994). A b-spline basis comprises a series of functions that take a positive value in part of the range of the data and are equal to zero elsewhere. When estimating a model with a b-spline basis, the linear combination of these functions that best fits the data is estimated (Perperoglou et al. 2019). B-splines were chosen over alternatives, such as penalised splines, because they are mathematically convenient; the

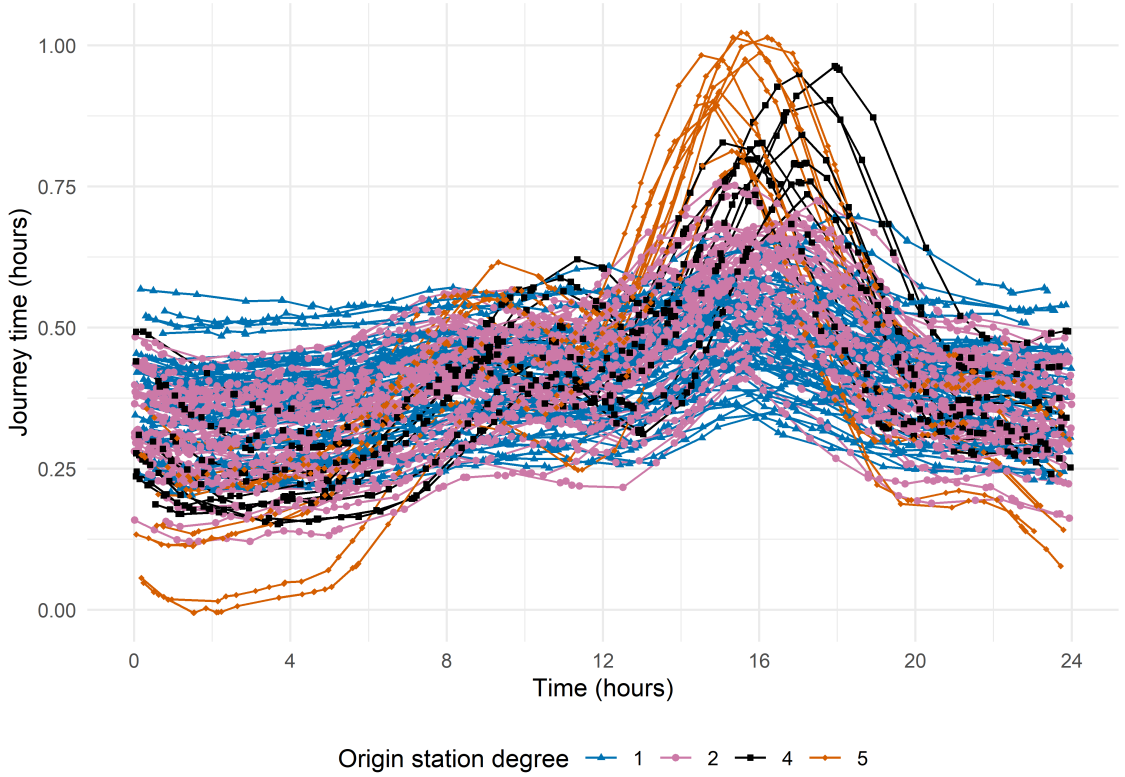


Figure 8.3: Example of simulated dataset representing temporal patterns of journey lengths between different origin and destination stations

value of each b-spline basis function can be pre-calculated for each observed measurement time and included in the model as a covariate, which is not the case for splines where a smoothing parameter must be estimated (Perperoglou et al. 2019).

The model (shown in Equation 8.1) for journey length ($y_{i,from,to}$) included an intercept ($\beta_{0,from,to}$), coefficients ($\beta_{n,from,to}$) for each of k basis functions ($f_n(t_i)$) that make up the b-spline basis and an error term ($e_{i,from,to}$).

$$
\begin{aligned}
y_{i,from,to} &= \beta_{0,from,to} + \sum_{n=1}^{k} \beta_{n,from,to} f_n(t_i) + e_{i,from,to} \\
\beta_{0,from,to} &= \beta_0 + u_{0,from} + u_{0,to} + v_{0,from} + v_{0,to} \\
\beta_{n,from,to} &= \beta_n + u_{n,from} + u_{n,to} + v_{n,from} + v_{n,to} \\
e_{i,from,to} &= \frac{a e_{i-1,from,to}}{(t_i - t_{i-1} + 1)}
\end{aligned}
\tag{8.1}
$$

The values of the intercept and all coefficients for each edge were allowed to vary ran-

domly between origin-destination pairs. This random variation was split into parts (lines two and three of Equation 8.1): aspatial random variation associated with the origin vertex ($u_{0,from}$); aspatial random variation associated with the destination vertex ($u_{0,to}$); spatial random variation associated with the origin vertex ($v_{0,from}$); and spatial random variation associated with the destination vertex ($v_{0,to}$). The aspatial random variation was constrained to follow a normal distribution with mean zero. The spatial random variation was set to follow an intrinsic Gaussian conditional autoregressive (CAR) distribution such that random effects associated with spatially close vertices would be more highly correlated than spatially distant ones (Besag, York, et al. 1991). This constrains the random effects to sum to zero. Having random effects associated with both the origin and destination stations meant that edges sharing an origin or destination vertex are likely to be more alike than those without a shared vertex.

In addition to the random effects, the error term in the model ($e_{i,from,to}$) was set to follow a first order continuous-time autoregressive structure to account for temporal autocorrelation of observations. This is detailed on the fourth line of Equation 8.1: the i^{th} error term for a given origin-destination pair ($e_{i,from,to}$) is a function of the previous error term for this origin-destination pair ($e_{i-1,from,to}$), multiplied by an autoregressive coefficient (a), divided by the time difference between the two observations plus one ($t_i - t_{i-1} + 1$ (the addition of one is to account for potential simultaneous observations)), plus a random error term, $w_{i,from,to}$, that is constrained to follow a normal distribution with mean zero. This means that the random variation for each pair of subsequent observations is correlated, with the extent of this correlation decreasing if the time difference between them increases (Goldstein et al. 1994).

The number of basis functions and the range in which each one is positive is specified using a series of knot points. The specification of these knot points was chosen separately for each known simulated pattern. Automatic knot point selection procedures are available for single level spline models (Yeh et al. 2020; Yuan et al. 2013). However, as no such procedure has been identified for a multilevel context, a selection of 40 possible knot placements were tested and the knot placement resulting in the lowest Deviance Information Criterion (a model fit criterion for Bayesian estimation) was chosen. The number of knots in the possible specifications ranged from one to 20, with knots either evenly spaced over time

or placed at evenly spaced quantiles of the distribution of observations over time. Placing knots at equal spaces along the time-axis and at quantiles of the data distribution are common knot placement strategies (Ramsay and Silverman 1997). An alternative method is to place more knots at times where curvature in temporal patterns is high (Holmes and Mallick 2003; Ramsay and Silverman 1997). In this case, equally spaced knot placement strategies were chosen as the areas of high curvature will vary between origin-destination pairs. Using quantiles guarantees that all splines cover an equal number of observations, meaning estimates for all spline coefficients are supported by the same amount of data even if data are not evenly distributed (Howe et al. 2013; Ramsay and Silverman 1997). In this simulation, we expect data to be distributed evenly along the time-axis. The knot specification chosen was 19 knots spaced evenly along the time-axis.

Models were fitted using a Bayesian Markov chain Monte Carlo estimation procedure in R and OpenBUGS (Lunn et al. 2009; R Core Team 2020a; Sturtz et al. 2005; A. Thomas et al. 2014). OpenBUGS is a modelling program that uses Markov chain Monte Carlo estimation (with a Gibbs sampling algorithm) to fit complex model structures (S. Geman and D. Geman 1984). This was chosen as Bayesian estimation methods are generally more effective at estimating complex model structures than maximum likelihood methods. OpenBUGS was chosen as alternative Bayesian engines did not have the capability to fit spatially correlated random effects (L. Muthén and B. Muthén 2010; Plummer 2003; Stan Development Team 2019). Bayesian estimation aims to estimate a probability distribution for model parameters, rather than a set of point values, and produces a series of posterior parameter samples. This is based on the model specification, data and prior distributions provided. Prior distributions represent current beliefs about which values are most likely for each parameter (Heck and S. Thomas 2015). In this case, loose priors that do strongly indicate any prior beliefs were specified as this is not an analysis intended to fully reflect Bayesian thinking, but rather one in which Bayesian estimation tools are the most suitable.

Initially, models with loose priors proved difficult to estimate, even with the aid of this tool. To aid estimation, models were initially run *without* the first order temporal autoregressive error structure included. This produced a set of posterior distributions that could be included as prior distributions for some parameters in the full model. As recommended when using intrinsic CAR random effects, the fixed effects parameters (β_0 and β_n in

Equation 8.1) were set to follow a flat prior in this model (Besag, York, et al. 1991; A. Thomas et al. 2014). The precision parameter for random effects distributions was set to follow a gamma prior distribution with shape and rate equal to two. The starting value for the precision parameter was set to one and starting values for fixed effects were sampled from a standard normal distribution. The initial model was run with four chains of 4000 iterations each, 3000 of which were discarded as model burn-in. This produced parameter distributions (to be used as priors in the final model) that allowed the final model (with temporally autocorrelated errors) to be consistently estimated with a low percentage of convergence failures (defined here as when the model estimation process failed to complete) in the simulations.

The posterior distributions for fixed effects from this initial model were used as priors and to choose starting values in the final model. The autocorrelation parameter was set to follow a gamma distribution with shape and rate equal to one, truncated at zero and one as positive autocorrelation was expected. The starting parameter was set to 0.5. The initial values and distributions for random effects precision parameters remained the same as the initial model. The final model was run with four chains of 20000 iterations each, 19000 of which were discarded as model burn-in. The remaining 4000 samples of parameter estimates provided empirical posterior distributions for each parameter and pattern feature from which means and 95% credible intervals were derived to compare with the known simulated values.

8.3.3 Comparison with simulated values

1000 simulations were generated, and the temporal patterns of edge properties analysed using the multilevel model specified above. The first derivative of the model for each individual edge was calculated to extract the maximum journey length, the time of this maximum, and the range of the function (the difference between the maximum and minimum) during the 24-hour period (Gadd, Comber, et al. 2021). The range of the function corresponds to the maximum delay that passengers experience travelling between a particular pair of stations. Extracting the pattern features was completed using mean model coefficients for each origin-destination pair to extract a point estimate. It was then repeated with coefficients from each of 4000 posterior samples. This generated posterior distributions of the maximum journey length, delay and time (with one value for each

posterior sample) from which quantiles could be taken to represent 95% credible intervals.

For each origin-destination pair, the maximum journey length, maximum time and maximum delay were recorded, along with the width of the 95% credible interval. The simulated and modelled maximum journey length, maximum time and maximum delay were compared. Whether the credible intervals for these pattern features included the pattern features from the simulated temporal patterns was recorded. The simulated maximum journey length, time and delay was also converted into a z-score based on the mean modelled value and 95% credible interval for each feature. If the posterior distributions from which the 95% credible intervals are calculated are normally distributed, and the 95% credible intervals are of an appropriate width then we would expect the distribution of these z-scores to follow a standard normal distribution.

This work was undertaken on ARC3, part of the High-Performance Computing facilities at the University of Leeds, UK. Code for simulation and analysis of data is included as Supplementary Material.

8.4 Results

Across 1000 simulations, 98.5% of models ran successfully with only 15 failing to complete estimation. Running one simulation with each chain of the multilevel model estimated in parallel took approximately two days of computing time. The full simulation was run with other processes running in parallel and took approximately 14 days.

Table 8.1 compares simulated and modelled values for each of the three pattern features: maximum journey length, maximum time and maximum delay. Maximum journey length is the maximum time taken to complete a journey between a pair of stations during the day examined. Maximum time is the time of day at which this maximum journey length occurred. Maximum delay is the difference between maximum journey length and the minimum journey length between a pair of stations during the day examined.

Figure 8.4 shows a Bland-Altman plot (Bland and Altman 1999) for the agreement between simulated and modelled maximum time, including the mean bias and 95% limits of agreement. Figure 8.5 shows a density plot of the 95% credible interval widths for estimates of maximum time. Figure 8.6 shows a density plot, with mean and standard

Feature	Maximum time	Maximum journey length	Maximum delay
Simulated mean (hh:mm:ss, hours)	15:50:49	0.618	0.306
Modelled mean (hh:mm:ss, hours)	15:50:02	0.596	0.294
Simulated SD (hours)	1.117	0.141	0.188
Modelled SD (hours)	0.844	0.106	0.159
Mean 95%CI width (hours)	2.095	0.066	0.089
SD 95% CI width (hours)	1.398	0.007	0.008
Median 95%CI width (hours)	1.90	0.066	0.088
IQR 95% CI width (hours)	(1.12,2.57)	(0.061,0.071)	(0.083,0.093)
Mean bias (hours)	-0.013	-0.022	-0.012
95% Limits of agreement (hours)	(-2.28,2.22)	(-0.154,0.116)	(-0.087,0.051)
Mean simulated value as z-score	0.023	1.333	0.532
SD simulated value as z-score	3.598	4.137	1.868
Percentage of real values within CI	59.533	36.174	68.613

Table 8.1: Table summarising and comparing simulated and estimated pattern feature values. SD = standard deviation, CI = credible interval, hh:mm:ss = hours, minutes, seconds time format. Mean bias is the mean difference between estimated and simulated values. 95% Limits of agreement represent limits of the bias we would expect to find in 95% of estimates of each feature.

deviation indicated, of simulated maximum time features as a z-score of the normal distribution described by the estimated maximum time and 95% credible interval estimates. Note that this assumes that the posterior distribution of the pattern feature estimates is normally distributed.

In Figure 8.4, the simulated maximum times are mostly in the afternoon, with the majority of values falling between approximately 12:00 and 20:00. This reflects the simulated data we saw in Figure 8.3 and the average pattern in Figure 8.2, where the maxima take place around the afternoon commuting period. The mean bias is close to zero, suggesting that estimates of maximum time are, on average, unbiased. However, there is a negative correlation between simulated maximum time and modelled minus simulated maximum time. This suggests that maxima taking place early in the time period are being estimated as later than they are and vice versa. This may indicate that the model is failing to capture the full amount of variation in maximum times between edges in the network.

The 95% limits of agreement indicate the limits of the bias we would expect in 95% of estimates of maximum time from these models. These are wide, with a range of just less than 2.3 hours either side of zero. This indicates that, while the mean bias is small, there is the potential for individual estimates of maximum time to have considerable inaccuracy.

The 95% credible interval widths of maximum time are also wide: the most common width

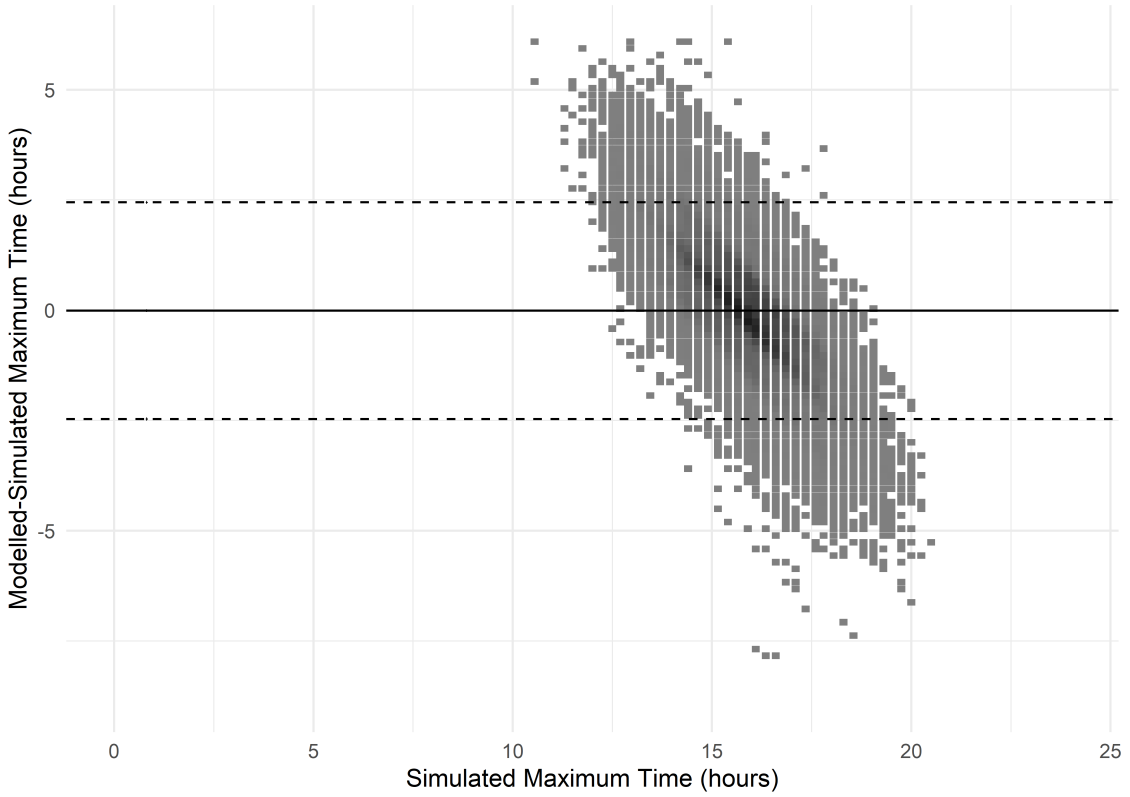


Figure 8.4: Bland-Altman plot showing the relationship between simulated maximum time and modelled minus simulated maximum time. The density of points is indicated by the colour of squares in the plot (darker grey equals greater density). The solid line indicates the mean bias and dashed lines indicated 95% limits of agreement.

is slightly less than 2.5 hours with a large proportion of intervals wider than this. This indicates that the estimates of maximum time from the model are also imprecise. The density plot for maximum time credible interval widths does not have a smooth shape but rather a series of peaks of common credible interval widths. The distance between these peaks appears to approximately correspond to the distance between knots in the b-spline basis (1.2 hours, see Supplementary Figure F.1).

If the 95% credible intervals had appropriate coverage, we would expect the distribution of simulated maximum times as z-scores of their estimated credible intervals to follow a standard normal distribution with mean zero and standard deviation one. The mean of the distribution is close to zero, as shown in Table 8.1, but the standard deviation, indicated by the dashed lines in Figure 8.6, is over 3.5. This suggests that the 95% credible intervals are too narrow and do not cover the appropriate amount of simulated maximum times. Table 8.1 shows that the percentage of simulated maximum times contained within the estimated credible intervals is just below 60%.

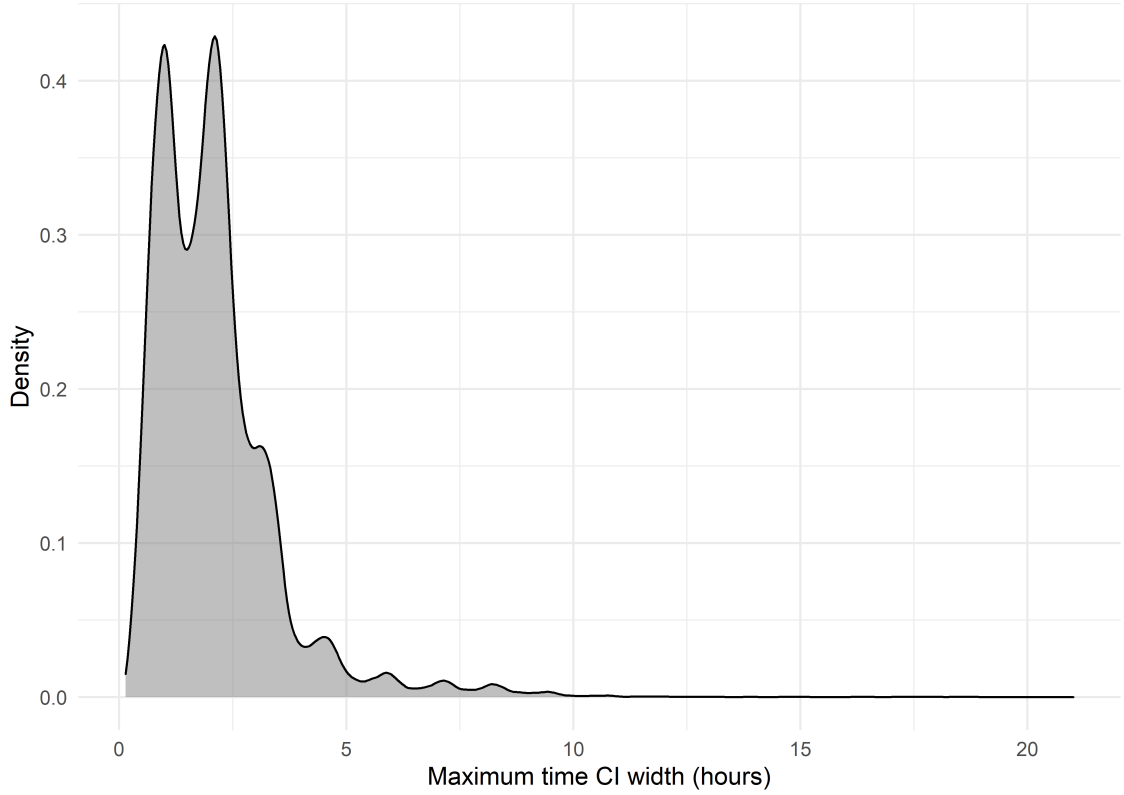


Figure 8.5: Density plot showing the distribution of 95% credible interval (CI) widths for maximum time estimates.

Figure 8.7 shows the Bland-Altman plot with limits of agreement for simulated and modelled maximum journey length. Figure 8.8 shows the distribution of credible interval widths for this feature and Figure 8.9 shows the distribution of simulated maximum journey lengths as a z-score of the credible interval distributions.

The distribution of simulated maximum journey length is multimodal as it contains lower maximum journey lengths for journeys originating at stations with few connections and higher values for journeys originating at stations with many connections. This matches the simulated data. In Figure 8.7 we see a negative correlation between the simulated maximum journey length and the degree of over- or under-estimation by the models, like in Figure 8.4. In this case, however, there is a small negative bias in the estimation of maximum journey length. On average, estimates from the models are 0.022 hours (1.32 minutes) below the simulated maximum journey length. The 95% limits of agreement are relatively narrow indicating that 95% of estimated maximum journey lengths are biased by less than 12 minutes in either direction from the true maximum journey length.

95% credible interval widths, shown in Figure 8.8, are also narrower for the maximum

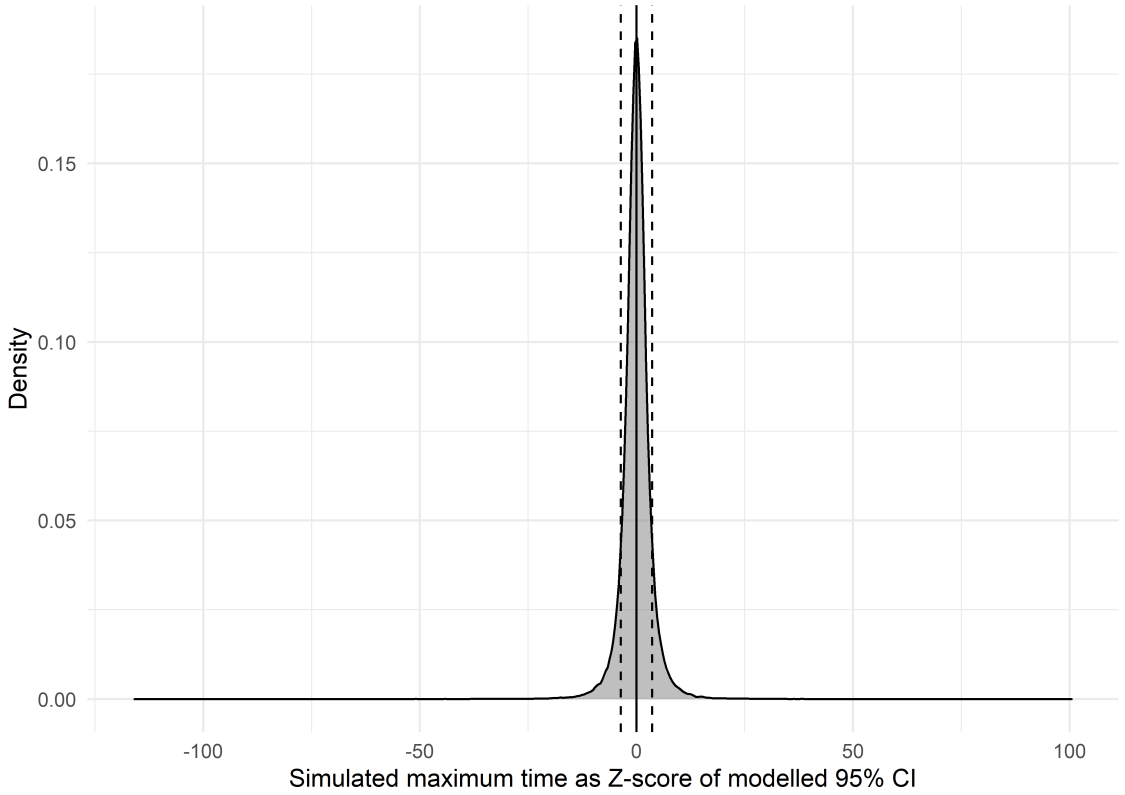


Figure 8.6: Density plot showing simulated maximum time values as a z-score of the relevant posterior distributions of estimates for maximum time. The solid line indicates the mean z-score and the dashed lines one standard deviation either side of the mean.

journey length compared to maximum time with almost all widths being under six minutes. However, the results in Figure 8.9 indicate that these may be somewhat too narrow. The standard deviation of simulated maximum journey lengths as a z-score of the distributions defined by their estimated credible interval widths is over four and the percentage of simulated values covered by the credible interval is only 36%. This indicates severe undercoverage of these credible intervals.

Figure 8.10 shows the Bland-Altman plot for simulated and modelled maximum delay with the distribution of credible interval widths shown in Figure 8.11. Figure 8.12 shows a density plot of the simulated maximum delay values as z-scores of the distributions described by their corresponding modelled maximum delay and 95% credible interval estimates.

There are four distinct groups of simulated maximum delays corresponding to the four different numbers of connections for stations in Figure 8.1. Over all these groups, a negative correlation between simulated maximum delay and modelled minus simulated maximum delay is present, similar to the other pattern features observed. However, within

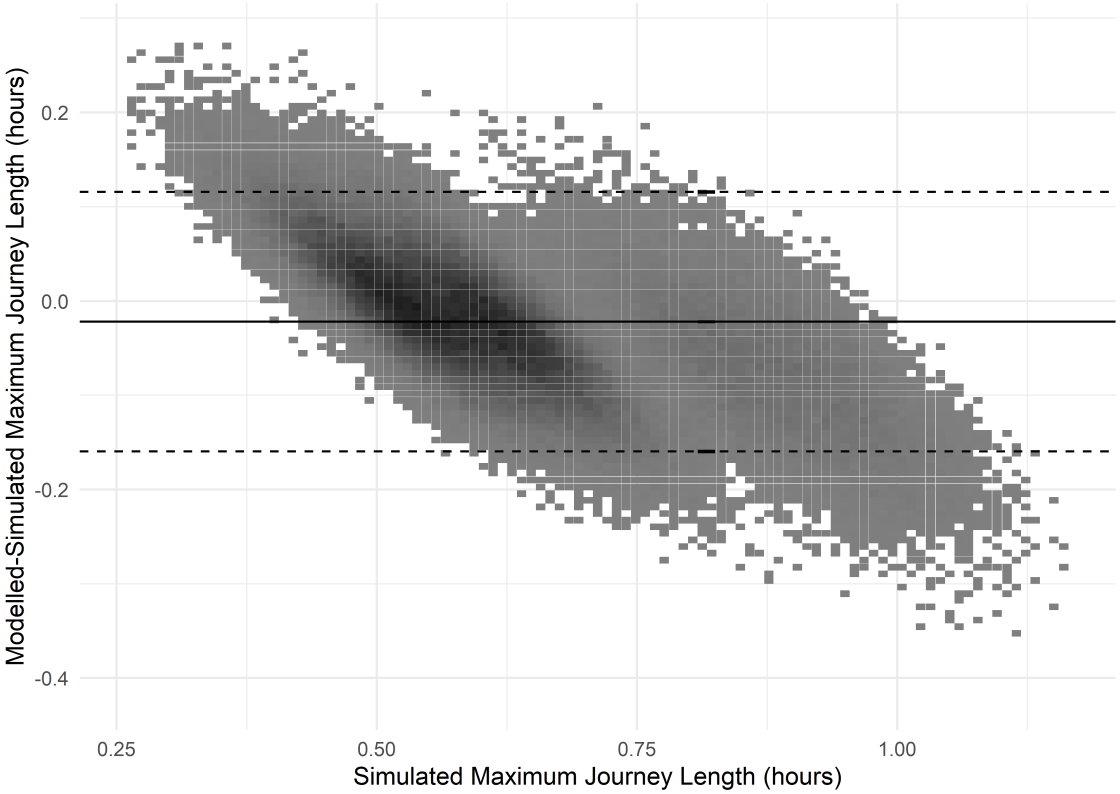


Figure 8.7: Bland-Altman plot showing the relationship between simulated maximum journey length and modelled minus simulated maximum journey length. The density of points is indicated by the colour of squares in the plot (darker grey equals greater density). The solid line indicates the mean bias and dashed lines indicated 95% limits of agreement.

some individual groups this correlation does not appear so strong. The mean bias is small and negative (-0.012 hours, less than 45 seconds) and the 95% limits of agreement are within six minutes either side of zero.

The majority of estimated 95% credible interval width for maximum delay were less than 7.5 minutes wide, however there is still an undercoverage problem with these intervals: 68% of simulated maximum delays were covered by the estimated credible intervals and the standard deviation of simulated maximum delay as a z-score of the 95% credible interval widths is 1.868. The extent of this problem is less, however, than for the other two pattern features which had wider z-score distributions and lower coverage percentages.

8.5 Discussion

The mean bias for all pattern features was relatively low, with maximum time, maximum journey length and maximum delay all being slightly underestimated by less than two

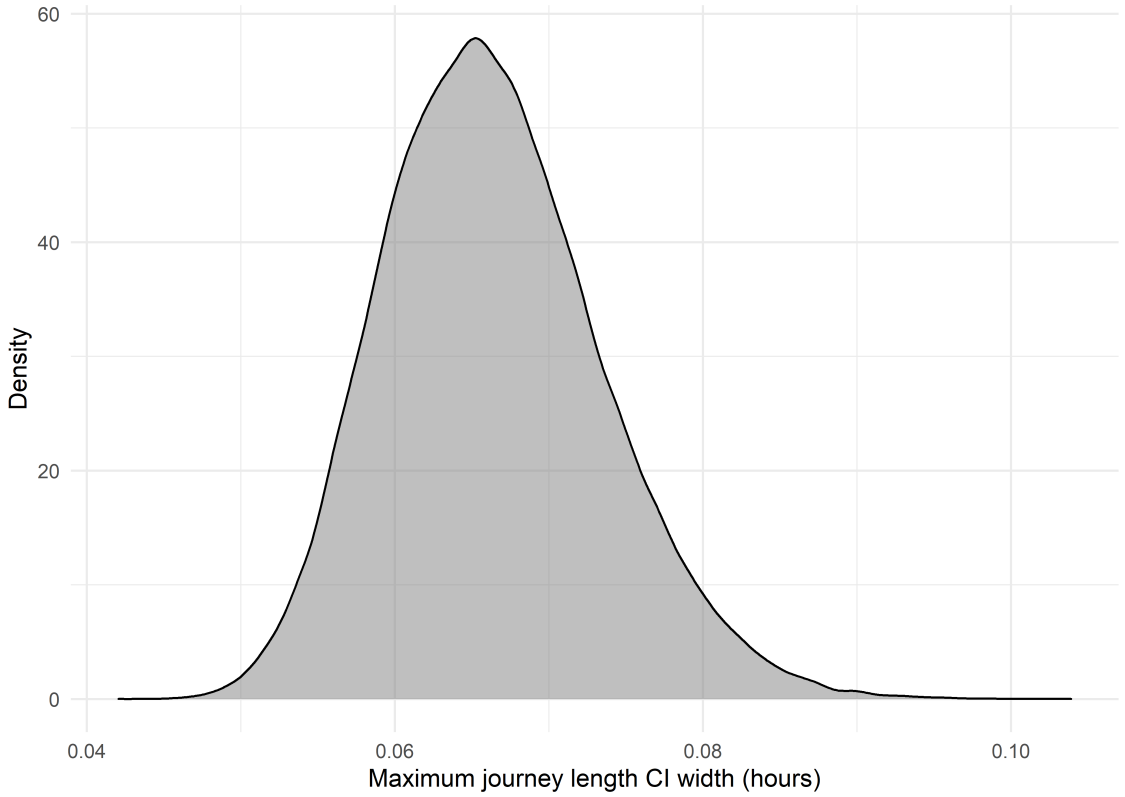


Figure 8.8: Density plot showing the distribution of 95% credible interval (CI) widths for maximum journey length estimates.

minutes on average. However, for all features, a negative correlation of the simulated pattern feature and the modelled minus simulated pattern feature was observed. This suggests that lower values are being overestimated and higher values underestimated. This may be an indication that the models are not capturing the full extent of variation in pattern features between observational units. This may partially be due to either a lack of flexibility in the model which prevents it from capturing more of the range of variability. Another cause may be shrinkage: this is an effect in which any model will, on average, predict less variation than is present in the observed data (Copas 1983; Greenland 2000). This prevents the model from capturing the *full* range of variation in pattern features in the data, but it is possible that some changes to the model specification may help to capture *more* of the variation.

Firstly, the model included no x-axis random variation and relied upon the flexibility of the underlying b-spline basis to capture variation between observational units in the x-axis. There are forms of multilevel models that include an x-axis random effect. For example, the Super Imposition by Translation And Rotation (SITAR) model is a form of

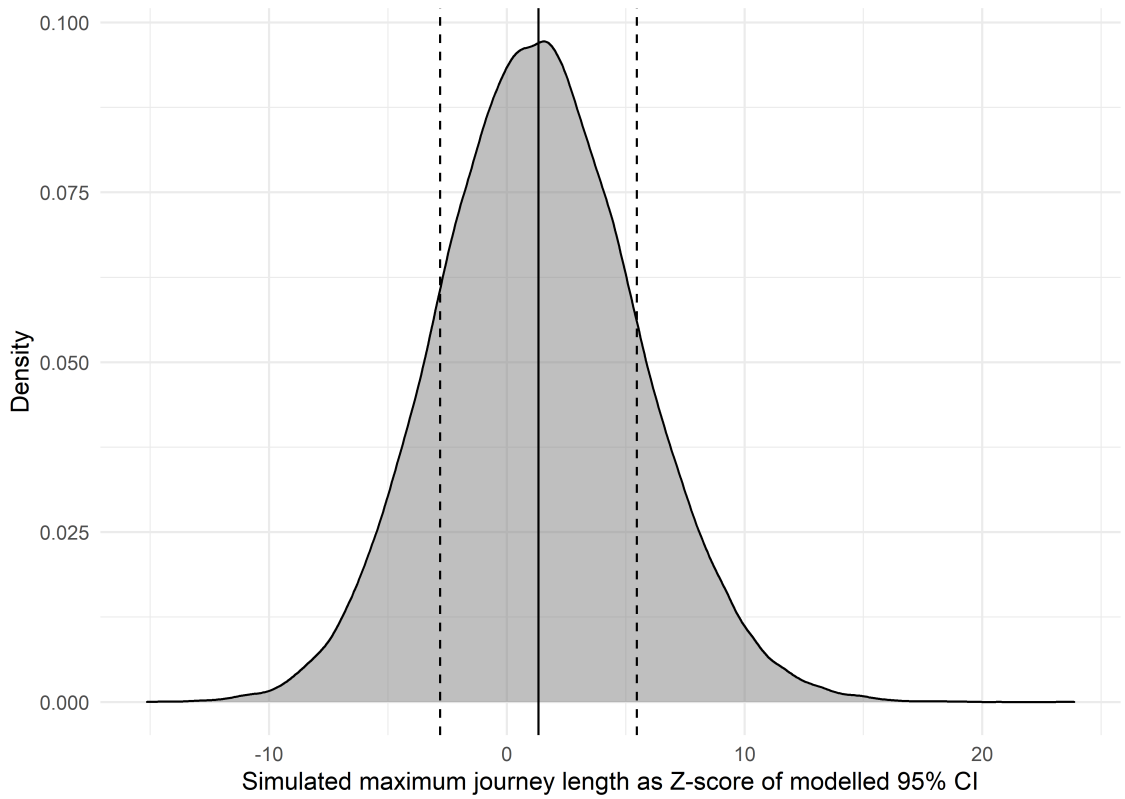


Figure 8.9: Density plot showing simulated maximum journey length values as a z-score of the relevant posterior distributions of estimates for maximum journey length. The solid line indicates the mean z-score and the dashed lines one standard deviation either side of the mean.

multilevel model explicitly designed to investigate childhood growth curves (Cole et al. 2010). It includes three random effects: a y-axis random intercept changing the mean size of a child across the time measured, an x-axis random intercept representing changes in the timing of when a growth spurt occurs, and a random effect for the velocity of growth. Incorporating an x-axis effect to this model may help it to capture more of the variation in maximum time values by allowing horizontal translations of the modelled patterns between observations units – this reflects the simulation process underlying the data.

Including random effects in the x-axis may present some difficulties in fitting the model due to increased complexity. Currently, the model code treats the b-spline basis as a series of fixed variables, the value of which at each observation time is calculated before estimating the model. This is convenient and represents an advantage of using b-splines over penalised alternatives. However, if an x-axis random intercept were included, the b-spline basis would be shifted horizontally for each edge in the network by an amount estimated by the model. This means that the value of each b-spline basis function at each

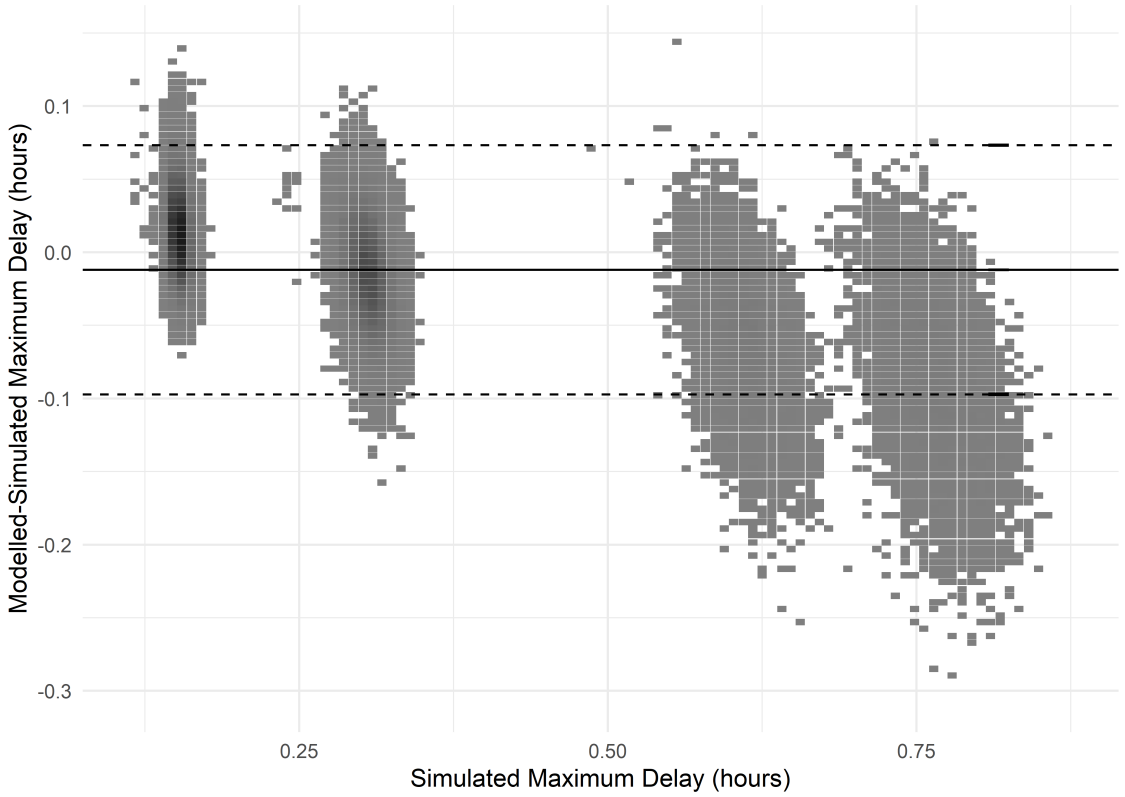


Figure 8.10: Bland-Altman plot showing the relationship between simulated maximum delay and modelled minus simulated maximum delay. The density of points is indicated by the colour of squares in the plot (darker grey equals greater density). The solid line indicates the mean bias and dashed lines indicated 95% limits of agreement.

observation time would depend on a quantity estimated in the model (the x-axis random effect) and could not be calculated before fitting. Incorporating such a random effect would likely be possible in OpenBUGS but would be complicated – this would probably increase computational intensity of the models and make them more difficult to construct.

However, it may be useful to consider alternative model specifications, particularly for the specification of random effects, that may allow an x-axis random effect to be included without increased complexity. In this simulation, spatial and aspatial random variation between observational units was incorporated into the coefficients for each b-spline function in the model. This aimed to allow variation in the form of the temporal patterns of journey lengths for each observational unit while allowing observations closer together to be more similar. This requires estimation of a large number of random effects coefficients. The coefficients for the b-splines, however, do not have a clear real-world interpretation (Stimson et al. 1978) so perhaps the incorporation of random effects in these coefficients does not correspond completely to random variation in the temporal patterns, particu-

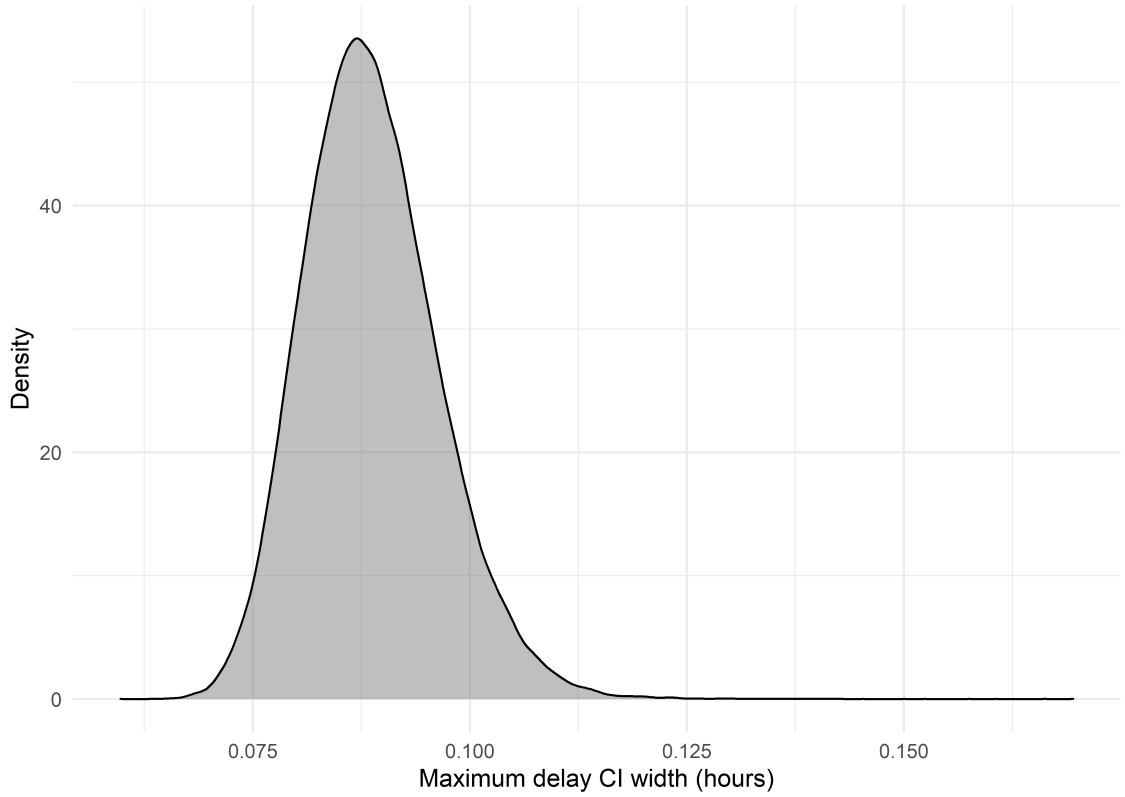


Figure 8.11: Density plot showing the distribution of 95% credible interval (CI) widths for maximum delay estimates.

larly spatial random variation, in the way we would expect. It may be more useful and interpretable to fix the b-spline coefficients and include a series of random effects to allow transformations of the average temporal pattern for each observational unit, for example including a random y-axis intercept and y-axis scaling factor, similar to the SITAR model (Cole et al. 2010). This would reduce the number of random effects that needed to be estimated by the model, allowing for additional x-axis random effects to be included without increasing model complexity, and also mean that the interpretation of these random effects was clearer – for example, similar coefficients for two observational units would clearly correspond to similarity in the temporal patterns of journey delays.

This approach, however, may have some shortcomings. The multilevel model in this paper assumes that the random coefficients for each b-spline are normally distributed (Besag, York, et al. 1991; Goldstein 2011). As the simulation has incorporated some latent or underlying groups in the data with different journey delays, the distributions of simulated maximum delay and maximum journey length are not normally distributed. As the model used in this example does not directly estimate the random variation in these parameters,

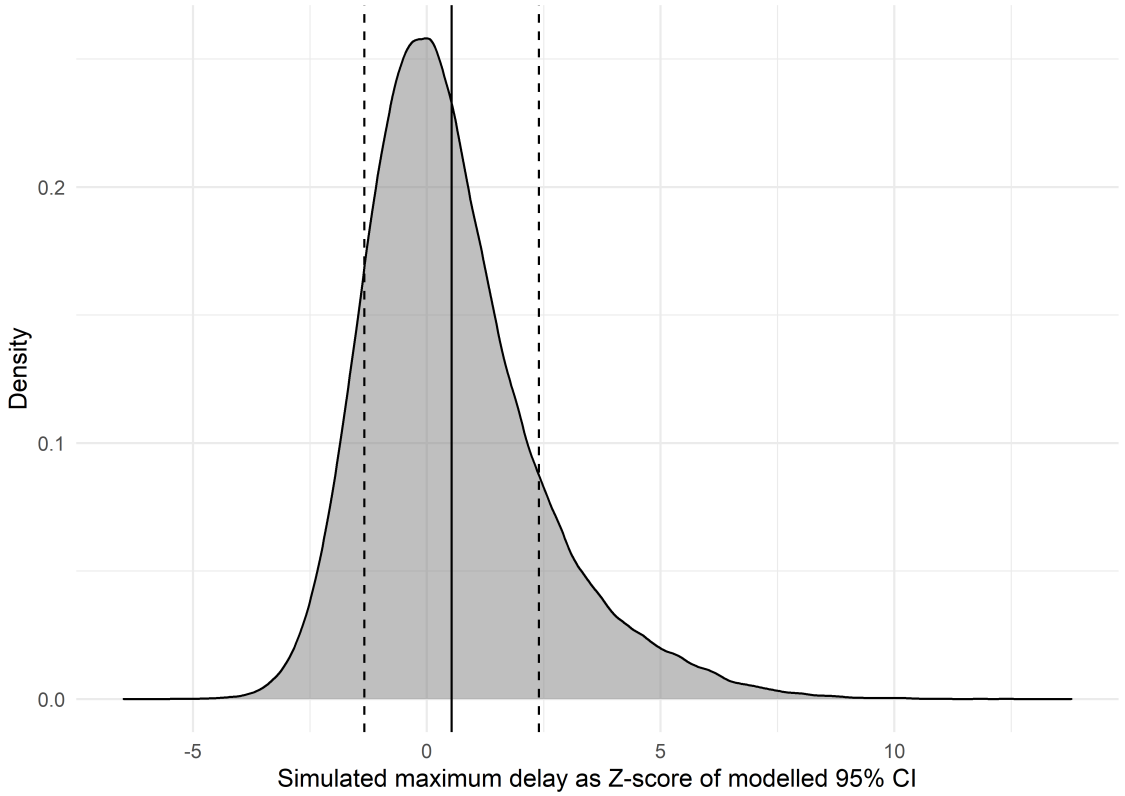


Figure 8.12: Density plot showing simulated maximum delay values as a z-score of the relevant posterior distributions of estimates for maximum delay. The solid line indicates the mean z-score and the dashed lines one standard deviation either side of the mean.

but rather random variation in the 'abstract' b-spline coefficients, this does not necessarily violate this assumption (though it may). However, if random effects more directly related to these pattern features were included this may more clearly violate these assumptions and cause the model to perform poorly (Besag, York, et al. 1991; Goldstein 2011). It may be possible to account for this, for example by using models with latent (unmeasured) variables included (Kaplan et al. 2009). It is important for researchers to examine data carefully to see if there are any structures like this. This allows the researcher to attempt to account for this where necessary.

The mean width of the credible intervals estimated for maximum time were particularly large in relation to the 24-hour period covered by the data. The majority of credible intervals in Figure 8.5 are also fairly wide, which means that the estimates from these models often do not give a particularly precise indication of maximum time. This may not be adequate if the models were to be used in a real situation, for example, to choose a time to implement measures to reduce congestion in stations, or times to include more trains in a timetable. The distribution of credible interval widths in Figure 8.5, unlike those

for maximum journey length and maximum delay, has a series of local maxima at fairly regular intervals indicating that there are some credible interval widths that are more common than others. The distance between these local maxima appears to be similar to the distance between knot points in the b-spline basis specified for this model (1.2 hours, see Supplementary Figure F.1). This may indicate that the structure of the temporal patterns estimated by the models strongly reflects the b-spline knot placements and that the shape of these patterns is influencing the width of posterior distributions that generate credible interval estimates. It would be useful for future work to investigate if this is the case and if alternative model specifications might change this pattern.

In this particular simulation, the simulated pattern has a clear maximum in the 24-hour period of interest that is far greater than the other local maxima included. However, it is possible that some temporal patterns may include multiple maxima with a smaller difference in their values, for example, patterns that include multiple days and have a periodic nature. It may be that, for patterns with multiple local maxima, the posterior distribution of maximum time estimates would contain more than one of these maxima, rather than the maximum with the largest value during the 24-hour period. This may produce a multimodal posterior distribution of maximum times which will result in a much wider credible interval than if the distribution is concentrated around a single maximum in the 24-hour period. While the usual interpretation of a credible interval based on a multimodal posterior distribution of maximum time would remain accurate, a 95% credible interval does not give such a precise description of the posterior distribution in this case and it may be assumed by readers that it is unimodal. It may be useful for future work to investigate if a multimodal posterior distribution for the maximum time pattern feature might be produced in certain situations and whether there is a more detailed way to represent uncertainty when this is the case.

For all the pattern features, even maximum time which had wide credible intervals, the coverage of credible intervals is far lower than the expected 95%. Coverage is lowest for maximum journey length and highest for maximum delay. This indicates that the estimates of uncertainty for the pattern features are not reliable. This could suggest that the estimated credible intervals are too narrow, or the estimates of maximum journey length are biased (or a combination of these). There is no clear indication here that the

estimated credible intervals are too narrow – in fact, relative to the amount of variation in the real maximum journey length, the credible intervals are relatively wide – but we cannot rule this out as a cause of the undercoverage of these estimates. Despite the mean bias estimates being low, there is clearly some bias in the estimation of low and high values of each pattern feature. This may have contributed to the size of the 95% limits of agreement for these features which are larger than half the mean credible interval width for all three pattern features. The size of the 95% limits of agreement in relation to the credible interval widths means that we would expect more than 5% of the credible intervals not to cover the simulated maximum journey length, which is indeed the case. Where the credible interval widths are smaller compared to the 95% limits of agreement, we also see a decreased coverage. This may indicate that bias does contribute to the undercoverage of credible intervals. Further work could be carried out to investigate the contributions of different factors to the undercoverage seen in this simulation study. This would indicate how best to approach fixing the problem.

8.6 Conclusion

This simulation tested the ability of continuous-time multilevel models to capture maxima in temporal patterns of edge properties for a spatio-temporal network. Only one scenario with one data structure was investigated – this particular scenario was chosen to reflect a real data example to which this model has previously been applied (Gadd, Comber, et al. 2021). Future work testing these models in the presence of different data structures (for example, the timing of measurements (evenly spaced, balanced/unbalanced, measured in waves), different degrees of temporal and spatial autocorrelation) should be carried out as the impact of different scenarios on the model performance should be assessed.

The models were able to capture the maximum time and maximum journey length with a small *mean* bias; they tended to overestimate low simulated values and underestimate high simulated values for both features, leading to biased estimates for individual edges in the network with low or high maximum times/journey lengths. The extent to which variation is not fully captured may have contributed to the relatively small proportion of credible intervals for these pattern features that contain the simulated pattern feature values. In practice, this means the credible intervals from this model in its current form cannot

be interpreted as the range of most likely values for the pattern features. Future work concerning continuous-time multilevel models for temporal patterns of edge properties in spatio-temporal networks should consider incorporation of an x-axis random intercept and alternative prior and initial value specifications, which could potentially improve this.

References

Abdelghany, A., Mahmassani, H., & Chiu, Y. (2001). Spatial microassignment of travel demand with activity trip chains. *Transportation Research Record*, *1777*(1), 36–46.

Anderson, T., & Dragićević, S. (2018). A geographic network automata approach for modeling dynamic ecological systems. *Geographical Analysis*, *52*, 3–27.

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), 1–20.

Bland, J., & Altman, D. (1999). Measuring agreement in method comparison studies. *Stat Methods Med Res*, *8*(2), 135–60.

Blonder, B., Wey, T., Dornhaus, A., James, R., & Sih, A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, *3*(6), 958–972.

Chen, S., Claramunt, C., & Ray, C. (2014). A spatio-temporal modelling approach for the study of the connectivity and accessibility of the Guangzhou metropolitan network. *Journal of Transport Geography*, *36*, 12–23.

Cheng, T., Wang, J., Haworth, J., Heydecker, B., & Chow, A. (2011). Modelling dynamic space-time autocorrelations of urban transport network. *Proceedings of the 11th international conference on Geocomputation 2011*, 215–210.

Cheng, T., Wang, J., Haworth, J., Heydecker, B., & Chow, A. (2014). A dynamic spatial weight matrix and localized space–time autoregressive integrated moving average for network modeling. *J Geographical Analysis*, *46*(1), 75–97.

Colak, S., Schneider, C., Wang, P., & Gonzalez, M. (2013). On the role of spatial dynamics and topology on network flows. *New Journal of Physics*, *15*, 113037.

Cole, T., Donaldson, M., & Ben-Shlomo, Y. (2010). SITAR - A useful instrument for growth curve analysis. *International Journal of Epidemiology*, *39*(6), 1558–1566.

Copas, J. (1983). Regression, prediction and shrinkage. *45*(3), 311–335.

Dubin, R. (1998). Spatial autocorrelation: A primer. *Journal of Housing Economics*, *7*(4), 304–327.

Freeman, J. (1989). Systematic sampling, temporal aggregation, and the study of political relationships. *Political Analysis*, *1*, 61–98.

Gadd, S., Comber, A., Gilthorpe, M., Suchak, K., & Heppenstall, A. (2021). Simplifying the interpretation of continuous-time models for spatio-temporal networks. *Journal of Geographical Systems*.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721–741.

Goldstein, H. (2011). *Multilevel statistical models* (Fourth edition). Wiley.

Goldstein, H., Healy, M., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *J Statistics in medicine*, *13*(16), 1643–1655.

Greenland, S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology*, *29*(1), 158–167.

Head, M., Holman, L., Lanfear, R., Kahn, A., & Jennions, M. (2015). The extent and consequences of P-hacking in science. *PLOS Biology*, *13*(3), e1002106.

Heck, R., & Thomas, S. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus* (Third edition). Taylor & Francis.

Holmes, C., & Mallick, B. (2003). Generalized nonlinear modeling with multivariate free-knot regression splines. *Journal of the American Statistical Association*, *98*(462), 352–368.

Howe, L., Tilling, K., Matijasevich, A., Petherick, E., Santos, A., Fairley, L., Wright, J., Santos, I., Barros, A., Martin, R., Kramer, M., Bogdanovich, N., Matush, L., Barros, H., & Lawlor, D. (2013). Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research*, *25*(5), 1854–1874.

Hwang, S. (2000). The effects of systematic sampling and temporal aggregation on discrete time long memory processes and their finite sample properties. *Econometric Theory*, *16*(3), 347–372.

Kaplan, D., Kim, J.-S., & Kim, S.-Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology*. SAGE Publications Ltd.

Li, T., & Liao, Q. (2016). Dynamic networks analysis and visualization through spatiotemporal link segmentation. *IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 209–214.

Li, X., & Griffin, W. (2013). Using ESDA with social weights to analyze spatial and social patterns of preschool children's behavior. *Applied Geography*, *43*, 67–80.

Liu, S., Wang, L., Liu, C., & Destech Publicat, I. (2018). Visualized social network analysis on spatial dynamics of international trade between China and League of Arab States. *2018 3rd international conference on computational modeling, simulation and applied mathematics* (pp. 270–277).

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine*, *28*, 3049–3067.

Morris, T., White, I., & Crowther, M. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102.

Muthén, L., & Muthén, B. (2010). *Mplus user's guide* (Sixth edition).

Neeson, T., Wiley, M., Adlerstein, S., & Riolo, R. (2012). How river network structure and habitat availability shape the spatial dynamics of larval sea lampreys. *Ecological Modelling*, *226*, 62–70.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256.

Newman, M. (2018). Fundamentals of network theory. *Networks* (Second edition). Oxford University Press.

Niezink, N., Snijders, T., & van Duijn, M. (2019). No longer discrete: Modeling the dynamics of social networks and continuous behavior. *Sociological Methodology*, *49*(1), 295–340.

Oud, J., Folmer, H., Patuelli, R., & Nijkamp, P. (2012). Continuous-time modeling with spatial dependence. *Geographical Analysis*, *44*(1), 29–46.

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. *BMC Medical Research Methodology*, *19*(1), 46.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*, *124*.

R Core Team. (2020a). R: A language and environment for statistical computing.

Ramsay, J., & Silverman, B. (1997). *Functional data analysis*. Springer.

Stan Development Team. (2019). RStan: The R interface to stan. R package version 2.19.2.

Stimson, J., Carmines, E., & Zeller, R. (1978). Interpreting polynomial regression. *Sociological Methods & Research*, *6*(4), 515–524.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, *12*(3), 1–16.

Thomas, A., Best, N., Lunn, D., Arnold, R., & Spiegelhalter, D. (2014). GeoBUGS user manual.

Transport for London. (2009a). London Underground map December 2009.

Transport for London. (2009b). London Underground map September 2009.

Weiss, A. (1984). Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics*, *26*(3), 271–281.

Yang, Y., Heppenstall, A., Turner, A., & Comber, A. (2019). A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. *Computers, Environment and Urban Systems*, *77*, 101361.

Yeh, R., Nashed, Y., Peterka, T., & Tricoche, X. (2020). Fast automatic knot placement method for accurate b-spline curve fitting. *Computer-Aided Design*, *128*, 102905.

Yuan, Y., Chen, N., & Zhou, S. (2013). Adaptive b-spline knot selection using multi-resolution basis set. *IIE Transactions*, *45*(12), 1263–1277.

Chapter 9

Discussion

9.1 Introduction

This research has explored the application of continuous-time models to examine temporal patterns of edge properties in spatio-temporal networks.

Initially, a review was undertaken to identify methods used to analyse spatio-temporal networks in the recent literature, with a particular focus on the specification of time as continuous or discrete. Almost all the papers identified in this review used discrete time specifications, which showed a clear need for investigation into the use of continuous-time methods for spatio-temporal networks and the challenges of making results from spatio-temporal network analyses interpretable and accessible to readers.

Purported continuous-time methods established in geography and epidemiology were critically reviewed in Chapter 3. This chapter identified that a number of these methods are not as adept at handling continuous time as they initially seem, particularly the most flexible form of latent growth curve models (Bollen and Curran 2005). This chapter also highlighted a potential pitfall of temporal data analysis – the practice of conditioning on the outcome. Little detailed information regarding the issue was found in the literature and this made clear the need to highlight this issue to discourage people from using this kind of methodology (Tu, Tilling, et al. 2013). Three potentially useful continuous-time methods were identified: multilevel models, functional data analysis, and stochastic differential equations. Of these, this thesis focused on multilevel models because of the ease with which they capture complex covariance structures and the wide availability of soft-

ware for fitting these complex models (Besag, York, et al. 1991; Browne, Draper, et al. 2002; R. Jones and Boadi-Boateng 1991).

Chapter 4 outlined in detail the multilevel model specification whose application to spatio-temporal network data would be investigated. It also outlined the simplification of results from these models by the extraction of 'pattern features'. Pattern features represent events that can be defined by some part of the temporal pattern, for example, maxima or minima. In addition, the use of these models to generate continuous temporal functions of network properties in the presence of data with measurement interval heterogeneity was outlined. The process outlined in Chapter 4 is, to the best of the author's knowledge, new to the geography literature, particularly the generation of continuous-time functions of network properties. This use of continuous-time multilevel models also fills a gap in the literature that was identified in Chapter 2. Work containing these models (Chapter 7) has been accepted for publication as Gadd, Comber, et al. (2021).

Chapter 5 described the data used in the investigation of these methods. Real data from Transport for London Open Data were introduced, but the focus was the development of a flexible and widely applicable simulation tool for longitudinal data. Simulating data can be a complex and difficult process, and the aim of developing such a function was to provide a relatively simple tool through which longitudinal data could be simulated with a flexible form and data structure, with known parameters at the population and individual level, and a pre-specified spatial and/or network dimension. This tool has been made freely available online and is a useful contribution to the field of temporal or longitudinal data analysis in which simulations can be used to evaluate methodology.

Chapter 6 included an investigation into the inappropriate interpretation of models and methods that aim to relate temporal patterns of a variable for individual observational units to a later, time-invariant outcome. The methods in question attempted to model and/or interpret temporal patterns conditional on this outcome, which resulted in inappropriate inferences of average values of the temporal variable within outcome groups at different time points as a continuous temporal pattern. This was included because, in an exploration of the use of continuous-time models to model temporal patterns at an individual level, many studies used these methods, particularly within the area of lifecourse epidemiology (Færch et al. 2013; Hulsegge et al. 2017; Oka et al. 2017; Sabia et al. 2017;

Shoaff et al. 2017). Little detailed information about this issue has been published to date, so a simulated example outlining some of the potential shortcomings and misunderstandings that can arise when aiming to model continuous temporal patterns was considered immportant to encourage researchers to avoid this methodological error when applying continuous-time models to spatio-temporal networks. To avoid the issue, models aiming to capture individual observational unit level temporal patterns should be fitted agnostic to any later outcome. Results from these models, including temporal pattern features extracted as in Chapters 7 and 8, can then be related to later outcomes. Producing a detailed demonstration of the issues associated with methods that condition on the outcome is an important contribution to research involving temporal models in any field and has been published as Gadd, Tennant, et al. (2019).

Chapter 7 provided an example application of the novel analysis approach outlined in Chapter 4 to both simulated and real data examples. In the examples described, the models were applied to explore temporal patterns of delays to journeys between different origin-destination station pairs in a railway network. More generally these models could be applied to any situation where mapping or correlating events defined by edge or node property fluctuations in a network with static topology is of interest. For example, the timing and size of rapid changes in river levels and their relation to weather events in different locations of the river basin, or the timing of congestion in different parts of a road traffic network in response to a road blockage (due to an accident, for example) in different locations.

Chapter 8 included a simulation analysis testing the accuracy and precision with which the process in Chapter 4 was able to extract features of temporal patterns of network properties. This identified some shortcomings of the novel approach and provided an indication of some directions for future research.

9.2 Review of spatio-temporal network methods

Chapter 2 reviewed literature that analysed spatio-temporal networks. Spatio-temporal networks were found in a wide range of different fields, which highlighted the importance of suitable analytical methods for these structures. However, a key finding from the review was that all but four of the eighty plus papers specified time as a discrete variable

(Abdelghany et al. 2001; Y. Fang et al. 2019; Y. Jin et al. 2019; Lengyel et al. 2020). There are a range of issues associated with using discrete-time models, not least that they do not represent the true, continuous nature of time (Oud et al. 2012). The use of discrete time steps or windows means that data must either be measured regularly and simultaneously for all observational units or must be aggregated – the process of aggregation has been shown to introduce bias (de Haan-Rietdijk et al. 2017; Hwang 2000). In addition, the size of the time steps is a key choice for discrete-time models that must be made based on knowledge of the process being analysed (Comber and Wulder 2019) – however, when data are measured regularly, the length of time steps is often predetermined (J. Freeman 1989). Too long or short a time-step can mask variation in temporal properties (J. Freeman 1989; Rossana and Seater 1995).

As discrete-time models are subject to this range of issues, the lack of use of continuous-time models highlighted a key area for further research – the use of continuous-time models in spatio-temporal networks. This was the key focus of research for this project.

Another additional finding in the review was the difficulties associated with visualising spatio-temporal network data. Presenting spatio-temporal network information in an interpretable way is difficult. Even static spatial network graphs often suffer from overcrowding of edges or nodes which can obscure important information (Graser et al. 2017; Wood et al. 2010); introducing a further dimension and more information to this only exacerbates the problem. A number of papers aimed to develop new methods or software for visualising spatio-temporal networks, but the resulting solutions were not ideal and sometimes unclear (T. Li and Liao 2016; G. Sun et al. 2017). One spatio-temporal visualisation did produce a clearly interpretable map in which the timing of key events, extracted from a spatio-temporal network model, for different edges in the network was presented (Colak et al. 2013). This was based on a discrete-time model, but did highlight the importance of simplification of results and focus on a key outcome of interest as ways to convey information about complex structures clearly.

9.3 Continuous-time models

Chapter 3 provided an overview and critical discussion of a range of continuous-time models and methods that could potentially be applied to spatio-temporal networks. This

drew on methods not just from the field of geography, but also on well-established methods in the field of lifecourse epidemiology.

A number of the methods identified in this chapter are presented as being continuous-time methods but are actually limited to treating time as a discrete variable. One such method was the latent-basis latent growth curve model – this is a form of latent growth curve model where the factor loadings (used to specify the relationship between a latent slope variable and observed measures of some temporal property, usually set to be equal to or some function of the measurement times) are allowed to be estimated freely (Bollen and Curran 2005). This essentially stretches the time-axis of the model allowing latent-basis latent growth curve models to fit non-parametric shapes. However, the amount of stretching of this axis is only estimated at the measurement times for the observed variable, not in between. This means that the latent-basis latent growth curve model is undefined in between observed measurements which is not conducive to producing a continuous temporal model of some property.

Another methodology that does not treat time as a continuous variable is conditioning on the outcome – this was identified as a common pitfall in which analyses included their outcome variable as a covariate when attempting to examine how the patterns of a temporal exposure variable may cause it. The aim is to identify 'average' patterns, but the result is a series of average values within each group at discrete time points that cannot be joined and called a 'pattern' or 'trajectory'. There was limited mention of this issue in the literature, despite it being a fairly prevalent problem, which highlighted a clear need for this issue to be addressed in detail to discourage further use of conditioning on the outcome methods (Tu, Tilling, et al. 2013).

There were three methods identified in Chapter 3 that could be suitable for application to spatio-temporal networks: multilevel models, functional data analysis and stochastic differential equations (Goldstein 2011; Ramsay and Silverman 1997; Van Kampen 1976). For this investigation, multilevel models were chosen due to the wider familiarity and availability of software for multilevel modelling that can also incorporate both spatial and temporal covariance structures.

Capturing the variance structures underlying data generation is important when modelling to seek causal understanding. Many standard statistical models make the default

assumption of independence of observations (Wolf and Best 2014a). In many situations, including when we have repeated measures of the same observational units, this will not hold; repeated measures taken within the same observational units are likely to be more similar to each other than measures made for other observational units (Goldstein et al. 1994). Failure to account for this will result in inappropriate estimates of uncertainty (e.g. credible or confidence intervals) from analyses (Ugrinowitsch et al. 2004). Estimates of uncertainty are a key part of the results from an analysis – they provide an indication of a range of probable values for an effect size and can help us understand if that effect may be of importance in the real world.

Additionally, results describing the complex variance structure underlying data generation may be of interest to researchers to address specific research questions that cannot be elicited from the fixed effects alone – for example, the extent of spatial correlation between patterns of a variable in observational units may be a key outcome of an investigation. However, estimates of uncertainty and results describing the covariance structure underlying data will only be useful to researchers if this covariance structure is captured accurately by the model. This is a complex task – researchers must carefully consider the data generation processes and knowledge of the system of interest and investigate different models for covariance structures that may be appropriate (Gilthorpe et al. 2014).

However, it must be acknowledged that multilevel models explicitly describe temporal patterns as a function of time and, in many contexts, this is not an accurate depiction of the causal relationships underlying the data generation process. Time itself does not cause temporal fluctuations in a variable, but rather indexes the underlying processes that do cause fluctuations in that variable (Oud et al. 2012). Describing temporal patterns of a variable as a function of time is a simplified representation that is used to consistently extract temporal features of interest across observational units. An alternative approach that uses a continuous representation of time is the stochastic differential equation. Stochastic differential equations describe the rate of change of a variable at a given time as a function of the observed value of the variable (and potentially other variables) at that time (Van Kampen 1976). This uses time as an index, to order the variable of interest over time, rather than as the cause of the variable of interest. The relationship described is that between the current value of the variable and the rate of change at each instant in

time, which may be a more appropriate description of the causal processes underlying the data.

However, extracting pattern features in the manner presented in Chapters 7 and 8 is essentially a process of interpolation (aiming to simplify interpretation). The specification of the temporal pattern of the variable of interest as a function of time is used in the process of extracting pattern features, but is not necessarily of importance to their interpretation – whether extracted by a multilevel model or a stochastic differential equation, the timing of a maximum point represents the same 'event'. When simplification is the goal, we must ask whether it is necessary to attempt to describe the causal relationships underlying temporal pattern generation when they are not necessarily of interest as a result. On the other hand, hypothesised causal relationships between variables often form an important basis to developing predictive models (Piccininni et al. 2020). If one variable is an important cause of another, they are likely to be highly correlated; having information on one of two highly correlated variables is likely to improve predictions of the other. This is likely to also be the case when interpolating within the range of the data, rather than aiming to generate out of sample predictions. Attempting to model the causal relationships underlying the generation of temporal patterns in data could potentially lead to more accurate recovery of pattern features. These two approaches (modelling the data generation processes and modelling temporal patterns as a function of time) should be compared in future work.

While it would be interesting to explore the use of stochastic differential equations to carry out analyses as described in Chapters 7 and 8, there are alternatives to multilevel models that also define temporal patterns of variables as a function of time. One method that was identified as potentially useful in Chapter 3 was functional data analysis (Ramsay and Silverman 1997). Functional data analysis can account for covariance between residuals, for example temporally autoregressive error structures, and has been used in spatial applications (Delicado et al. 2010). Functional data analysis also has its main aim in recovering temporal patterns which is similar to the process of extracting temporal pattern features outlined in Chapter 4. As part of this, it includes a process called registration which involves aligning the features of these temporal patterns (Ramsay and Silverman 1997) – this is, in a sense, akin to including a random x-axis intercept and could be very useful in extracting temporal pattern features more accurately. The application of functional data

analysis to spatio-temporal networks should also be explored in future work.

9.4 Continuous-time model specification for spatio-temporal network data

Chapter 4 introduced the specification of a continuous-time multilevel model that could be used to capture temporal patterns of edge properties in a spatial network and a process for simplifying the results to present them in an interpretable way. The process for simplifying results included the extraction of pattern features, for example maxima, from the temporal functions similar to Colak et al. (2013). In addition, a novel process to generate continuous-time functions of whole-network properties was introduced. This method has wide ranging potential applicability for studying the evolution of network structures in continuous-time. Calculating network properties at regular discrete time intervals is a common tactic but it relies on regularly measured data or temporal aggregation of data (Blonder, Wey, et al. 2012). Producing a continuous-time function using continuous-time models enables network properties to be estimated at all points in time, including interpolating properties at times with no observations, in the presence of interval heterogeneity without the process of aggregation (which can introduce bias (de Haan-Rietdijk et al. 2017; Hwang 2000)).

These models could be applied to wide range of circumstances where temporal patterns of edge (or node) properties are of interest. Chapter 7 illustrates an application in the transport sector. In this scenario, the timing and extent of the maximum delays to journeys recorded for each origin-destination pair in a rail network could be used to identify where and when to add new rail services to the network timetable in order to reduce delays as much as possible for the money expended. Other applications could be considered in other areas, for example, examining the extent and timing of changes in river levels for different reaches of a river network in response to a rainfall event in the river basin, which could be used to inform safety warning systems for those taking part in river-based leisure activities.

9.4.1 Advantages

There are two major advantages of the framework outlined in Chapter 4, regardless of the type of continuous-time model used to capture temporal patterns. First, the models

treat time as a continuous variable and, second, the improved interpretability of model results through the extraction of temporal pattern features and continuous whole-network property functions. These two factors are heavily linked to each other – the interpolation in continuous time is necessary to extract pattern features and continuous whole-network property functions would not be possible if a discrete-time model was used, as these are only defined at each discrete time step and not in between them. Producing continuous-time temporal functions of whole-network properties would be difficult without the use of continuous-time models of observational unit level properties. If data were measured simultaneously for all observational units within the network being studied, it would be possible to calculate the whole-network property of interest at each time point and then model these results over time. However, if the measurements were not simultaneous, which is likely to be the case in many circumstances, interpolation of the data between measurements at the observational unit level (using models such as that in Chapter 4) is necessary before whole-network properties can be calculated at any time point; data from all observation-units is required to calculate the whole-network properties.

Using a continuous-time model also more accurately represents the real nature of time (Oud et al. 2012). For all practical purposes, time does not advance in discrete steps but can be divided into infinitely small units. While a discrete variable, such as the number of passengers in a tube station, might change at a single instant in time, the world is also full of continuous variables that change continuously over continuous time periods. For example, the speed of a car travelling on roads changes continuously as it accelerates and decelerates, or the temperature inside a London Underground train carriage will change continuously throughout each day rather than instantaneously increasing by discrete units.

As discussed in more detail in Chapters 2 and 3, continuous-time models can also cope with unevenly spaced measurement intervals (within and between observational units) without the aggregation of measurements into time segments. Aggregating measurements over time essentially changes the time value associated with each observation. Many statistical models assume that there is no error in the measurement of covariates, only in the outcome (Wolf and Best 2014c). Introducing error by changing the 'timestamps' of measurements through aggregation clearly violates this assumption. Inappropriate aggregation has been shown to bias inferences from autoregressive models, as discussed in

Chapter 2, and should be avoided (de Haan-Rietdijk et al. 2017; Hwang 2000; Rossana and Seater 1995). Continuous-time models provide a convenient way of circumventing this issue.

The process of simplifying temporal patterns by extracting pattern features is also an important and potentially useful part of this research. Various attempts to visualise spatio-temporal network data were identified in Chapter 2. Those that did not simplify temporal patterns into features were not clear (T. Li and Liao 2016; G. Sun et al. 2017). Including a large quantity of data with at least three dimensions (two-dimensional space, and time) is a huge challenge for visualisation, particularly if avoiding the use of animations and interactive graphs. Visualisations with all these dimensions and associated information are prone to becoming overcrowded and difficult to read and understand, which results in no useful information being transmitted to the reader (Graser et al. 2017; Wood et al. 2010). Extracting pattern features gives us just a small number of pieces of information to map, model or summarise. By condensing a temporal pattern into salient pieces of information, we do not have to display the time dimension in the data which reduces clutter in visualisations. Continuous-time functions of whole-network properties also reduce the dimensions of data we must display – by summarising across the whole network we no longer have to visualise space, which makes these much easier to display and interpret.

One other potential advantage of considering specific pattern features and network properties is the focusing of analyses on specific information and outcomes of interest. This may reduce practices such as p-hacking, which aim to mine the data for a "statistically significant" result. These practices often produce poor quality research with little consideration of the underlying data generating processes (Head et al. 2015; Simmons et al. 2011). If a researcher is focused on a particular feature of interest, like a pattern feature or whole-network property, they can tailor the analysis to appropriately reflect the processes underlying the data generation (Comber and Wulder 2019). Models will never be able to fully reflect these processes and must compromise on the detail they include and the parsimony of the model (Box 1987). Even when modelling two different outcomes within the same system, the processes particularly salient to the generation of these outcomes may differ; using different model specifications for these two purposes may be the best modelling strategy.

9.4.2 Limitations

The novel multilevel model approach introduced in this research thesis has limitations. The model explicitly included spatial and temporal autocorrelation, using the conditional autoregressive prior distribution (spatial) and an autoregressive error structure (temporal) (Besag, York, et al. 1991; Goldstein et al. 1994). However, the model did not include autocorrelation structures related to the network (geodesic) distances between edges (Newman 2002). It did allow edges that shared an origin or destination station to be more similar than those that did not, which allows some correlation of observations related to the network structure by allowing journeys with the same origin or destination to be more alike than journeys without those similarities; as travelling through the origin and destination stations (after tapping into the London Underground system or before tapping out) constitutes part of the journey length recorded, it is likely that journeys sharing an origin or destination station will have some similarity – particularly if congestion is occurring within the station, so this is an important factor that has been accounted for.

However, there are other factors relating to the structure of the London Underground network itself that may be of importance. Firstly, the distance between stations along network edges, which could induce spatial autocorrelation in a similar way to geographical distance, is not accounted for. It would be possible to account for the network distance between stations, for example, by including an additional random effect with a conditional autoregressive prior based on weights describing the network topology (Besag, York, et al. 1991; Ermagun and Levinson 2018). However, this would introduce additional complexity to the model – researchers may wish to consider which model components are of the highest priority and remove some to increase parsimony. This can improve the estimation of the model.

Another potentially important factor is whether journeys share the same London Underground lines and are therefore likely to experience similar delays or train speeds. This would require more detailed work to account for, as there are often several possible routes between most origin and destination stations in rail networks – this is the case for the London Underground network. Identifying the most likely route of travel for an origin-destination pair with a given journey time could potentially include processes such as identifying all different potential routes, estimating travel times, if possible, and account-

ing for closures of certain lines or segments of the network at different times (Gordon et al. 2018; Kwan 2013; Nassir et al. 2015). Accounting for the lines used to complete a journey would require a further multilevel structure to be incorporated into the model – possibly a multiple membership model where observations of journey lengths can belong to one or more different train line clusters. In this type of model, the 'contribution' of each train line cluster to a journey length observation is weighted (with the sum of weights equal to one) – these weights could reflect the proportion of the journey (in distance or time) that is carried out on each contributing train line (Fielding and Goldstein 2006). Exploring the application of a model that accounts for the sharing of Underground lines would be extremely interesting and could potentially provide insight into which lines experience the worst delays and when.

The project only explores the application of multilevel models to systems with static network topology and geography, that is, the configuration of nodes and edges in the network does not change and they cannot move around. This limitation on networks is realistic in many situations. For example, while rail networks can change, mostly they do so through the closure or opening of an old or new station, and the construction or destruction of rail lines. This tends to happen relatively infrequently.

However, in some situations, changes occur more regularly. There may be certain stations that are closed for part of each day, for example. In other types of networks, like social or ecological networks, vertices may move through space over time. For example, if connections between animals are of interest, the animals will be represented by vertices which will regularly move (S. Chen, Ilany, et al. 2015). In certain places, some river segments flow only during certain seasons or weather conditions, meaning the topology of stream networks can change rapidly in a way that might be important to researchers (Peirce and Lindsay 2015). If topological or geographical changes in networks are considered important to the process under consideration, researchers must try and account for them when modelling the system.

There are discrete-time models for networks that aim to account for temporally dynamic network topology, for example, LSTARIMA (T. Cheng et al. 2014). This is a space-time autoregressive moving average model with a dynamic spatial weight matrix specifying the relationship between edges in a road network that leads to spatial autocorrelation amongst

traffic conditions. While the road network itself is static, it is expected that the effect of one segment on another adjacent one will change depending on the traffic conditions. Thus, traffic conditions at a given time t affect the spatial weight matrix which partially determines traffic conditions at time $t+1$. The discrete-time specification in LSTARIMA enables the relatively easy specification of this 'back-and-forth' relationship between traffic conditions and the spatial weight matrix. It may be possible to accommodate changing network structure in continuous time by jointly modelling the property of interest along with the evolving network structure over time, but this would be challenging (Ibrahim et al. 2010; W. Zhang, Leng, et al. 2015).

As a whole, the model specification used in this work is quite complex: it includes spatial, temporal and network autocorrelation structures and a spline specification to allow time to be treated as a continuous variable. There is scope for the specification of this model to be simplified of further complicated depending on the circumstances in which it was used. Having a more complex model may mean that the covariance structure specified more closely matches the processes underlying the data - this will result in more appropriate estimates of uncertainty and potentially a better fitting model (Goldstein 2011; Wolf and Best 2014a). However, it is not possible to specify a model that completely reproduces all the processes underlying the data - all models are a simplification of reality (Box 1987). Simplifying the model further may reduce estimation times and make the models more widely accessible. However, oversimplification of the covariance structure in the data will lead to underestimation of uncertainty associated with results (Goldstein 2011; Wolf and Best 2014a).

The choice of how complex to make the model specification can only be made based on the research question at hand and knowledge of the processes underlying the data in question, however, there are two complex elements of this model that may be particularly important. First, the cross-classified random structure in which observations are grouped by origin and destination stations. This allows patterns of edge properties to be examined individually for different origin-destination pairs, rather than as an 'average trajectory' (which should be discouraged (Gadd, Tennant, et al. 2019)). The inclusion of this hierarchical structure also accounts for similarity between observations associated with the same origin or destination stations - this is a key component of the network structure of

the data and ignoring this would likely result poor estimation of uncertainty (Goldstein 1994). Second, the continuous-time structure of this model is extremely important in allowing for interpolation of pattern features and continuous-time whole network properties - if these types of measures are useful for the research question, then a continuous-time model should be considered over a discrete-time alternative. However, in situations where this is not the focus, it may be appropriate to consider simpler discrete-time models as an alternative. Researchers should also consider the measurement scheme in the data and if, in the presence of irregularly measured data, aggregation of measurements into time windows for discrete-time models may introduce bias (J. Freeman 1989).

9.4.3 Future work

Overall, there are a range of different model types and model specifications that could potentially form an important part of the framework for analysing spatio-temporal networks and interpreting the results from these analyses set out in Chapter 4. Future work should aim to compare the utility and performance of stochastic differential equations and functional data analysis with each other and multilevel models in a range of different situations. Additionally, investigating different specifications of models for the covariance structure in models should be considered: like any model we cannot expect to fully capture the covariance structure in the data and we must consider and investigate to what extent we must incorporate this complexity into our model in order to generate accurate results with appropriate estimates of uncertainty (Box 1987; Gilthorpe et al. 2014).

Further aspects of model specification that should be investigated are different spatial and temporal weight specifications and alternative spline basis functions. The model in Chapters 4, 7 and 8 uses inverse weighting schemes for temporal autocorrelation processes. This means that the strength of the relationship between two locations or times is inversely proportional to the distance between them (in space or time) (Dubin 2009). One of the major issues with this type of weighting scheme is that the inverse weight ($\frac{1}{distance}$) is undefined when the distance is equal to zero. While this is less likely to be an issue in space, repeated measures in time may occur simultaneously, especially if they occur when an event such as the start of a rail journey takes place. To attempt to account for this, the model simply added a constant (one) to the temporal distances between observations – preserving the differences between them but meaning all inverse weights were defined.

This approach was chosen after experimentation with different weighting schemes as it lead to improved model convergence, but is probably not ideal, and alternative weighting schemes may provide better model performance. For example, the negative exponential weighting system which is mentioned in Chapter 4 (Dubin 2009).

Even with an alternative weighting scheme, there may still be issues with simultaneous measurements – particularly if more than two measurements occur simultaneously, as there is no clear ordering of these measurements in the temporal autocorrelation structure. Under a first order temporal autocorrelation structure, each measurement only influences a single subsequent measurement. This means that if three simultaneous measurements occur, their bivariate correlations would not be equal (arbitrarily ordered, the 'first' measurement will be more strongly correlated with the 'second' than the 'third'). It would probably be expected that three measurements occurring simultaneously would be equally correlated with one another, and this type of autoregressive structure is not able to reflect that. Future work should compare model performance, as well as computation time and ease of estimation, under different weighting schemes. However, the weighting scheme in any model is a parameterisation choice that should be considered not only in terms of performance but also knowledge about the underlying data generation process (Comber and Wulder 2019).

9.5 Simulated data

Chapter 5 outlined the data used in this project. This included real data from Transport for London Open Data, but the major contribution of this chapter is the longitudinal simulation tool which was used in Chapters 7 and 8. This tool is a highly flexible R function for generating repeated measurements of temporal data which can be associated with a network, locations or origin-destination pairs, and grouped into clusters. The tool also produces known temporal patterns at the individual, cluster, and population level, and can incorporate complex error structures. Simulating complex datasets can be difficult but is often necessary when testing the performance of models or methods. It is also important for the user to know what they have simulated. Therefore, this simulation tool is potentially very useful to researchers looking to simulate longitudinal data when evaluating methods, as it removes some of the time commitment to learning to generate

the complex structures it can produce.

It is important to note that simulation studies have some shortcomings. In particular, they can only evaluate the performance of analytical techniques in a specific scenario (Morris et al. 2019). As only one simulated scenario was included in Chapter 8, we only have information about the performance of the model in these particular circumstances. It is of importance that future work addresses a range of different scenarios, testing the ability of models like that in Chapter 4 to capture different pattern features (step changes, rates of change, volume of rush hour traffic) from various different pattern shapes (multiple maxima, different amplitudes, varying complexity) with different data structures (auto-correlation structures, underlying causal relationships, data measurement schemes).

9.6 Conditioning on the outcome

Chapter 3 identified methods that conditioned on the outcome as a common pitfall in temporal data analysis, however very little literature outlining this pitfall and discouraging the use of methods that condition on the outcome could be identified. This highlighted a clear need for a paper emphasising these issues. The key message of Chapter 6 was that, when aiming to look at the relationship between temporal patterns of a property and some later outcome, one should capture information about the patterns agnostic to this outcome and then relate the two, as opposed to examining patterns conditional on the outcome variable.

It is possible that the models in this project could be used to capture pattern features which could then go on to be related to a later outcome in a two-step process. This is an approach that has been used in epidemiology, for example to relate patterns of childhood growth to later experience of skin rashes (Aris et al. 2017; M. Gao et al. 2017; Tilling et al. 2011). In the context of a rail network, such an approach could be used to explore a variety of aspects of how, when, where or why passengers use certain parts of the rail network. For example: to investigate the relationship of origin or destination station locations with peak travel times and peak congestion; to evaluate the effect of measures taken to reduce congestion on the extent of delays to journeys taken from, to or passing through a particular station; or to examine the relationship between passenger perceptions of commuter-period delays and congestion on particular rail lines or stations and the real

delays and congestion on these lines and stations during commuter periods. Studies like this could be used to inform decisions about where and when to prioritise making changes to the rail network infrastructure, rail charges, or public information about the rail network to improve the transport system on a potentially limited budget.

As this approach of relating patterns to a later outcome is a potential use of the methods in this project, it was highly important to demonstrate a major potential pitfall in temporal modelling to discourage it and improve the quality of future research of this type.

9.7 Example application of multilevel models to real and simulated spatio-temporal network data

Chapter 7 included an example application of the methods in Chapter 4 to a real and simulated dataset. This included the application of continuous-time multilevel models, the extraction of pattern features (maximum delays to journeys in a rail network) and the extraction of continuous-time whole-network properties.

The estimation of pattern features and continuous temporal functions for network properties formed a key part of providing interpretable results (with estimates of uncertainty) for these models. The b-spline function on which they are based does not produce interpretable coefficients with a real-world meaning. In addition, the visualisation of temporal patterns within a two (or more) dimensional space proves difficult, leading to cluttered, uninterpretable graphs, or a reliance on animation and interactive graphics which are not accessible to all (Graser et al. 2017; Wood et al. 2010). Chapter 7 includes some visualisations of point estimates of pattern features in relation to network properties of the railway system under analysis, and some visualisations of temporal network property functions. In the opinion of the author, the temporal pattern features and continuous-time network property functions have a clear interpretation with a real-world meaning. The ease of interpretation of a result or measure, however, is subjective – it would be interesting to investigate how easily people from different backgrounds are able to interpret and understand results from analyses like those in this project, for example through a survey.

It is also important to note that visualisation of this information, in particular temporal pattern features, is subject to all the usual fallacies and difficulties of presenting other data.

One aspect of this is the visualisation of uncertainty, particularly with regard to mapping temporal pattern features. Figure 9.1 shows an example map including only the point estimates of maximum flow between nodes in a network, without any indication of the width of the 95% credible interval for this feature. There are some potential solutions to visualising uncertainty presented in Figures 9.2 and 9.3 – Figure 9.2 includes a second map showing credible interval widths which would be presented alongside the map with point estimates; Figure 9.3 includes adding 'triple stripe' lines for each edge in the map, with each line representing the upper limit of the 95% credible interval, the point estimate, and the lower limit . Both solutions have issues: processing two images together can be difficult and including three pieces of information for each edge results in a cluttered map. Even when not attempting to visualise uncertainty alongside pattern feature point estimates, visualising edge properties in networks without cluttering and potentially obscuring useful or important information is particularly difficult, and this approach does not attempt to provide any solutions to this issue (Graser et al. 2017; Wood et al. 2010).

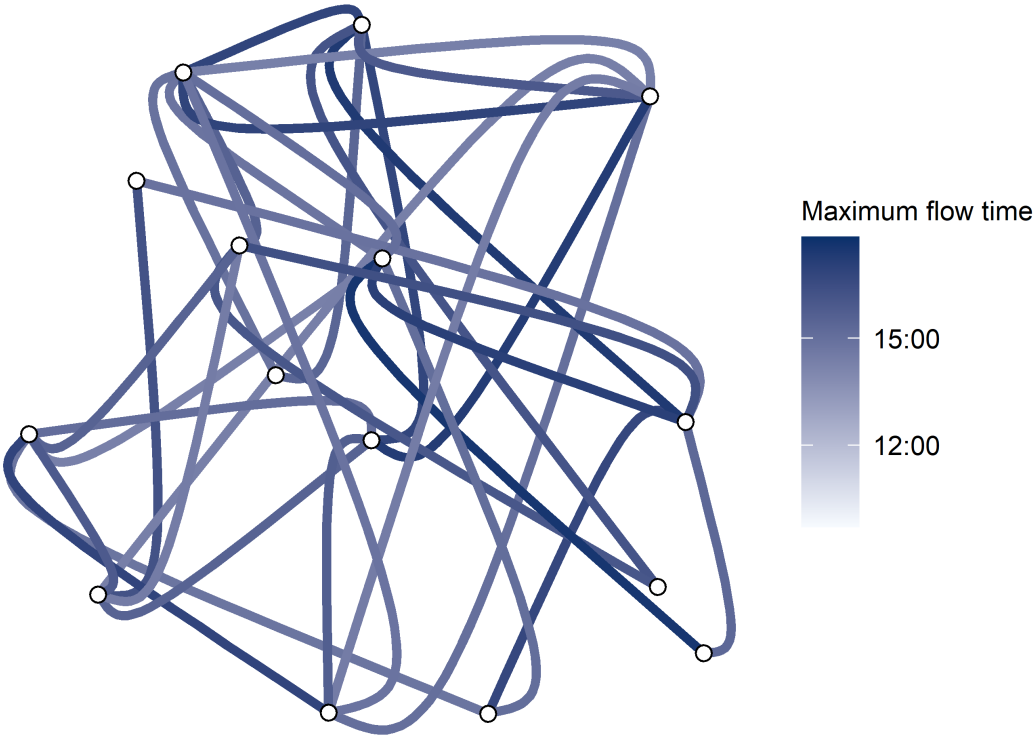


Figure 9.1: Example map of times of maximum flow along edges in a network with no representation of uncertainty

Chapters 4 and 7 also outlined the use of temporal functions of network properties to generate continuous temporal functions of whole-network properties, but this was not further

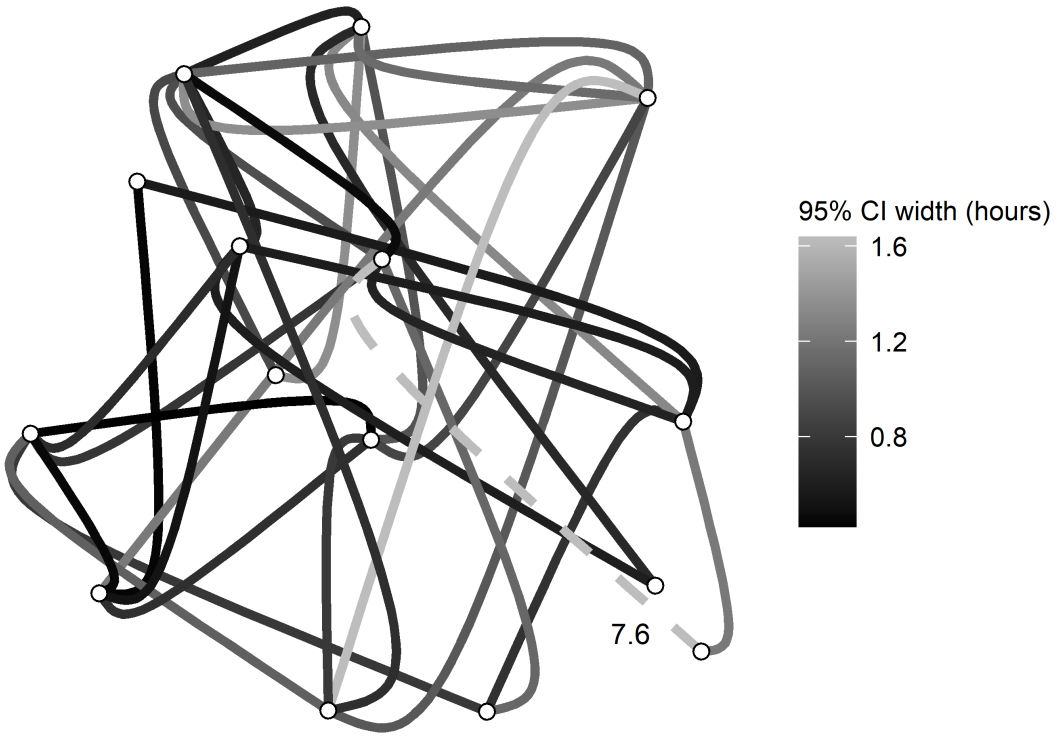


Figure 9.2: Example map showing the width of 95% credible intervals associated with maximum flow times in Figure 9.1

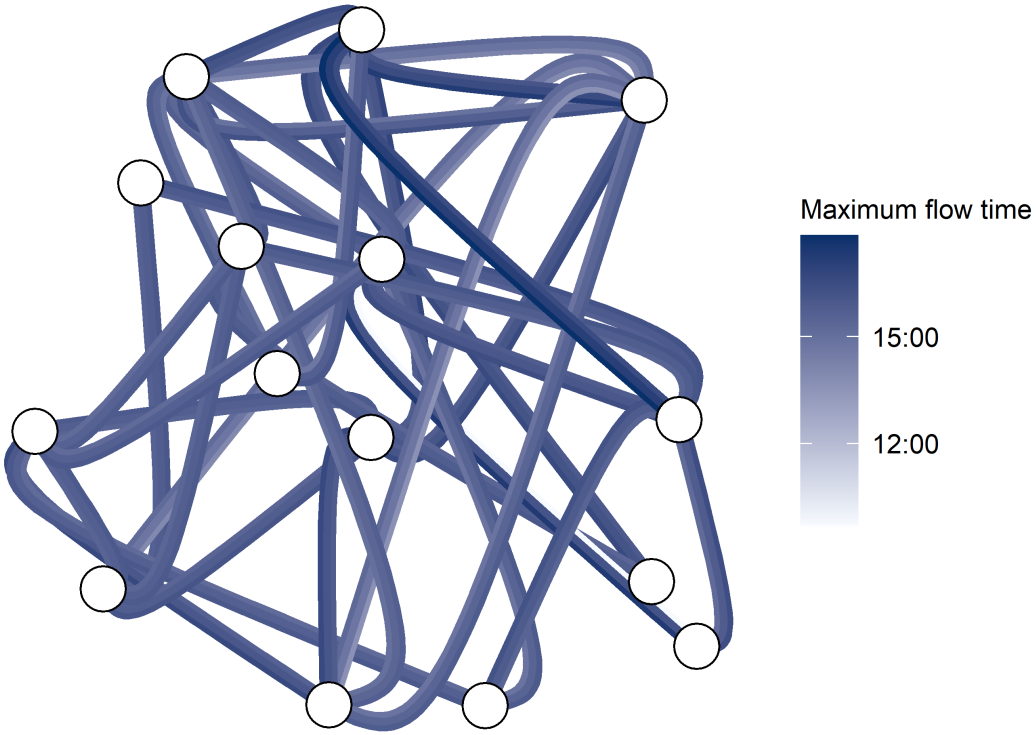


Figure 9.3: Example map of times of maximum flow along edges in a network. Each edge shows three colours – the point estimate and upper and lower 95% credible interval bounds.

investigated in the simulations in Chapter 8. This is a potentially useful and interesting way of summarising spatio-temporal networks over time, and enabling interpolation of network properties at any time point in the presence of non-simultaneous measurements without the loss of accuracy associated with temporal data aggregation. However, to investigate how useful this is, the ability of the models to recover temporal functions of network properties accurately and with appropriate estimates of uncertainty should be investigated. There are several papers in the literature that test coverage probabilities of confidence bands using simulations. In Francisco-Fernández and Quintela-del-Río (2016), for example, confidence bands were constructed using a series of simultaneous confidence intervals at fixed points along the x-axis of a function. Simultaneous confidence intervals are point-wise confidence intervals spaced along a function and corrected for multiple testing. To test coverage probabilities, confidence bands were constructed for a function estimated from 1000 simulated datasets. For each simulation, whether the simulated function exceeded the bounds of any of the simultaneous confidence intervals at the fixed points in question was recorded. This was used to calculate the percentage of confidence bands that fully covered the simulated function (Francisco-Fernández and Quintela-del-Río 2016).

This method of generated confidence bands using simultaneous confidence intervals at fixed points is the same as the process described in Chapter 4. Chapter 4 uses the Bonferroni correction for multiple testing which tends to result in conservative estimates of uncertainty (wider than necessary for the coverage probability desired) (Francisco-Fernández and Quintela-del-Río 2016; Neumann and Polzehl 1998). It would be useful for future work to evaluate the true coverage probabilities of the simultaneous credible intervals constructed in Chapters 4 and 7 and to investigate alternative corrections such as the iterative procedures presented in Francisco-Fernández and Quintela-del-Río (2016).

This evaluation of the confidence bands generated from a series of simultaneous confidence intervals, however, does highlight a potential issue with the interpretation of this approach. Often papers present a confidence band constructed in this way as a continuous region along the x- or time-axis of a graph (Francisco-Fernández and Quintela-del-Río 2016; Hong et al. 2010; Lenhoff et al. 1999). However, we only have information about the uncertainty of the estimated function at the locations along this axis where we calculated

simultaneous confidence intervals. Drawing a line joining these intervals implies that we have information about uncertainty in locations where we do not. If these lines pass through an area of sparse data, for example, we expect greater uncertainty (as we do not have much information) that is not reflected. Perhaps, in a similar way to dealing with the issues in Chapter 6, it would be prudent for researchers to present simultaneous confidence intervals as a series of discrete intervals, rather than lines showing continuous confidence bands. This would highlight the nature of construction of these confidence bands to readers and give a clear indication of where along the function uncertainty has (and has not) been estimated in the form of simultaneous credible or confidence intervals. It would be useful for this potential issue to be investigated in future work, for example, by reviewing the extent to which confidence bands constructed from simultaneous confidence intervals are presented in this way and perhaps the density of simultaneous confidence intervals along the x- or time-axis used to construct them. This may give an indication of the extent to which this is an issue and provide some insight into whether it should be further addressed.

9.7.1 Further work

In the modelled temporal patterns of whole-network properties in Chapter 7 we saw particularly wide credible intervals towards the boundaries of the data. This may be a result of the erratic behaviour of some of the posterior estimated functions used to define these intervals. The model specified in this thesis used a b-spline basis to capture complex curves. This was chosen as b-splines are computationally easy to deal with and were uncomplicated to implement, allowing focus on fitting the spatial and temporal autocorrelation structure in the model (Perperoglou et al. 2019). The real and simulated data used in Chapter 7 spanned only a period of 24 hours and did not show any signs of a periodic structure. There are a range of alternative spline specifications that could be used in this situation. Natural cubic splines are discussed in Chapter 4. These are similar to b-splines, in that they are mathematically convenient (they can be calculated before fitting the model and incorporated as a covariate), but they include additional constraints that reduce erratic behaviour towards the boundaries of the data (Perperoglou et al. 2019). If the erratic behaviour of b-splines is a cause of the wide credible intervals, using a spline specification that reduces this erratic behaviour may result in narrower credible intervals

towards the boundaries of the data.

Another alternative group of splines is penalised splines. Penalised splines use an estimation procedure that maximises a likelihood function multiplied by a penalty for 'roughness'. For example, p-splines are a type of penalised spline which are based on cubic b-splines with many, equally spaced knots. The roughness penalty is based on the second order difference between the model value at these knots which can be thought of as an approximation to the rate of change or 'wiggliness' (Kauermann and Kuhlenkasper 2013). Increasing the effect of the penalty on the model estimation will increase the smoothness. The use of a roughness penalty means the shape of the model depends less on knot placement than when using b-splines or natural cubic splines (Kauermann and Kuhlenkasper 2013). However, penalised splines use an altered estimation procedure which may make them less convenient when estimating complex models. It would be interesting to investigate the use of penalised splines with multilevel models and Bayesian estimation tools.

In a situation with periodic data, alternative basis functions should be considered. The Fourier basis system is one of these, which was mentioned in Section 3.6.1. A Fourier basis comprises a series of sine and cosine functions of increasing frequency (Ramsay and Silverman 1997). Wavelets are another type of basis system that can be used for periodic data but also capture abrupt changes better than Fourier basis systems. They are made of oscillatory functions with mean zero that rapidly decay over a fixed period (rather than the infinite period of Fourier bases) (Ramsay and Silverman 1997; Ullah and Finch 2013). Wavelet functions can be scaled and translated along the time-axis (Ullah and Finch 2013). Future work should investigate the application of the analysis framework outlined in Chapter 4 to periodic systems requiring the use of these types of basis system.

9.8 Simulation evaluating performance of multilevel models with spatio-temporal network data

Chapter 8 used simulations to test the accuracy and precision with which the methods in Chapter 4 were able to extract temporal pattern features. This used the same scenario as Chapter 7, but repeated simulations which can be used to provide information on model performance.

9.8.1 Model performance

The models in Chapters 7 and 8 were reasonably computationally expensive. In Chapter 8 the data generation, modelling and extraction of pattern features for one simulated dataset took approximately 48 hours with four parallel model chains and 40 parallel units calculating pattern features for each of 4000 posterior parameter samples. In Chapter 7, generation and analysis of simulated data without parallelisation took less than 48 hours. Processing and analysis of Oyster Card journey data took under 48 hours with four parallel model chains and a parallel process collecting pattern features and temporal network property functions from posterior parameter samples. All analyses required the use of a high performance computing unit which is not universally available and the computational intensity of this analysis is a disadvantage of the approach, especially when working with large datasets. The process of extracting posterior distributions for temporal pattern features was particularly time consuming. It would be interesting to explore the extent to which this could be sped up by improved coding practices and the use of alternative programming languages which are less computationally expensive than R, for example by completing data processing or the extraction of posterior distributions in shell script (Soundararaj et al. 2019).

The results in Chapter 8 show that the models need some development to improve their accuracy. While all three pattern features included in this simulation (time of maximum delay, maximum journey length and maximum delay) were estimated with a reasonably small mean bias, low values were overestimated and high values underestimated for all three features. This may be an indication that the models do not capture variation in pattern features fully.

Credible intervals for time were also wide. Appendix G shows the posterior distributions of pattern features for all edges for one simulation from the thousand conducted in Chapter 8. While the distributions for maximum journey length and maximum delay appear to have converged well, the distributions for maximum time are not as smooth and are often multimodal. This may indicate some poor mixing of the four chains used to estimate the model for this feature and will likely be an additional contributor to the particularly wide credible intervals for this feature (Gilks and Roberts 1995).

Across all three pattern features, 95% credible interval coverage was below the expected

level (95%). This may in part be because of the bias in estimates of low and high values for each pattern feature, with 95% limits of agreement being greater than twice as wide as the average credible intervals for two of three pattern features. It is important to re-evaluate the coverage of the 95% credible intervals calculated as described in Chapter 4 in a situation where the models perform with little or no systematic bias – this would give a clearer indication of whether their coverage is appropriate.

Considering an alternative model specification that includes x-axis random effects may help to capture temporal variation in maximum time and other temporal pattern features more fully, and potentially lead to better convergence of the posterior distribution for this pattern feature and reduced bias for extreme values of the features. However, incorporating an x-axis random effect adds complexity to the model which may make it more difficult to estimate. It may be useful to consider an alternative specification of the random effects, as discussed in Section 8.5. Here it was suggested that, instead of incorporating random effects into b-spline coefficients which do not have a direct interpretation, random coefficients that transform the whole temporal function (x- and y-axis translation and stretching) could be incorporated, similar to the SITAR model specification (Cole et al. 2010). This may be of use not only in producing a more interpretable and meaningful random structure, but also improving the accuracy with which the model captures temporal pattern features.

One of the major potential issues with the approach outlined in Chapter 4, and many other approaches aiming to predict or interpolate properties for individual observational units, is the aim of predicting information at an individual level from a population model. This is a reasonably common aim, particularly, for example, in the field of 'personalised medicine', but there are major issues with it. Statistical models, including machine learning models and discrete-time models mentioned in earlier chapters, estimate average expected values for the whole population (Wilkinson et al. 2020). In the context of these models, the expected temporal pattern of a property is estimated for the whole population. This can also be estimated for groups, but the precision of estimates at any lower level (than the whole population) is decreased (Wilkinson et al. 2020). The estimates of temporal patterns at the level of individual edges in Chapters 7 and 8 are comprised of the expected values for two groups – the origin and destination groups of the edge. This means that

we cannot necessarily expect models to precisely recover temporal patterns and temporal pattern features at the level of individual observational units, if these estimates are based on higher group level information. This is one of the potential reasons we see in Chapter 8 that estimates of observational unit level pattern features are on average unbiased but biased for low and high values of the pattern features.

9.8.2 Further work

The simulation study in Chapter 8 tests the application of the analysis framework in only one specific circumstance. If we are to fully understand the potential applications of the framework outlined in Chapters 4 and 7, we must test its application in a range of scenarios. This necessarily involves choosing appropriate basis functions to reflect the underlying data generating processes (for example periodicity), which may differ from b-splines and explore some of the possibilities outlined in Section 9.7.1.

There is a huge range of different scenarios that need to be tested to understand when it is useful and appropriate to use a particular method. For this analysis, one particularly important aspect is the complexity of the temporal patterns that models are attempting to recover. A 'wigglier' pattern will likely be more difficult to capture accurately. In addition, it is likely that some patterns will include multiple maxima or minima with more similar values to those in the simulated pattern in Chapter 8. In this scenario, it will be harder for models to capture the correct maximum across the whole time period and this should be tested. It is also possible that in this scenario the posterior distribution for the time of a maximum, estimated by a model, could be distributed bi- or multi-modally if different posterior parameter sets estimate different local simulated maxima to have the greatest value within the range of the data. While this is a valid result from the model, summarising a bi- or multi-modal posterior distribution with 95% credible intervals does not give any information about the shape of the distribution between the bounds of this interval – most readers would probably assume the distribution was a wide unimodal distribution.

It would be useful for future work to explore whether this scenario would occur frequently and how it is best to summarise bi- or multi-modal posterior distributions of pattern features while balancing simplicity and detailed description. Code for simulations that test the ability of the model in Chapter 4 to recover the time and value maxima from a

range of different temporal pattern shapes is available in the Online Appendix. This is based on abstract scenarios, rather than a realistic hypothetical situation, and includes 24 simulated patterns with different maximum 'shapes' (smooth and wide or narrow and pointed), different numbers of maxima and different pattern amplitudes. This could be adapted to investigate the possibility of multimodal posterior distributions for pattern features. This has not been run as part of this research due to heavy computational demands and therefore time constraints – running one simulation for a single scenario takes approximately two days, 1000 simulations for a single scenario takes at least two weeks with a large degree of parallelisation. Running simulations for 24 different scenarios would likely take almost a year.

Other similar simulations should be carried out to investigate the effect of changing data structure as well as changes in temporal pattern shapes. In particular, changes to the extent of temporal and spatial autocorrelation and the data measurement scheme should be investigated. Measurements in the simulations in Chapter 8 are uniformly distributed across the study period; however, in real life, measurements of journey lengths will be taken at the time that the journey is undertaken. People will not take journeys at uniformly distributed times throughout the day but will be more likely to travel in the morning and evening commute times, for example. The uneven distribution of measurements throughout time will affect the model and it is important to investigate the extent of this effect on the recovery of temporal patterns (Liang et al. 2009; Song et al. 2012). In addition, simulation studies could be used to investigate the effect of different measurement schemes such as wave and regular measurements on model performance, when compared to uniformly distributed measurements. These measurement schemes are outlined in Chapter 5 and can be simulated using the simulation tool from that chapter. The simulation tool can also produce data with different temporal and spatial autocorrelation strengths but has not been extended to include different parameterisations, such as different weighting schemes, and is thus based mainly on geographical distance.

Future work should also consider another area that this work has not explored: whether it would be useful to incorporate other variables, as well as time, into the multilevel models used to extract temporal pattern features. Chapter 6 explains that including covariates in models for temporal patterns and interpreting the average values over time, with respect

to these covariates, as a temporal pattern is not a valid interpretation. While this is true, the incorporation of covariates could be useful in improving the accuracy with which pattern features are recovered within individual observational units. For example, in the simulated scenarios in Chapters 7 and 8, there was a latent fixed effect simulated in the data. A hypothetical unmeasured variable affected the range of the function. Not accounting for this variable may have led to a violation of some assumptions underlying the model, particularly that of normally distributed random effects coefficients. This may have contributed to some of the issues with accuracy and credible interval coverage that we saw in Chapter 8. If the variable that defined these effects had been hypothetically measured, incorporating it into the model may have improved fit and potentially overcome some of the model's shortcomings. Of course, including more variables and complexity in a model comes at a cost: decreased degrees of freedom, which can make estimation more difficult and introduce some bias (Browne and Draper 2000). When modelling any process, it is never possible to fully capture the highly complex systems that surround us and generate the data we collect – researchers must strike a balance between model parsimony and detailed reflection of these complex data generating systems at play in order to produce a model that is "useful" in the particular situation they are focused on (Box 1987).

9.9 Conclusion

This research clearly identified a lack of use of continuous-time models in the spatio-temporal network literature. There are a range of issues with the use of discrete-time models that have been outlined throughout this thesis, and research into the use of continuous-time models in this context was identified as an important area. After considering a range of different possible continuous-time models, a novel approach using multilevel models to capture temporal patterns of edge properties in a network was introduced. This included the simplification of results to improve their interpretability by a) extracting temporal pattern features and b) generating continuous temporal functions of network properties. In particular, this second approach to simplification was new to the literature.

An example application of these methods to real and simulated data was produced and the extraction of temporal pattern features was tested using simulations. The simulations

identified some shortcomings of the approach, but the model has great potential for wide ranging applications in the many fields that analyse spatio-temporal network structures. Future work should be carried out to test alternative model specifications and model performance in different scenarios.

The production of a simulation tool for generating simulated temporal data with a highly flexible structure was also an important contribution of this project – this could be an important and useful tool for future work in this area as it simplifies the process of simulating data from scratch.

There are some caveats to be aware of when analysing temporal data, a key one being the issue of conditioning on the outcome which was identified as a reasonably common approach when analysing temporal data in Chapter 3. Chapter 6 included a simulation that described some of the issues with this approach to discourage its use – this is an important contribution to the literature as avoiding methods that do not perform as they purport to is key to ensuring the quality of future analyses and research.

Overall, this research introduces and tests a novel idea with potentially wide-ranging applications that fills the gap in the spatio-temporal network literature for continuous-time analyses. Several suggestions for future work have been made, including testing the methods in different scenarios and with alternative continuous-time models and specifications. Exploring the interpretability of results through qualitative work could also be an interesting avenue for future development to pursue.

References

Abdelghany, A., Mahmassani, H., & Chiu, Y. (2001). Spatial microassignment of travel demand with activity trip chains. *Transportation Research Record*, *1777*(1), 36–46.

Aris, I., Bernard, J., Chen, L., Tint, M., Pang, W., Lim, W., Soh, S., Saw, S., Godfrey, K., Gluckman, P., Chong, Y., Yap, F., Kramer, M., & Lee, Y. (2017). Infant body mass index peak and early childhood cardio-metabolic risk markers in a multi-ethnic Asian birth cohort. *International Journal of Epidemiology*, *46*(2), 513–525.

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), 1–20.

Blonder, B., Wey, T., Dornhaus, A., James, R., & Sih, A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, *3*(6), 958–972.

Bollen, K., & Curran, P. (2005). *Latent curve models: A structural equation perspective*. John Wiley & Sons.

Box, G. (1987). *Empirical model-building and response surfaces*. Wiley.

Browne, W., & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics*, *15*(3), 391–420.

Browne, W., Draper, D., Goldstein, H., & Rasbash, J. (2002). Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. *Computational Statistics & Data Analysis*, *39*(2), 203–225.

Chen, S., Ilany, A., White, B., Sanderson, M., & Lanzas, C. (2015). Spatial-temporal dynamics of high-resolution animal networks: What can we learn from domestic animals? *PLOS One*, *10*(6), e0129253.

Cheng, T., Wang, J., Haworth, J., Heydecker, B., & Chow, A. (2014). A dynamic spatial weight matrix and localized space–time autoregressive integrated moving average for network modeling. *J Geographical Analysis*, *46*(1), 75–97.

Colak, S., Schneider, C., Wang, P., & Gonzalez, M. (2013). On the role of spatial dynamics and topology on network flows. *New Journal of Physics*, *15*, 113037.

Cole, T., Donaldson, M., & Ben-Shlomo, Y. (2010). SITAR - A useful instrument for growth curve analysis. *International Journal of Epidemiology*, *39*(6), 1558–1566.

Comber, A., & Wulder, M. (2019). Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS*, *23*(5), 879–891.

de Haan-Rietdijk, S., Voelkle, M., Keijsers, L., & Hamaker, E. (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*, 1849–1849.

Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics*, *21*(3-4), 224–239.

Dubin, R. (2009). Spatial weights. In A. S. Fotheringham & P. A. Rogerson (Eds.), *The SAGE handbook of spatial analysis*. SAGE Publications, Ltd.

Ermagun, A., & Levinson, D. (2018). An introduction to the network weight matrix. *Geographical Analysis*, *50*(1), 76–96.

Færch, K., Witte, D., Tabák, A., Perreault, L., Herder, C., Brunner, E., Kivimäki, M., & Vistisen, D. (2013). Trajectories of cardiometabolic risk factors before diagnosis of three subtypes of type 2 diabetes: A post-hoc analysis of the longitudinal Whitehall II cohort study. *Lancet Diabetes Endocrinol*, *1*(1), 43–51.

Fang, Y., Wang, Z., Cheng, R., Li, X., Luo, S., Hu, J., & Chen, X. (2019). On spatial-aware community search. *IEEE Transactions on Knowledge and Data Engineering*, *31*(4), 783–798.

Fielding, A., & Goldstein, H. (2006). Cross-classified and multiple membership structures in multilevel models : An introduction and review (Research Report No. 791).

Francisco-Fernández, M., & Quintela-del-Río, A. (2016). Comparing simultaneous and pointwise confidence intervals for hydrological processes. *PLOS ONE*, *11*(2), e0147505.

Freeman, J. (1989). Systematic sampling, temporal aggregation, and the study of political relationships. *Political Analysis*, *1*, 61–98.

Gadd, S., Comber, A., Gilthorpe, M., Suchak, K., & Heppenstall, A. (2021). Simplifying the interpretation of continuous-time models for spatio-temporal networks. *Journal of Geographical Systems*.

Gadd, S., Tennant, P., Heppenstall, A., Boehnke, J., & Gilthorpe, M. (2019). Analysing trajectories of a longitudinal exposure: A causal perspective on common methods in lifecourse research. *PLOS ONE*, *14*(12), e0225217.

Gao, M., Brannstrom, L., & Almquist, Y. (2017). Exposure to out-of-home care in childhood and adult all-cause mortality: A cohort study. *International Journal of Epidemiology*, *46*(3), 1010–1017.

Gilks, W., & Roberts, G. (1995). Strategies for improving MCMC. In W. Gilks, S. Richardson, & D. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice*. CRC Press.

Gilthorpe, M., Dahly, D., Tu, Y., Kubzansky, L., & Goodman, E. (2014). Challenges in modelling the random structure correctly in growth mixture models and the impact this has on model mixtures. *Journal of Developmental Origins of Health and Disease*, *5*(3), 197–205.

Goldstein, H. (1994). Multilevel cross-classified models. *J Sociological Methods Research*, *22*(3), 364–375.

Goldstein, H. (2011). *Multilevel statistical models* (Fourth edition). Wiley.

Goldstein, H., Healy, M., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *J Statistics in medicine*, *13*(16), 1643–1655.

Gordon, J., Koutsopoulos, H., & Wilson, N. (2018). Estimation of population origin– interchange–destination flows on multimodal transit networks. *Transportation Research Part C: Emerging Technologies*, *90*, 350–365.

Graser, A., Schmidt, J., Roth, F., & Brändle, N. (2017). Untangling origin-destination flows in geographic information systems. *Information Visualization*, *18*(1), 153–172.

Head, M., Holman, L., Lanfear, R., Kahn, A., & Jennions, M. (2015). The extent and consequences of P-hacking in science. *PLOS Biology*, *13*(3), e1002106.

Hong, Y., Escobar, L., & Meeker, W. (2010). Coverage probabilities of simultaneous confidence bands and regions for log-location-scale distributions. *Statistics & Probability Letters*, *80*(7), 733–738.

Hulsegge, G., Spijkerman, A., van der Schouw, Y., Bakker, S., Gansevoort, R., Smit, H., & Verschuren, W. (2017). Trajectories of metabolic risk factors and biochemical markers prior to the onset of type 2 diabetes: The population-based longitudinal Doetinchem study. *Nutr Diabetes*, *7*(5), e270.

Hwang, S. (2000). The effects of systematic sampling and temporal aggregation on discrete time long memory processes and their finite sample properties. *Econometric Theory*, *16*(3), 347–372.

Ibrahim, J., Chu, H., & Chen, L. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, *28*(16), 2796–2801.

Jin, Y., Peng, R., & Shi, J. (2019). Population dynamics in river networks. *Journal of Nonlinear Science*, *29*(6), 2501–2545.

Jones, R., & Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics*, *47*(1), 161–175.

Kauermann, G., & Kuhlenkasper, T. (2013). Penalized splines and multilevel models. In M. A. Scott, J. S. Simonoff, & B. D. Marx (Eds.), *The SAGE handbook of multilevel models*. SAGE Publications Ltd.

Kwan, M.-P. (2013). Beyond space (as we knew it): Toward temporally integrated geographies of segregation, health, and accessibility. *Annals of the Association of American Geographers*, *103*(5), 1078–1086.

Lengyel, B., Bokanyi, E., Di Clemente, R., Kertesz, J., & Gonzalez, M. (2020). The role of geography in the complex diffusion of innovations. *Scientific Reports*, *10*(1).

Lenhoff, M., Santner, T., Otis, J., Peterson, M., Williams, B., & Backus, S. (1999). Bootstrap prediction and confidence bands: A superior statistical method for analysis of gait data. *Gait & Posture*, *9*(1), 10–17.

Li, T., & Liao, Q. (2016). Dynamic networks analysis and visualization through spatiotemporal link segmentation. *IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 209–214.

Liang, Y., Lu, W., & Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, *65*(2), 377–384.

Morris, T., White, I., & Crowther, M. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102.

Nassir, N., Hickman, M., & Ma, Z.-L. (2015). Activity detection and transfer identification for public transit fare card data. *Transportation*, *42*(4), 683–705.

Neumann, M., & Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, *9*(4), 307–333.

Newman, M. (2002). The structure and function of networks. *Computer Physics Communications*, *147*(1), 40–45.

Oka, R., Shibata, K., Sakurai, M., Kometani, M., Yamagishi, M., Yoshimura, K., & Yoneda, T. (2017). Trajectories of postload plasma glucose in the development of type 2 diabetes in Japanese adults. *J Diabetes Res*, *2017*, 5307523.

Oud, J., Folmer, H., Patuelli, R., & Nijkamp, P. (2012). Continuous-time modeling with spatial dependence. *Geographical Analysis*, *44*(1), 29–46.

Peirce, S., & Lindsay, J. (2015). Characterizing ephemeral streams in a southern Ontario watershed using electrical resistance sensors. *Hydrological Processes*, *29*(1), 103–111.

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. *BMC Medical Research Methodology*, *19*(1), 46.

Piccininni, M., Konigorski, S., Rohmann, J., & Kurth, T. (2020). Directed acyclic graphs and causal thinking in clinical risk prediction modeling. *BMC Medical Research Methodology*, *20*(1), 179.

Ramsay, J., & Silverman, B. (1997). *Functional data analysis*. Springer.

Rossana, R., & Seater, J. (1995). Temporal aggregation and economic time series. *Journal of Business & Economic Statistics*, *13*(4), 441–451.

Sabia, S., Dugravot, A., Dartigues, J., Abell, J., Elbaz, A., Kivimaki, M., & Singh-Manoux, A. (2017). Physical activity, cognitive decline, and risk of dementia: 28 year follow-up of Whitehall II cohort study. *Bmj-British Medical Journal*, *357*.

Shoaff, J., Papandonatos, G., Calafat, A., Ye, X., Chen, A., Lanphear, B., Yolton, K., & Braun, J. (2017). Early-life phthalate exposure and adiposity at 8 years of age. *Environmental Health Perspectives*, *125*(9).

Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*, *22*(11), 1359–66.

Song, X., Mu, X., & Sun, L. (2012). Regression analysis of longitudinal data with time-dependent covariates and informative observation times. *Scandinavian Journal of Statistics*, *39*(2), 248–258.

Soundararaj, B., Cheshire, J., & Longley, P. (2019). Medium data toolkit - A case study on smart street sensor project. *GISRUK*.

Sun, G., Tang, T., Peng, T., Liang, R., & Wu, Y. (2017). SocialWave: Visual analysis of spatio-temporal diffusion of information on social media. *ACM Transactions on Intelligent Systems and Technology*, *9*(2), 2157–6904.

Tilling, K., Davies, N., Nicoli, E., Ben-Shlomo, Y., Kramer, M., Patel, R., Oken, E., & Martin, R. (2011). Associations of growth trajectories in infancy and early childhood with later childhood outcomes. *The American Journal of Clinical Nutrition*, *94*(6 Suppl), 1808S–1813S.

Tu, Y.-K., Tilling, K., Sterne, J., & Gilthorpe, M. (2013). A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. *International Journal of Epidemiology*, *42*(5), 1327–1339.

Ugrinowitsch, C., Fellingham, G., & Ricard, M. (2004). Limitations of ordinary least squares models in analyzing repeated measures data. *36*(12), 2144–2148.

Ullah, S., & Finch, C. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, *13*(1), 43.

Van Kampen, N. (1976). Stochastic differential equations. *Physics Reports*, *24*(3), 171–228.

Wilkinson, J., Arnold, K., Murray, E., van Smeden, M., Carr, K., Sippy, R., de Kamps, M., Beam, A., Konigorski, S., Lippert, C., Gilthorpe, M., & Tennant, P. (2020). Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health*, *2*(12), e677–e680.

Wolf, C., & Best, H. (2014a). Linear regression. In C. Wolf & H. Best (Eds.), *The SAGE handbook of regression analysis and causal inference*. SAGE Publications Ltd.

Wolf, C., & Best, H. (2014c). Regression analysis: Assumptions and diagnostics. In C. Wolf & H. Best (Eds.), *The SAGE handbook of regression analysis and causal inference*. SAGE Publications Ltd.

Wood, J., Dykes, J., & Slingsby, A. (2010). Visualisation of origins, destinations and flows with OD maps. *The Cartographic Journal*, *47*(2), 117–129.

Zhang, W., Leng, C., & Tang, C. (2015). A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *77*(1), 219–238.

References

Abdelghany, A., Mahmassani, H., & Chiu, Y. (2001). Spatial microassignment of travel demand with activity trip chains. *Transportation Research Record*, *1777*(1), 36–46.

Aber, M., & McArdle, J. (1991). Latent growth curve approaches to modelling the development of competence. In M. J. Chandler & M. Chapman (Eds.), *Criteria for competence: Controversies in the conceptualization and assessment of children's abilities*. L. Erlbaum Associates.

Aigner, W., Miksch, S., Müller, W., Schumann, H., & Tominski, C. (2007). Visualizing time-oriented data - A systematic view. *Computers & Graphics*, *31*(3), 401–409.

Albers, H., Hall, K., Lee, K., Taleghan, M., & Dietterich, T. (2018). The role of restoration and key ecological invasion mechanisms in optimal spatial-dynamic management of invasive species. *Ecological Economics*, *151*, 44–54.

Albert, P. (1999). Longitudinal data analysis (repeated measures) in clinical trials. *18*(13), 1707–1732.

Anderson, T., & Dragicevic, S. (2018). Network-agent based model for simulating the dynamic spatial network structure of complex ecological systems. *Ecological Modelling*, *389*, 19–32.

Anderson, T., & Dragićević, S. (2018). A geographic network automata approach for modeling dynamic ecological systems. *Geographical Analysis*, *52*, 3–27.

Anderson, T., & Dragićević, S. (2020a). NEAT approach for testing and validation of geospatial network agent-based model processes: Case study of influenza spread. *International Journal of Geographical Information Science*, *34*(9), 1792–1821.

Anderson, T., & Dragićević, S. (2020b). Representing complex evolving spatial networks: Geographic network automata. *ISPRS International Journal of Geo-Information*, *9*(4).

Angeloudis, P., & Fisk, D. (2006). Large subway systems as complex networks. *Physica A: Statistical Mechanics and its Applications*, *367*, 553–558.

Apparicio, P., Abdelmajid, M., Riva, M., & Shearmur, R. (2008). Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics*, *7*(1), 7.

Aris, I., Bernard, J., Chen, L., Tint, M., Pang, W., Lim, W., Soh, S., Saw, S., Godfrey, K., Gluckman, P., Chong, Y., Yap, F., Kramer, M., & Lee, Y. (2017). Infant body mass index peak and early childhood cardio-metabolic risk markers in a multi-ethnic Asian birth cohort. *International Journal of Epidemiology*, *46*(2), 513–525.

Austwick, M., O'Brien, O., Strano, E., & Viana, M. (2013). The structure of spatial networks and communities in bicycle sharing systems. *PLOS One*, *8*(9).

Aydin, B., Leite, W., & Algina, J. (2014). The consequences of ignoring variability in measurement occasions within data collection waves in latent growth models. *Multivariate Behavioral Research*, *49*(2), 149–160.

Bagchi, M., & White, P. (2005). The potential of public transport smart card data. *Transport Policy*, *12*(5), 464–474.

Barker, D., Osmond, C., Forsén, T., Kajantie, E., & Eriksson, J. (2005). Trajectories of growth among children who have coronary events as adults. *New England Journal of Medicine*, *353*(17), 1802–1809.

Basile, R., Parteka, A., & Pittiglio, R. (2018). Export diversification and economic development: A dynamic spatial data analysis. *Review of International Economics*, *26*(3), 634–650.

Baybeck, B., & Huckfeldt, R. (2002). Urban contexts, spatially dispersed networks, and the diffusion of political information. *Political Geography*, *21*(2), 195–220.

Ben-Shlomo, Y., & Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, *31*(2), 285–293.

Besag, J., & Kooperberg, C. (1995). On conditional and intrinsic autoregression. *Biometrika*, *82*(4), 733–746.

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), 1–20.

Bhargava, S., Sachdev, H., Fall, C., Osmond, C., Lakshmy, R., Barker, D., Biswas, S., Ramji, S., Prabhakaran, D., & Reddy, K. (2004). Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood. *350*(9), 865–875.

Bi, Q., Goodman, K., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American Journal of Epidemiology*, *188*(12), 2222–2239.

Bjork, T. (2004). *Arbitrage theory in continuous time*. Oxford University Press.

Bland, J., & Altman, D. (1999). Measuring agreement in method comparison studies. *Stat Methods Med Res*, *8*(2), 135–60.

Blonder, B., & Dornhaus, A. (2011). Time-ordered networks reveal limitations to information flow in ant colonies. *PLOS one*, *6*(5), e20298.

Blonder, B., Wey, T., Dornhaus, A., James, R., & Sih, A. (2012). Temporal dynamics and network analysis. *Methods in Ecology and Evolution*, *3*(6), 958–972.

Blozis, S., & Cho, Y. (2008). Coding and centering of time in latent curve models in the presence of interindividual time heterogeneity. *Structural Equation Modeling-a Multidisciplinary Journal*, *15*(3), 413–433.

Blozis, S., Conger, K., & Harring, J. (2007). Nonlinear latent curve models for multivariate longitudinal data. *International Journal of Behavioral Development*, *31*(4), 340–346.

Bollen, K., & Curran, P. (2004). Autoregressive latent trajectory (ALT) models a synthesis of two traditions. *Sociological Methods & Research*, *32*(3), 336–383.

Bollen, K., & Curran, P. (2005). *Latent curve models: A structural equation perspective*. John Wiley & Sons.

Box, G. (1987). *Empirical model-building and response surfaces*. Wiley.

Browne, W., & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics*, *15*(3), 391–420.

Browne, W., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *J Bayesian analysis*, *1*(3), 473–514.

Browne, W., Draper, D., Goldstein, H., & Rasbash, J. (2002). Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. *Computational Statistics & Data Analysis*, *39*(2), 203–225.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 28.

Burnett, J., & Zhao, X. [X.]. (2017). Spatially explicit prediction of wholesale electricity prices. *International Regional Science Review*, *40*(2), 99–140.

Burton, A., Altman, D., Royston, P., & Holder, R. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, *25*(24), 4279–4292.

Byrne, B., & Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics*, *2*(3), 177–203.

Cai, H., Rahman, A., Su, X., & Zhang, H. (2014). A GIS-microscopic simulation approach for optimizing road barrier placement and configuration in university campus emer-

gency evacuation. *International Journal of Disaster Resilience in the Built Environment*, *5*(4), 362–379.

Carlsson, A., & Kindvall, O. (2001). Spatial dynamics in a metapopulation network: Recovery of a rare grasshopper Stauroderus scalaris from population refuges. *Ecography*, *24*(4), 452–460.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *J Journal of Statistical Software*, *76*(1), 32.

Carter, N., Baeza, A., & Magliocca, N. (2020). Emergent conservation outcomes of shared risk perception in human-wildlife systems. *Conservation Biology*, *34*(4), 903–914.

Caruso, M. (2007). Disruptive dynamics: The spatial dimensions of the spanish networks in the spread of monitorial schooling (1815-1825). *Paedagogica Historica*, *43*(2), 271–282.

Chen, L., Han, K., Yin, Q., & Cao, Z. (2020). GDCRN: Global diffusion convolutional residual network for traffic flow prediction. https://doi.org/10.1007/978-3-030-55393-7_39

Chen, S., Claramunt, C., & Ray, C. (2014). A spatio-temporal modelling approach for the study of the connectivity and accessibility of the Guangzhou metropolitan network. *Journal of Transport Geography*, *36*, 12–23.

Chen, S., Ilany, A., White, B., Sanderson, M., & Lanzas, C. (2015). Spatial-temporal dynamics of high-resolution animal networks: What can we learn from domestic animals? *PLOS One*, *10*(6), e0129253.

Cheng, S., Xie, B., Bie, Y., Zhang, Y., & Zhang, S. (2018). Measure dynamic individual spatial-temporal accessibility by public transit: Integrating time-table and passenger departure time. *Journal of Transport Geography*, *66*, 235–247.

Cheng, T., Wang, J., Haworth, J., Heydecker, B., & Chow, A. (2011). Modelling dynamic space-time autocorrelations of urban transport network. *Proceedings of the 11th international conference on Geocomputation 2011*, 215–210.

Cheng, T., Wang, J., Haworth, J., Heydecker, B., & Chow, A. (2014). A dynamic spatial weight matrix and localized space–time autoregressive integrated moving average for network modeling. *J Geographical Analysis*, *46*(1), 75–97.

Colak, S., Schneider, C., Wang, P., & Gonzalez, M. (2013). On the role of spatial dynamics and topology on network flows. *New Journal of Physics*, *15*, 113037.

Cole, T., Donaldson, M., & Ben-Shlomo, Y. (2010). SITAR - A useful instrument for growth curve analysis. *International Journal of Epidemiology*, *39*(6), 1558–1566.

Comber, A., & Wulder, M. (2019). Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS*, *23*(5), 879–891.

Congdon, P. (2014). *Applied Bayesian modelling*. Wiley.

Copas, J. (1983). Regression, prediction and shrinkage. *45*(3), 311–335.

Crooks, A., & Heppenstall, A. (2011). Introduction to agent-based modelling. In A. J. Heppenstall, M. Batty, A. T. Crooks, & L. M. See (Eds.), *Agent-based models of geographical systems* (First edition, pp. 95–105). Springer Netherlands.

Davidian, M., & Giltinan, D. (1995). *Nonlinear models for repeated measurement data* (Vol. 62). Chapman & Hall.

De Boor, C. (2001). *A practical guide to splines* (Revised edition). Springer.

de Haan-Rietdijk, S., Voelkle, M., Keijsers, L., & Hamaker, E. (2017). Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*, 1849–1849.

Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics*, *21*(3-4), 224–239.

Diao, Z., Wang, X., Zhang, D., Liu, Y., Xie, K., & He, S. (2019). Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 890–897.

Ditlevsen, S., & De Gaetano, A. (2005). Mixed effects in stochastic differential equation models. *REVSTAT-Statistical Journal*, *3*(2), 137–153.

Donfouet, H., Jeanty, P., & Malin, E. (2018). Analysing spatial spillovers in corruption: A dynamic spatial panel data approach. *Papers in Regional Science*, *97*, S63–S78.

Dubin, R. (1998). Spatial autocorrelation: A primer. *Journal of Housing Economics*, *7*(4), 304–327.

Dubin, R. (2009). Spatial weights. In A. S. Fotheringham & P. A. Rogerson (Eds.), *The SAGE handbook of spatial analysis*. SAGE Publications, Ltd.

Duncan, T., Duncan, S., & Strycker, L. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (Second edition). Lawrence Erlbaum Associates.

Dung, V., & Tjahjowidodo, T. (2017). A direct method to solve optimal knots of b-spline curves: An application for non-uniform b-spline curves fitting. *PLOS ONE*, *12*(3), e0173857.

Eriksson, J. [J.], Forsén, T., Tuomilehto, J., Osmond, C., & Barker, D. (2000). Fetal and childhood growth and hypertension in adult life. *Hypertension*, *36*(5), 790–794.

Eriksson, J. [J.G.], Forsen, T., Osmond, C., & Barker, D. (2003). Pathways of infant and childhood growth that lead to type 2 diabetes. *Diabetes Care*, *26*(11), 3006.

Ermagun, A., & Levinson, D. (2018). An introduction to the network weight matrix. *Geographical Analysis*, *50*(1), 76–96.

Erős, T., Schmera, D., & Schick, R. (2011). Network thinking in riverscape conservation – A graph-based approach. *Biological Conservation*, *144*(1), 184–192.

Evans, L. (2013). *An introduction to stochastic differential equations* (Vol. 82). American Mathematical Society.

Færch, K., Witte, D., Tabák, A., Perreault, L., Herder, C., Brunner, E., Kivimäki, M., & Vistisen, D. (2013). Trajectories of cardiometabolic risk factors before diagnosis of three subtypes of type 2 diabetes: A post-hoc analysis of the longitudinal Whitehall II cohort study. *Lancet Diabetes Endocrinol*, *1*(1), 43–51.

Fan, X., Li, X., Yin, J., & Liang, J. (2019). Temporal characteristics and spatial homogeneity of virtual water trade: A complex network analysis. *Water Resources Management*, *33*(4), 1467–1480.

Fang, S., Zhang, Q., Meng, G., Xiang, S., & Pan, C. (2019). GSTnet: Global spatial-temporal network for traffic flow prediction. *2019-August*, 2286–2293.

Fang, Y., Wang, Z., Cheng, R., Li, X., Luo, S., Hu, J., & Chen, X. (2019). On spatial-aware community search. *IEEE Transactions on Knowledge and Data Engineering*, *31*(4), 783–798.

Feng, D., Wu, Z., Zhang, J., & Wu, Z. (2020). Dynamic global-local spatial-temporal network for traffic speed prediction. *IEEE Access*, *8*, 209296–209307.

Ferenc, S., Tamas, P., & Jozsef, B. (2016). Complex analysis of the dynamic effects of car population along the trajectories. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*.

Fielding, A., & Goldstein, H. (2006). Cross-classified and multiple membership structures in multilevel models : An introduction and review (Research Report No. 791).

Fitzmaurice, G., & Ravichandran, C. (2008). A primer in longitudinal data analysis. *Circulation*, *118*(19), 2005–2010.

Forsén, T., Osmond, C., Eriksson, J., & Barker, D. (2004). Growth of girls who later develop coronary heart disease. *Heart*, *90*(1), 20.

Fotheringham, A., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis*, *47*(4), 431–452.

Francisco-Fernández, M., & Quintela-del-Río, A. (2016). Comparing simultaneous and pointwise confidence intervals for hydrological processes. *PLOS ONE*, *11*(2), e0147505.

Freeman, J. (1989). Systematic sampling, temporal aggregation, and the study of political relationships. *Political Analysis*, *1*, 61–98.

Freeman, L. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Freni-Sterrantino, A., Ventrucci, M., & Rue, H. (2018). A note on intrinsic conditional autoregressive models for disconnected graphs. *Spatial and Spatio-temporal Epidemiology*, *26*, 25–34.

Gadd, S., Comber, A., Gilthorpe, M., Suchak, K., & Heppenstall, A. (2021). Simplifying the interpretation of continuous-time models for spatio-temporal networks. *Journal of Geographical Systems*.

Gadd, S., Tennant, P., Heppenstall, A., Boehnke, J., & Gilthorpe, M. (2019). Analysing trajectories of a longitudinal exposure: A causal perspective on common methods in lifecourse research. *PLOS ONE*, *14*(12), e0225217.

Gao, M., Brannstrom, L., & Almquist, Y. (2017). Exposure to out-of-home care in childhood and adult all-cause mortality: A cohort study. *International Journal of Epidemiology*, *46*(3), 1010–1017.

Garbin, S., Celegon, E., Fanton, P., & Botter, G. (2019). Hydrological controls on river network connectivity. *Royal Society Open Science*, *6*(2).

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721–741.

Gesler, W. (1986). The uses of spatial analysis in medical geography: A review. *Social Science & Medicine*, *23*(10), 963–973.

Gilks, W., & Roberts, G. (1995). Strategies for improving MCMC. In W. Gilks, S. Richardson, & D. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice*. CRC Press.

Gilthorpe, M., Dahly, D., Tu, Y., Kubzansky, L., & Goodman, E. (2014). Challenges in modelling the random structure correctly in growth mixture models and the impact this has on model mixtures. *Journal of Developmental Origins of Health and Disease*, *5*(3), 197–205.

Glass, T., Goodman, S., Hernán, M., & Samet, J. (2013). Causal inference in public health. *Annual review of public health*, *34*, 61–75.

Gneiting, T., Genton, M., & Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability, and full symmetry. In B. Finkenstadt, L. Held, & V. Isham (Eds.), *Statistical methods for spatio-temporal systems*. CRC Press.

Golding, J. (2004). The Avon longitudinal study of parents and children (ALSPAC) - Study design and collaborative opportunities. *European Journal of Endocrinology*, *151*(Suppl 3), U119.

Goldstein, H. (1994). Multilevel cross-classified models. *J Sociological Methods Research*, *22*(3), 364–375.

Goldstein, H. (2011). *Multilevel statistical models* (Fourth edition). Wiley.

Goldstein, H., Healy, M., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *J Statistics in medicine*, *13*(16), 1643–1655.

Gordon, J., Koutsopoulos, H., & Wilson, N. (2018). Estimation of population origin– interchange–destination flows on multimodal transit networks. *Transportation Research Part C: Emerging Technologies*, *90*, 350–365.

Graser, A., Schmidt, J., Roth, F., & Brändle, N. (2017). Untangling origin-destination flows in geographic information systems. *Information Visualization*, *18*(1), 153–172.

Greenland, S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology*, *29*(1), 158–167.

Griffith, D. (2012). Space, time, and space-time eigenvector filter specifications that account for autocorrelation. *J Estadística española*, *54*(177), 7–34.

Grimm, K., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*(5), 1357–1371.

Guo, D., Zou, Y., Wu, Y., & Liu, M. (2013). Altruistic behavior in complex network based on reputation and future expectation. *2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics*, *2*, 35–38.

Guo, S., Lin, Y., Feng, N., Song, C., Wan, H., & Aaai. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Thirty-Third Aaai Conference on Artificial Intelligence / Thirty-First Innovative Applications of Artificial Intelligence Conference / Ninth Aaai Symposium on Educational Advances in Artificial Intelligence*, 922–929.

Guo, Y., Yang, L., & Gao, J. (2020). Coordinated perimeter control for multiregion heterogeneous networks based on optimized transfer flows. *Mathematical Problems in Engineering*, *2020*.

Hägerstraand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, *24*(1), 6–21.

Hall, K., Albers, H., Taleghan, M., & Dietterich, T. (2018). Optimal spatial-dynamic management of stochastic species invasions. *Environmental & Resource Economics*, *70*(2), 403–427.

Han, X., Hsieh, C., & Ko, S. (2019). Spatial modeling approach for dynamic network formation and interactions. *Journal of Business & Economic Statistics*.

Hawawini, G. (1978). A note on temporal aggregation and serial correlation. *Economics Letters*, *1*(3), 237–242.

He, B., Xu, Z., Xu, Y., Hu, J., & Ma, Z. (2020). Integrating semantic zoning information with the prediction of road link speed based on taxi GPS data. *Complexity*, *2020*.

Head, M., Holman, L., Lanfear, R., Kahn, A., & Jennions, M. (2015). The extent and consequences of P-hacking in science. *PLOS Biology*, *13*(3), e1002106.

Heck, R., & Thomas, S. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus* (Third edition). Taylor & Francis.

Herrett, E., Gallagher, A., Bhaskaran, K., Forbes, H., Mathur, R., van Staa, T., & Smeeth, L. (2015). Data resource profile: Clinical practice research datalink (CPRD). *International Journal of Epidemiology*, *44*(3), 827–836.

Hidas, P., & Wagner, P. (2004). Review of data collection methods for microscopic traffic simulation. *World Conference on Transport Research (WCTR)*.

Higham, D. J., Batty, M., Bettencourt, L. M. A., Greetham, D. V., & Grindrod, P. (2017). An overview of city analytics. *Royal Society Open Science*, *4*, 161063.

Hoffman, M., & 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.

Holmes, C., & Mallick, B. (2003). Generalized nonlinear modeling with multivariate free-knot regression splines. *Journal of the American Statistical Association*, *98*(462), 352–368.

Hong, Y., Escobar, L., & Meeker, W. (2010). Coverage probabilities of simultaneous confidence bands and regions for log-location-scale distributions. *Statistics & Probability Letters*, *80*(7), 733–738.

Horvath, S. (2011). *Weighted network analysis: Applications in genomics and systems biology*. Springer.

Howe, L., Tilling, K., Matijasevich, A., Petherick, E., Santos, A., Fairley, L., Wright, J., Santos, I., Barros, A., Martin, R., Kramer, M., Bogdanovich, N., Matush, L., Barros, H., & Lawlor, D. (2013). Linear spline multilevel models for summarising childhood growth trajectories: A guide to their application using examples from five birth cohorts. *Statistical Methods in Medical Research*, *25*(5), 1854–1874.

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, *24*(3), 383–401.

Hulsegge, G., Spijkerman, A., van der Schouw, Y., Bakker, S., Gansevoort, R., Smit, H., & Verschuren, W. (2017). Trajectories of metabolic risk factors and biochemical markers prior to the onset of type 2 diabetes: The population-based longitudinal Doetinchem study. *Nutr Diabetes*, *7*(5), e270.

Hwang, S. (2000). The effects of systematic sampling and temporal aggregation on discrete time long memory processes and their finite sample properties. *Econometric Theory*, *16*(3), 347–372.

Ibrahim, J., Chu, H., & Chen, L. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, *28*(16), 2796–2801.

Iotti, B., Antonioni, A., Bullock, S., Darabos, C., Tomassini, M., & Giacobini, M. (2017). Infection dynamics on spatial small-world network models. *Physical Review E*, *96*(5), 052316.

Jacoby, D., Brooks, E., Croft, D., & Sims, D. (2012). Developing a deeper understanding of animal movements and spatial dynamics through novel application of network analyses. *Methods in Ecology and Evolution*, *3*(3), 574–583.

Jeong, H., & Lee, L. (2020). Spatial dynamic models with intertemporal optimization: Specification and estimation. *Journal of Econometrics*, *218*(1), 82–104.

Jia, J., Tan, S., Ji, G., & Zhao, B. (2019). Discovering hotspots in dynamic spatial networks using mobility data, 102–107.

Jin, G., Cui, Y., Zeng, L., Tang, H., Feng, Y., & Huang, J. (2020). Urban ride-hailing demand prediction with multiple spatiotemporal information fusion network. *Transportation Research Part C-Emerging Technologies*, *117*.

Jin, Y., Peng, R., & Shi, J. (2019). Population dynamics in river networks. *Journal of Nonlinear Science*, *29*(6), 2501–2545.

John, F. (2009). An introduction to calculus. *A mathematical primer for social statistics*. SAGE Publications, Inc.

Johnson, W. (2014). Analytical strategies in human growth research. *American Journal of Human Biology*, *27*(1), 69–83.

Jones, P. (2017). Markov chains. *Stochastic processes: An introduction*. Chapman & Hall.

Jones, P., & Smith, P. (2018). Markov chains.

Jones, R., & Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics*, *47*(1), 161–175.

Jovanovic, T., Hale, R., Gironas, J., & Mejia, A. (2019). Hydrological functioning of an evolving urban stormwater network. *Water Resources Research*, *55*(8), 6517–6533.

Kaplan, D., Kim, J.-S., & Kim, S.-Y. (2009). Multilevel latent variable modeling: Current research and recent developments. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology*. SAGE Publications Ltd.

Kauermann, G., & Kuhlenkasper, T. (2013). Penalized splines and multilevel models. In M. A. Scott, J. S. Simonoff, & B. D. Marx (Eds.), *The SAGE handbook of multilevel models*. SAGE Publications Ltd.

Ke, J., Zheng, H., Yang, H., & Chen, X. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*, *85*, 591–608.

Kim, K., Sentürk, D., & Li, R. (2011). Recent history functional linear models for sparse longitudinal data. *Journal of statistical planning and inference*, *141*(4), 1554–1566.

Kraak, M.-J. (2003). The space-time cube revisited from a geovisualization perspective. *Proc. 21st International Cartographic Conference*, 1988–1996.

Kuersteiner, G., & Prucha, I. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, *88*(5), 2109–2146.

Kwan, M.-P. (2013). Beyond space (as we knew it): Toward temporally integrated geographies of segregation, health, and accessibility. *Annals of the Association of American Geographers*, *103*(5), 1078–1086.

Leenders, R. (2002). Modeling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks*, *24*(1), 21–47.

Leigh, C., Kandanaarachchi, S., McGree, J., Hyndman, R., Alsibai, O., Mengersen, K., & Peterson, E. (2019). Predicting sediment and nutrient concentrations from high-frequency water-quality data. *J BioRxiv*, 599712.

Lengyel, B., Bokanyi, E., Di Clemente, R., Kertesz, J., & Gonzalez, M. (2020). The role of geography in the complex diffusion of innovations. *Scientific Reports*, *10*(1).

Lenhoff, M., Santner, T., Otis, J., Peterson, M., Williams, B., & Backus, S. (1999). Bootstrap prediction and confidence bands: A superior statistical method for analysis of gait data. *Gait & Posture*, *9*(1), 10–17.

Li, G., Knoop, V., & Lint, H. (2020). Dynamic graph filters networks: A gray-box model for multistep traffic forecasting.

Li, T., & Liao, Q. (2016). Dynamic networks analysis and visualization through spatiotemporal link segmentation. *IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 209–214.

Li, W., Wang, X., Zhang, Y., & Wu, Q. (2021). Traffic flow prediction over muti-sensor data correlation with graph convolution network. *Neurocomputing*, *427*, 50–63.

Li, X., & Griffin, W. (2013). Using ESDA with social weights to analyze spatial and social patterns of preschool children's behavior. *Applied Geography*, *43*, 67–80.

Li, Y., Bi, J., & Sun, H. (2008). Spatial price dynamics: From complex network perspective. *Physica a-Statistical Mechanics and Its Applications*, *387*(23), 5852–5856.

Li, Z., Li, L., Peng, Y., & Tao, X. (2020). A two-stream graph convolutional neural network for dynamic traffic flow forecasting. *2020-November*, 355–362.

Liang, Y., Lu, W., & Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, *65*(2), 377–384.

Liggett, T. (2010). *Continuous time markov processes : An introduction.* American Mathematical Society.

Lin, H., Bai, R., Jia, W., Yang, X., & You, Y. (2020). Preserving dynamic attention for long-term spatial-temporal prediction, 36–46.

Littell Ramon, C., Pendergast, J., & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine*, *19*(13), 1793–1819.

Liu, C., Gao, C., & Xin, Y. (2018). Measuring the diversity and dynamics of mobility patterns using smart card data. *Knowledge Science, Engineering and Management"*, 438–451.

Liu, S., Wang, L., Liu, C., & Destech Publicat, I. (2018). Visualized social network analysis on spatial dynamics of international trade between China and League of Arab States. *2018 3rd international conference on computational modeling, simulation and applied mathematics* (pp. 270–277).

Locantore, N., Marron, J., Simpson, D., Tripoli, N., Zhang, J., Cohen, K., Boente, G., Fraiman, R., Brumback, B., Croux, C., Fan, J., Kneip, A., Marden, J., Peña, D., Prieto, J., Ramsay, J., Valderrama, M., Aguilera, A., Locantore, N., ... Cohen, K. (1999). Robust principal component analysis for functional data. *Test*, *8*(1), 1–73.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique, and future directions. *Statistics in Medicine*, *28*, 3049–3067.

Luptáková, I., & Pospichal, J. (2018). Community detection based clustering. *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*, 000289–000294.

Ma, X., Li, Y., & Chen, P. (2020). Identifying spatiotemporal traffic patterns in large-scale urban road networks using a modified nonnegative matrix factorization algorithm. *Journal of Traffic and Transportation Engineering-English Edition*, *7*(4), 529–539.

Malaba, T., Phillips, T., Le Roux, S., Brittain, K., Zerbe, A., Petro, G., Ronan, A., McIntyre, J., Abrams, E., & Myer, L. (2017). Antiretroviral therapy use during pregnancy and adverse birth outcomes in South African women. *International Journal of Epidemiology*, *46*(5), 1678–1689.

Mateus-Anzola, J., Wiratsudakul, A., Rico-Chavez, O., & Ojeda-Flores, R. (2019). Simulation modeling of influenza transmission through backyard pig trade networks in a wildlife/livestock interface area. *Tropical Animal Health and Production*, *51*(7), 2019–2024.

Matthews, A., Herrett, E., Gasparrini, A., Van Staa, T., Goldacre, B., Smeeth, L., & Bhaskaran, K. (2016). Impact of statin related media coverage on use of statins: Interrupted time series analysis with uk primary care data. *BMJ*, *353*, i3283.

McLachlan, G., & Peel, D. (2000). Hidden markov models. *Finite mixture models* (pp. 326–341).

Mehta, P., & West, S. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*(1), 23–43.

Mendoza-Cota, J., & Torres-Preciado, V. (2019). The impact of regional remittances on economic growth in mexico: A dynamic space-time panel approach. *Papeles De Poblacion*, *25*(101), 113–144.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122.

Morris, T., White, I., & Crowther, M. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102.

Muthén, L., & Muthén, B. (2010). *Mplus user's guide* (Sixth edition).

Muthukrishnan, R., & Rohini, R. (2016). LASSO: A feature selection technique in predictive modeling for machine learning. *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, 18–20.

Naimi, A., Cole, S., & Kennedy, E. (2017). An introduction to g methods. *International Journal of Epidemiology*, *46*(2), 756–762.

Nassir, N., Hickman, M., & Ma, Z.-L. (2015). Activity detection and transfer identification for public transit fare card data. *Transportation*, *42*(4), 683–705.

National Institute for Health and Care Excellence. (2016). Clinical knowledge summary: Diabetes - type 2. https://cks.nice.org.uk/diabetes-type-2#!diagnosissub

Neale, M., Boker, S., Xie, G., & Maes, H. (2003). *Mx: Statistical modeling.* (Sixth edition).

Neeson, T., Wiley, M., Adlerstein, S., & Riolo, R. (2012). How river network structure and habitat availability shape the spatial dynamics of larval sea lampreys. *Ecological Modelling*, *226*, 62–70.

Neto, P., Friesz, T., & Han, K. (2016). Electric power network oligopoly as a dynamic stackelberg game. *Networks and Spatial Economics*, *16*(4), 1211–1241.

Neumann, M., & Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, *9*(4), 307–333.

Newman, M. (2002). The structure and function of networks. *Computer Physics Communications*, *147*(1), 40–45.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256.

Newman, M. (2018). Fundamentals of network theory. *Networks* (Second edition). Oxford University Press.

Niezink, N., Snijders, T., & van Duijn, M. (2019). No longer discrete: Modeling the dynamics of social networks and continuous behavior. *Sociological Methodology*, *49*(1), 295–340.

Oka, R., Shibata, K., Sakurai, M., Kometani, M., Yamagishi, M., Yoshimura, K., & Yoneda, T. (2017). Trajectories of postload plasma glucose in the development of type 2 diabetes in Japanese adults. *J Diabetes Res*, *2017*, 5307523.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251.

Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks*, *31*(2), 155–163.

Ou, J., Sun, J., Zhu, Y., Jin, H., Liu, Y., Zhang, F., Huang, J., & Wang, X. (2020). STP-TrellisNets: Spatial-temporal parallel trellisnets for metro station passenger flow prediction, 1185–1194.

Oud, J., Folmer, H., Patuelli, R., & Nijkamp, P. (2012). Continuous-time modeling with spatial dependence. *Geographical Analysis*, *44*(1), 29–46.

Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *J Evaluation review*, *30*(1), 66–85.

Parent, O., & LeSage, J. (2010). A spatial dynamic panel model with random effects applied to commuting times. *Transportation Research Part B-Methodological*, *44*(5), 633–645.

Patuelli, R., Griffith, D., Tiefelsdorf, M., & Nijkamp, P. (2006). The use of spatial filtering techniques: The spatial and space-time structure of german unemployment data. *Tinbergen Institute Discussion Papers*.

Peirce, S., & Lindsay, J. (2015). Characterizing ephemeral streams in a southern Ontario watershed using electrical resistance sensors. *Hydrological Processes*, *29*(1), 103–111.

Peng, H., Wang, H., Du, B., Bhuiyan, M., Ma, H., Liu, J., Wang, L., Yang, Z., Du, L., Wang, S., & Yu, P. (2020). Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. *Information Sciences*, *521*, 277–290.

Perez, L., & Dragicevic, S. (2009). An agent-based approach for modeling dynamics of contagious disease spread. *International Journal of Health Geographics*, *8*(1).

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. *BMC Medical Research Methodology*, *19*(1), 46.

Perry, G., & Lee, F. (2019). How does temporal variation in habitat connectivity influence metapopulation dynamics? *Oikos*, *128*(9), 1277–1286.

Peterson, E., & Ver Hoef, J. (2010). A mixed-model moving-average approach to geostatistical modeling in stream networks. *J Ecology*, *91*(3), 644–651.

Pfeifer, P., & Deutsch, S. (1980). A STARIMA model-building procedure with application to description and regional forecasting. *J Transactions of the Institute of British Geographers*, 330–349.

Piccininni, M., Konigorski, S., Rohmann, J., & Kurth, T. (2020). Directed acyclic graphs and causal thinking in clinical risk prediction modeling. *BMC Medical Research Methodology*, *20*(1), 179.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. C. (2018). nlme: Linear and nonlinear mixed effects models.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*, *124*.

Pons, P., & Latapy, M. (2005). Computing communities in large networks using random walks, 284–293.

Porta, M. (2014). *A dictionary of epidemiology*. Oxford University Press, Incorporated.

Pullenayegum, E., & Lim, L. (2016). Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. *25*(6), 2992–3014.

Qu, Y., Zhu, Y., Zang, T., Xu, Y., & Yu, J. (2020). Modeling local and global flow aggregation for traffic flow forecasting. https://doi.org/10.1007/978-3-030-62005-9_30

R Core Team. (2020a). R: A language and environment for statistical computing.

R Core Team. (2020b). R: The exponential distribution. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/Exponential.html

Rabash, J., Steele, F., Browne, W., & Goldstein, H. (2020). *A user's guide to MLwiN, v3.05*. Centre for Multilevel Modelling, University of Bristol.

Ramachandran, K., Sikdar, B., & Ieee. (2007a). Modeling malware propagation in networks of smart cell phones with spatial dynamics. *IEEE infocom 2007* (pp. 2516–2520).

Ramachandran, K., Sikdar, B., & Ieee. (2007b). On the stability of the malware free equilibrium in cell phones networks with spatial dynamics. *2007 IEEE International Conference on Communications*, *1-14*, 6169–6174.

Ramsay, J., & Silverman, B. (1997). *Functional data analysis*. Springer.

Rasmussen, P., Staggs, M., Beard Jr., T., & Newman, S. (1998). Bias and confidence interval coverage of creel survey estimators evaluated by simulation. *127*(3), 469–480.

Rossana, R., & Seater, J. (1995). Temporal aggregation and economic time series. *Journal of Business & Economic Statistics*, *13*(4), 441–451.

Royston, P., & Altman, D. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *43*(3), 429–467.

Russo, F. (2012). Public health policy, evidence, and causation: Lessons from the studies on obesity. *Medicine, Health Care and Philosophy*, *15*(2), 141–151.

Ryser, R., Haussler, J., Stark, M., Brose, U., Rall, B., & Guill, C. (2019). The biggest losers: Habitat isolation deconsructs complex food webs from top to bottom. *Proceedings of the Royal Society B-Biological Sciences*, *286*(1908).

Saberi, M., Mahmassani, H., Brockmann, D., & Hosseini, A. (2017). A complex network perspective for characterizing urban travel demand patterns: Graph theoretical analysis of large-scale origin–destination demand networks. *Transportation*, *44*(6), 1383–1402.

Sabia, S., Dugravot, A., Dartigues, J., Abell, J., Elbaz, A., Kivimaki, M., & Singh-Manoux, A. (2017). Physical activity, cognitive decline, and risk of dementia: 28 year follow-up of Whitehall II cohort study. *Bmj-British Medical Journal*, *357*.

Sayers, A., Heron, J., Smith, A., Macdonald-Wallis, C., Gilthorpe, M., Steele, F., & Tilling, K. (2017). Joint modelling compared with two stage methods for analysing longitudinal data and prospective outcomes: A simulation study of childhood growth and BP. *Stat Methods Med Res*, *26*(1), 437–452.

Scott, J. (1988). Social network analysis. *Sociology*, *22*(1), 109–127.

Semboloni, F. (2000). The growth of an urban cluster into a dynamic self-modifying spatial pattern. *Environment and Planning B: Planning and Design*, *27*(4), 549–564.

Shoaff, J., Papandonatos, G., Calafat, A., Ye, X., Chen, A., Lanphear, B., Yolton, K., & Braun, J. (2017). Early-life phthalate exposure and adiposity at 8 years of age. *Environmental Health Perspectives*, *125*(9).

Shrier, I., & Platt, R. (2008). Reducing bias through directed acyclic graphs. *BMC Medical Research Methodology*, *8*(1), 70.

Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*, *22*(11), 1359–66.

Smith, A., Hardy, R., Heron, J., Joinson, C., Lawlor, D., Macdonald-Wallis, C., & Tilling, K. (2016). A structured approach to hypotheses involving continuous exposures over the life course. *International Journal of Epidemiology*, *45*(4), 1271–1279.

Smith, A., Heron, J., Mishra, G., Gilthorpe, M., Ben-Shlomo, Y., & Tilling, K. (2015). Model selection of the effect of binary exposures over the life course. *Epidemiology*, *26*(5), 719–26.

Soetaert, K., & Herman, P. (2009). rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations.

Song, X., Mu, X., & Sun, L. (2012). Regression analysis of longitudinal data with time-dependent covariates and informative observation times. *Scandinavian Journal of Statistics*, *39*(2), 248–258.

Soundararaj, B., Cheshire, J., & Longley, P. (2019). Medium data toolkit - A case study on smart street sensor project. *GISRUK*.

Sperrin, M., Petherick, E., & Badrick, E. (2017). Informative observation in health data: Association of past level and trend with time to next measurement. *Stud Health Technol Inform*, *235*, 261–265.

Spiegelhalter, D., Best, N., Carlin, B., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.

Srivastava, A., & Salapaka, S. (2020). Simultaneous facility location and path optimization in static and dynamic networks. *IEEE Transactions on Control of Network Systems*, *7*(4), 1700–1711.

Stan Development Team. (2019). RStan: The R interface to stan. R package version 2.19.2.

Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *171*(1), 5–19.

Sterba, S. (2014). Fitting nonlinear latent growth curve models with individually varying time points. *Structural Equation Modeling-a Multidisciplinary Journal*, *21*(4), 630–647.

Stimson, J., Carmines, E., & Zeller, R. (1978). Interpreting polynomial regression. *Sociological Methods & Research*, *6*(4), 515–524.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, *12*(3), 1–16.

Sun, B., Zhao, D., Shi, X., & He, Y. (2021). Modeling global spatial-temporal graph attention network for traffic prediction. *IEEE Access*, *9*, 8581–8594.

Sun, G., Tang, T., Peng, T., Liang, R., & Wu, Y. (2017). SocialWave: Visual analysis of spatio-temporal diffusion of information on social media. *ACM Transactions on Intelligent Systems and Technology*, *9*(2), 2157–6904.

Tabák, A., Jokela, M., Akbaraly, T., Brunner, E., Kivimäki, M., & Witte, D. (2009). Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: An analysis from the Whitehall II study. *The Lancet*, *373*(9682), 2215–2221.

Thomas, A., Best, N., Lunn, D., Arnold, R., & Spiegelhalter, D. (2014). GeoBUGS user manual.

Thompson, S. (1919). *Calculus made easy: Being a very-simplest introduction to those beautiful methods of rekoning which are generally called by the terrifying names of the differential calculus and the integral calculus* (Second edition). Macmillan.

Tian, K., Guo, J., Ye, K., & Xu, C. (2020). ST-MGAT: Spatial-temporal multi-head graph attention networks for traffic forecasting. *2020-November*, 714–721.

Tilling, K., Davies, N., Nicoli, E., Ben-Shlomo, Y., Kramer, M., Patel, R., Oken, E., & Martin, R. (2011). Associations of growth trajectories in infancy and early childhood with later childhood outcomes. *The American Journal of Clinical Nutrition*, *94*(6 Suppl), 1808S–1813S.

Tobler, W. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, *46*, 234–240.

Transport for London. (2009a). London Underground map December 2009.

Transport for London. (2009b). London Underground map September 2009.

Transport for London. (2020a). Our open data. https://tfl.gov.uk/info-for/open-data-users/our-open-data

Transport for London. (2020b). Transport for London unified API. https://api.tfl.gov.uk/

Tu, Y.-K., D'Aiuto, F., Baelum, V., & Gilthorpe, M. (2009). An introduction to latent growth curve modelling for longitudinal continuous data in dental research. *European Journal of Oral Sciences*, *117*(4), 343–350.

Tu, Y.-K., & Gilthorpe, M. (2010). *Statistical thinking in epidemiology*. Chapman & Hall/CRC.

Tu, Y.-K., Tilling, K., Sterne, J., & Gilthorpe, M. (2013). A critical evaluation of statistical approaches to examining the role of growth trajectories in the developmental origins of health and disease. *International Journal of Epidemiology*, *42*(5), 1327–1339.

Ugrinowitsch, C., Fellingham, G., & Ricard, M. (2004). Limitations of ordinary least squares models in analyzing repeated measures data. *36*(12), 2144–2148.

Ullah, S., & Finch, C. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, *13*(1), 43.

Van Kampen, N. (1976). Stochastic differential equations. *Physics Reports*, *24*(3), 171–228.

Van Wissen, L., & Meurs, H. (1989). The Dutch mobility panel: Experiences and evaluation. *Transportation*, *16*(2), 99–119.

Ver Hoef, J., & Peterson, E. (2010). A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association*, *105*(489), 6–18.

Voelkle, M., Oud, J., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*(2), 176–192.

Wang, J.-L., Chiou, J.-M., & Müller, H.-G. (2016). Functional data analysis. *3*(1), 257–295.

Wang, W., & Yan, J. (2020). splines2: Regression spline functions and classes.

Wang, X., Ge, J., Wei, W., Li, H., Wu, C., & Zhu, G. (2016). Spatial dynamics of the communities and the role of major countries in the international rare earths trade: A complex network analysis. *PLOS One*, *11*(5), e0154575.

Wanzenböck, I., & Piribauer, P. (2018). R&d networks and regional knowledge production in europe: Evidence from a space-time model. *Papers in Regional Science*, *97*, S1–S24.

Ward, J., Evans, A., & Malleson, N. (2016). Dynamic calibration of agent-based models using data assimilation. *Royal Society Open Science*, *3*(4), 150703.

Wei, V., Wong, R., & Long, C. (2020). Architecture-intact oracle for fastest path and time queries on dynamic spatial networks, 1841–1856.

Wei, W. (1981). Effect of systematic sampling on ARIMA models. *Communications in Statistics - Theory and Methods*, *10*(23), 2389–2398.

Weiss, A. (1984). Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics*, *26*(3), 271–281.

Wilkinson, J., Arnold, K., Murray, E., van Smeden, M., Carr, K., Sippy, R., de Kamps, M., Beam, A., Konigorski, S., Lippert, C., Gilthorpe, M., & Tennant, P. (2020). Time to reality check the promises of machine learning-powered precision medicine. *The Lancet Digital Health*, *2*(12), e677–e680.

Wolf, C., & Best, H. (2014a). Linear regression. In C. Wolf & H. Best (Eds.), *The SAGE handbook of regression analysis and causal inference*. SAGE Publications Ltd.

Wolf, C., & Best, H. (2014b). Non-linear and non-additive effects in linear regression. In C. Wolf & H. Best (Eds.), *The SAGE handbook of regression analysis and causal inference*. SAGE Publications Ltd.

Wolf, C., & Best, H. (2014c). Regression analysis: Assumptions and diagnostics. In C. Wolf & H. Best (Eds.), *The SAGE handbook of regression analysis and causal inference*. SAGE Publications Ltd.

Wood, J., Dykes, J., & Slingsby, A. (2010). Visualisation of origins, destinations and flows with OD maps. *The Cartographic Journal*, *47*(2), 117–129.

World Health Organisation. (2015). *The case for investing in public health* (Report).

Wu, J., Roy, J., & Stewart, W. (2010). Prediction modeling using EHR data: Challenges, strategies, and a comparison of machine learning approaches. *Medical Care*, *48*(6), S106–S113.

Xiao, N. (2016). Basic geometric operations. *GIS algorithms: Theory and applications for geographic information science & technology*. SAGE Publications Ltd.

Xu, M., & Gao, Z. (2009). Chaos in a dynamic model of urban transportation network flow based on user equilibrium states. *Chaos Solitons & Fractals*, *39*(2), 586–598.

Yang, B., Kang, Y., Li, H., Zhang, Y., Yang, Y., & Zhang, L. (2020). Spatio-temporal expand-and-squeeze networks for crowd flow prediction in metropolis. *IET Intelligent Transport Systems*, *14*(5), 313–322.

Yang, Y., Heppenstall, A., Turner, A., & Comber, A. (2019). A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. *Computers, Environment and Urban Systems*, *77*, 101361.

Yao, H., Tang, X., Wei, H., Zheng, G., Li, Z., & Aaai. (2019). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. *Thirty-Third Aaai Conference on Artificial Intelligence / Thirty-First Innovative Applications of Ar-*

tificial Intelligence Conference / Ninth Aaai Symposium on Educational Advances in Artificial Intelligence, 5668–5675.

Yeh, R., Nashed, Y., Peterka, T., & Tricoche, X. (2020). Fast automatic knot placement method for accurate b-spline curve fitting. *Computer-Aided Design*, *128*, 102905.

Yu, H., Fang, Z., Lu, F., Murray, A., Zhao, Z., Xu, Y., & Yang, X. (2019). Massive automatic identification system sensor trajectory data-based multi-layer linkage network dynamics of maritime transport along 21st-century maritime silk road. *Sensors*, *19*(19).

Yu, W., Guan, M., & Chen, Z. (2019). Analyzing spatial community pattern of network traffic flow and its variations across time based on taxi gps trajectories. *Applied Sciences-Basel*, *9*(10).

Yuan, Y., Chen, N., & Zhou, S. (2013). Adaptive b-spline knot selection using multi-resolution basis set. *IIE Transactions*, *45*(12), 1263–1277.

Zambrano-Monserrate, M., Ruano, M., Ormeño-Candelario, V., & Sanchez-Loor, D. (2020). Global ecological footprint and spatial dependence between countries. *Journal of Environmental Management*, *272*.

Zhang, C., Yu, J., & Liu, Y. (2019). Spatial-temporal graph attention networks: A deep learning approach for traffic forecasting. *IEEE Access*, *7*, 166246–166256.

Zhang, H., Zhang, L., Che, F., Jia, J., & Shi, B. (2020). Revealing urban traffic demand by constructing dynamic networks with taxi trajectory data. *IEEE Access*, *8*, 147673–147681.

Zhang, W., Leng, C., & Tang, C. (2015). A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *77*(1), 219–238.

Zhang, W., Tian, Z., Zhang, G., & Dong, G. (2019). Spatial-temporal characteristics of green travel behavior based on vector perspective. *Journal of Cleaner Production*, *234*, 549–558.

Zhang, X., Chen, G., Wang, J., Li, M., & Cheng, L. (2019). A GIS-based spatial-temporal autoregressive model for forecasting marine traffic volume of a shipping network. *Scientific Programming*, *2019*.

Zhang, Z., Li, M., Lin, X., Wang, Y., & He, F. (2019). Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies. *Transportation Research Part C-Emerging Technologies*, *105*, 297–322.

Zhao, N., Ye, Z., Pei, Y., Liang, Y., & Niyato, D. (2020). Spatial-temporal attention-convolution network for citywide cellular traffic prediction. *IEEE Communications Letters*, *24*(11), 2532–2536.

Zhao, X. [X.Y.], & Pu, Y. (2018). The complex system of interprovincial migration flows in China, 1985-2015: A spatial dynamic panel modeling approach. In S. Hu, X. Ye, K. Yang, & H. Fan (Eds.), *2018 26th international conference on geoinformatics*.

Zheng, L., Yang, J., Chen, L., Sun, D., & Liu, W. (2020). Dynamic spatial-temporal feature optimization with ERI big data for short-term traffic flow prediction. *Neurocomputing*, *412*, 339–350.

Zhou, F., Yang, Q., Zhang, K., Trajcevski, G., Zhong, T., & Khokhar, A. (2020). Reinforced spatiotemporal attentive graph neural networks for traffic forecasting. *IEEE Internet of Things Journal*, *7*(7), 6414–6428.

Zhou, S., & Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association*, *96*(453), 247–259.

Zhu, X., & Guo, D. (2014). Mapping large spatial flow data with hierarchical clustering. *18*(3), 421–435.

Zoppou, C. (2001). Review of urban storm water models. *Environmental Modelling & Software*, *16*(3), 195–231.

Zyphur, M., & Oswald, F. (2015). Bayesian estimation and inference: A user's guide. *Journal of management*, *41*(2), 390–420.

Appendix A

Table of reviewed papers

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
A. F. Abdelghany, H. S. Mahmassani and Y. C. Chiu	2001	Spatial microassignment of travel demand with activity trip chains	Transport	Prediction	Stochastic simulation	Discrete, Continuous	Hours	Partial city	Actor	Actor, Component	Simulation works in steps, but expected trip time is continuous
H. J. Albers, K. M. Hall, K. D. Lee, M. A. Taleghan and T. G. Dietterich	2018	The Role of Restoration and Key Ecological Invasion Mechanisms in Optimal Spatial-Dynamic Management of Invasive Species	Ecology, Economics	Hypothetical	Stochastic simulation	Discrete	Unspecified	Unspecified	Actor	Component	One of a number of invasive species management - incorporated ecological and economic models for 7 abstract river reaches
T. Anderson and S. Dragićević	2018	A Geographic Network Automata Approach for Modeling Dynamic Ecological Systems	Ecology	Hypothetical	Stochastic simulation	Discrete	Years	Regional	Component	Whole network	Insect infestation trees - network evolution, node (tree) properties determine whether link made by insects
T. M. Anderson and S. Dragicevic	2018	Network-agent based model for simulating the dynamic spatial network structure of complex ecological systems	Ecology	Hypothetical	Stochastic simulation	Discrete	Months	Town	Actor	Component, Whole network	Beetles and trees, links are movemnt of beetle from tree I to j, trees and beetles both agents
T. Anderson and S. Dragićević	2020	NEAT approach for testing and validation of geospatial network agent-based model processes: case study of influenza spread	Epidemiology	Hypothetical	Stochastic simulation	Discrete	Months	Partial city	Component	Whole network	Model validation approach for ABMS. Influenza spread. Agents are adults, childs and places (homes, schools, businesses). Part of vancouver. 270 days, 6480 time steps (hours). Looks like model validation includes network measures.
T. Anderson and S. Dragićević	2020	Representing complex evolving spatial networks: Geographic network automata	Network methods	Hypothetical	Stochastic simulation	Discrete	Abstract	Abstract	Component	Whole network	Geographic network automata to represent evolution of networks. Several examples in different areas. Transition rules and connection costs

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
R. Basile, A. Parteka and R. Pittiglio	2018	Export diversification and economic development: A dynamic spatial data analysis	Economics	Relationships	Statistical model	Discrete	Years	Multi-country	Component	Component	Spatial spillover in neighbour relations was network components. Dynamic spatial panel data model
B. Baybeck and R. Huckfeldt	2002	Urban contexts, spatially dispersed networks, and the diffusion of political information	Sociology	Description	Statistical model	Discrete	Months	Regional	Component	Whole network	Diffusion of information (political). MLM local net characteristics for individuals and who they talk to
J. W. Burnett and X. Zhao	2017	Spatially Explicit Prediction of Wholesale Electricity Prices	Economics	Prediction	Statistical model	Discrete	Years	Multi-region	Component	Component	Predicting electricity prices with spatial and non-spatial panel data. Electricity network provides network basis. Spatial autoregressive model with temporal lag
H. Cai, A. Rahman, X. Su and H. Zhang	2014	A GIS-microscopic simulation approach for optimizing road barrier placement and configuration in university campus emergency evacuation	Transport, Urban planning	Hypothetical	Stochastic simulation	Discrete	Hours	Partial city	Actor	Component, Whole network	Sort of disaster management. Series of simulations estimating travel demand based on emergency timing (on a uni campus) and then optimising placement of road barriers to get people out. Road network.
A. Carlsson and O. Kindvall	2001	Spatial dynamics in a metapopulation network: recovery of a rare grasshopper Stauroderus scalaris from population refuges	Ecology	Hypothetical	Stochastic simulation	Discrete	Years	Single country	Component	Component	4 meaures of patchy grasshopper population over years. Metapopulatrion model - spatial dynamics, habitat mosaic. Species management option conseuences.

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
N. H. Carter, A. Baeza and N. R. Magliocca	2020	Emergent conservation outcomes of shared risk perception in human-wildlife systems	Ecology	Hypothetical	Stochastic simulation	Discrete	Years	Regional	Component, Whole-network	Whole network	Human wildlife systems (conservation outcomes). WHISL model - wildlife human interactions in shared landscapes. Social network influence on human response to encounters. Agent-based model. Changing landscape and social influence to test outcome - wildlife distributions damages and habitat quality
M. Caruso	2007	Disruptive dynamics: The spatial dimensions of the Spanish networks in the spread of monitorial schooling (1815-1825)	History, Sociology	Description	Visualisation	Discrete	Years	Single country	Component	Whole network	Visualisation of network of people involved in monitorial schooling in spain in history. Written description of process, network maps used as illustration
S. Chen, A. Ilany, B. J. White, M. W. Sanderson and C. Lanzas	2015	Spatial-Temporal Dynamics of High-Resolution Animal Networks: What Can We Learn from Domestic Animals?	Ecology	Description	Discrete network methods	Discrete	Days	Small area	Component	Whole network	Cow interactions in different spatial areas of pen. Looked at network properties at different points in time from aggregated interactions between cows
L. Chen, K. Han, Q. Yin and Z. Cao	2020	GDCRN: Global Diffusion Convolutional Residual Network for Traffic Flow Prediction	Transport	Prediction	Machine learning	Discrete	Months	Partial city	Component	Component	Traffic flow prediction machine learning. "Local and global spatio-temporal dependencies". Time slices
T. Cheng, J. Wang, J. Haworth, B. Heydecker and A. Chow	2014	A Dynamic Spatial Weight Matrix and Localized Space-Time Autoregressive Integrated Moving Average for Network Modeling	Transport	Prediction	Statistical model	Discrete	Months	Partial city	Component	Component	L-STARIMA dynamic spatial weight matrix depends on current traffic conditions.

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
S. Cheng, B. Xie, Y. Bie, Y. Zhang and S. Zhang	2018	Measure dynamic individual spatial-temporal accessibility by public transit: Integrating time-table and passenger departure time	Transport	Description	Discrete network methods	Discrete	Days	Citywide	Component	Component	Public transport accessibility based on current state of network - travel time for each edge depends on time of day. Network potential path area for origin destination and time budget
S. Colak, C. M. Schneider, P. Wang and M. C. Gonzalez	2013	On the role of spatial dynamics and topology on network flows	Transport	Hypothetical	Stochastic simulation	Discrete	Hours	Citywide	Actor, Component	Component, Whole network	Network topology effect on network particle flows and when reach congestion on differnet links. Map shows WHEN CONGESTION REACHED. Very small time intervals of 10sec
Z. L. Diao, X. Wang, D. F. Zhang, Y. R. Liu, K. Xie, S. Y. He and Aaai	2019	Dynamic Spatial-Temporal Graph Convolutional Neural Networks for Traffic Forecasting	Transport	Prediction	Machine learning	Discrete	Months	Partial city	Actor	Component	Graph convolutional neural networks modelling dynamic spatial dependence - related to structural changes - estimate matrix at different times.
H. P. P. Donfouet, P. W. Jeanty and E. Malin	2018	Analysing spatial spillovers in corruption: A dynamic spatial panel data approach	Economics, Sociology	Relationships	Statistical model	Discrete	Years	Multi-country	Component	Whole network	Dynamic spatial panel data model - nearest neighbour relationship matrix
X. H. Fan, X. X. Li, J. L. Yin and J. C. Liang	2019	Temporal Characteristics and Spatial Homogeneity of Virtual Water Trade: A Complex Network Analysis	Economics	Description	Discrete network methods	Discrete	Years	Multi-country	Component	Whole network	Annual network properties of trade network
S. Fang, Q. Zhang, G. Meng, S. Xiang and C. Pan	2019	GSTnet: Global spatial-temporal network for traffic flow prediction	Transport	Prediction	Machine learning	Discrete	Months	Citywide	Actor, Component	Component	Predicting traffic flow, considering global and local correlations. Tested model on taxi GPS, Beijing bus and Beijing subway.

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
Y. Fang, Z. Wang, R. Cheng, X. Li, S. Luo, J. Hu and X. Chen	2019	On Spatial-Aware Community Search	Sociology	Description	Continous network methods	Continuous	Years	Regional	Component	Component	Spatial aware community search for node that moves, approximate running SACS at every location change (in response to event rather than aggregated)
D. Feng, Z. Wu, J. Zhang and Z. Wu	2020	Dynamic Global-Local Spatial-Temporal Network for Traffic Speed Prediction	Transport	Prediction	Machine learning	Discrete	Months	Partial city, Regional	Component	Component	Traffic prediction with global and local information. Traffic speed prediction.
S. Ferenc, P. Tamas, B. Jozsef and Asme	2016	COMPLEX ANALYSIS OF THE DYNAMIC EFFECTS OF CAR POPULATION ALONG THE TRAJECTORIES	Transport	Prediction	Stochastic simulation	Discrete	Days	Citywide	Actor	Actor	Compled model for traffic conditions. Using a model of network state to look at indiviudal trajectories and predict accidents.
S. Garbin, E. A. Celegon, P. Fanton and G. Botter	2019	Hydrological controls on river network connectivity	Ecology	Hypothetical	Stochastic simulation	Discrete	Abstract	Abstract	Whole-network	Component, Whole network	Connectivity in river network. Causes and consequences of depth dynamics in river networks. Seasonal probability distro of streamflows. N rain events per unit time/area (poisson). Calculate streamflow pointwise and temporal and spatial variability is derived.
D. Guo, Y. Zou, Y. Wu and M. Liu	2013	Altruistic behavior in complex network based on reputation and future expectation	Economics	Hypothetical	Stochastic simulation	Discrete	Unspecified	Unspecified	Component	Whole network	Economic game to do with neighbour cooperation - neighbours in grid and allows selection of another connection not immediately adjacent. Evoluton of cooperation over time, time slice visualisation

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
S. N. Guo, Y. F. Lin, N. Feng, C. Song, H. Y. Wan and Aaai	2019	Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting	Transport	Prediction	Machine learning	Discrete	Months	Partial city	Actor	Component	Traffic flow prediction - attention based spatio-temporal graph Convolutional Neural Network. Capture dynamic spatio-temporal correlations and spatial patterns and temporal features. Recent effects, daily periodic effects, weekly effects. Frequent time slices - simultaneous measurements.
Y. J. Guo, L. C. Yang and J. Gao	2020	Coordinated Perimeter Control for Multiregion Heterogeneous Networks Based on Optimized Transfer Flows	Transport	Hypothetical	Stochastic simulation	Discrete	Hours	Partial city	Actor, Component	Actor, Component, Whole network	Congestion control. 'perimiter control strategy'. Testing out method using simulation in case study.
K. M. Hall, H. J. Albers, M. A. Taleghan and T. G. Dietterich	2018	Optimal Spatial-Dynamic Management of Stochastic Species Invasions	Ecology, Economics	Hypothetical	Stochastic simulation	Discrete	Years	Unspecified	Actor, Component	Component	Invasive species management on stylised river network with spatial heterogeneity, competition and management actions included in model. Node = reach of river= management unit
X. Y. Han, C. S. Hsieh and S. I. M. Ko	2019	Spatial Modeling Approach for Dynamic Network Formation and Interactions	Sociology	Relationships	Statistical model	Discrete	Years	Unspecified	Component	Component	Longitudinal net data and activity outcomes for high school student friendship. Waves of network and general questionnaires. Peer effects on academic performance (related to outdegree) Difficulty due to different time periods for 2 questionnaires. Spatial dynamic panel data model.

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
B. He, Z. F. Xu, Y. J. Xu, J. X. Hu and Z. W. Ma	2020	Integrating Semantic Zoning Information with the Prediction of Road Link Speed Based on Taxi GPS Data	Transport	Prediction	Machine learning	Discrete	Days	Citywide	Actor	Component	Predicting traffic speed, global and local correlations
B. Iotti, A. Antonioni, S. Bullock, C. Darabos, M. Tomassini and M. Giacobini	2017	Infection dynamics on spatial small-world network models‘	Epidemiology, Network methods	Hypothetical	Stochastic simulation	Discrete	Abstract	Unspecified	Component	Whole network	Uses infection dynamics as example. Model for generating synthetic network (REDS) with two types fo rewiring to get small world network. Shared neighbours and distance influence connection cost, each node has finite social energy. Tested infection model.
D. M. P. Jacoby, E. J. Brooks, D. P. Croft and D. W. Sims	2012	Developing a deeper understanding of animal movements and spatial dynamics through novel application of network analyses	Ecology	Description	Discrete network methods, Statistical model	Discrete	Months	Small area	Actor	Component	Anamial movements through network analysis application. Electronic tag data on sharks. Location defined by detector location and range, connections are shark movement. Compared node properties to random network.
H. Jeong and L. F. Lee	2020	Spatial dynamic models with intertemporal optimization: Specification and estimation	Economics	Relationships	Statistical model	Discrete	Years	Regional	Component	Whole network	Spatial dynamic model - alternative to traditional SDPD model. Counties' public safety spending in North Carolina - spillover effects considering 'forward looking' decision making

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
J. Jia, S. Tan, G. Ji and B. Zhao	2019	Discovering hotspots in dynamic spatial networks using mobility data	Network methods, Transport	Description	Discrete network methods	Discrete	Hours	Citywide	Actor	Component, Whole network	Hotspots of dense activity - find these using mobility data. Edges have weight sequence for discreteintervals. Hotspots are connected subgraphs with edges that have been increasingly crowded for several time intervals. 'Closed hotspot' contains all the increasingly crowded edges.
Y. Jin, R. Peng and J. P. Shi	2019	Population Dynamics in River Networks	Ecology	Description	Statistical model	Continuous	Abstract	Abstract	Component	Component, Whole network	River network represented as a graph, Population moves in a river network.
G. Y. Jin, Y. Cui, L. Zeng, H. B. Tang, Y. H. Feng and J. C. Huang	2020	Urban ride-hailing demand prediction with multiple spatiotemporal information fusion network	Transport	Prediction	Machine learning	Discrete	Months, Years	Citywide	Actor	Component	Predicinting ride hailing demand in taxis using oroigin destination data.
T. Jovanovic, R. L. Hale, J. Gironas and A. Mejia	2019	Hydrological Functioning of an Evolving Urban Stormwater Network	Urban planning	Description	Discrete network methods, Stochastic simulation	Discrete	Years	Partial city	Actor	Component, Whole network	Event based distributed hydrological model. Simulate hydrological flixes and states in the catchment. Model outputs analysed as network. Simulations based on previous rainfall data - so simulating changes in the network based on stuff that actually happened.
G. M. Kuersteiner and I. R. Prucha	2020	Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity	Network methods	Relationships	Statistical model	Discrete	Abstract	Abstract	Component	Whole network	Time varying spatial weight matrices in dynamic spatial panel data models - for data with network structure. No network formation model.

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
B. Lengyel, E. Bokanyi, R. Di Clemente, J. Kertesz and M. C. Gonzalez	2020	The role of geography in the complex diffusion of innovations	Sociology	Prediction, Description	Discrete network methods, Statistical model, Stochastic simulation	Discrete, Continuous	Abstract	Single country	Component	Whole network	Social media platform in hungary 2002-2014. Compares a differential equation to predict adoption of social media platform in different towns with an Agent based model.
Y. L. Li, J. T. Bi and H. J. Sun	2008	Spatial price dynamics: From complex network perspective	Economics	Hypothetical	Stochastic simulation	Discrete	Abstract	Unspecified	Whole network	Component	Spatial price problem. How trade network structure affects process when the rules for this problem used for simulation. Look at changes in price, when reaches equilibrium and different community structures.
X. Li and W. A. Griffin	2013	Using ESDA with social weights to analyze spatial and social patterns of preschool children's behavior	Sociology	Relationships	Statistical model	Discrete	Months	Small area	Component	Component, Whole network	Classroom children behaviour. Had 4 social types of behaviour and 15 activities along with a range of locations. Divided into spatial grid. Social network interactions noted for social network autocorrelations. Exploratory spatial data analysis with social weights to look at spatial autocorrelation of types of social behvaiouors and different activities.
T. Li and Q. Liao	2016	Dynamic networks analysis and visualization through spatiotemporal link segmentation	Network methods	Description	Visualisation	Discrete	Days	Unspecified	Component	Component	Visualising dynamic networks with spatio-temporal link segmentations. Links divided according to spatio-temp dimentsions to avoid animation. Segments also subdivided in same axis to show multiple properties which is kind of confusing. Time period is discrete. No consistent time unit size

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
G. Li, V. L. Knoop and H. V. Lint	2020	Dynamic Graph Filters Networks: A Gray-box Model for Multistep Traffic Forecasting	Transport	Prediction	Machine learning	Discrete	Months	Partial city	Component	Component	Traffic speed prediction
Z. Li, L. Li, Y. Peng and X. Tao	2020	A Two-Stream Graph Convolutional Neural Network for Dynamic Traffic Flow Forecasting	Transport	Prediction	Machine learning	Discrete	Months	Partial city, Regional	Component	Component	Traffic flow forecasting
W. Li, X. Wang, Y. Zhang and Q. Wu	2021	Traffic flow prediction over muti-sensor data correlation with graph convolution network	Transport	Prediction	Machine learning	Discrete	Months	Partial city	Component	Component	Traffic flow prediction with machine learning
H. Lin, R. Bai, W. Jia, X. Yang and Y. You	2020	Preserving Dynamic Attention for Long-Term Spatial-Temporal Prediction	Communications, Network methods, Transport	Prediction	Machine learning	Discrete	Months	Citywide, Regional	Actor, Component	Component	Machine learning traffic prediction
C. Liu, C. Gao and Y. Xin	2018	Measuring the diversity and dynamics of mobility patterns using smart card data	Transport	Relationships, Description	Discrete network methods, Statistical model	Discrete	Days	Citywide	Actor	Component, Whole network	Smart card trip data. Mobility patterns in week days, diversity for days. Temporal profile of stations, correelation with land use. Temporal clusters - network. Alluvial diagram, hour periods for temporal clusters.
S. G. Liu, L. Wang, C. Y. Liu and I. Destech Publicat	2018	Visualized Social Network Analysis on Spatial Dynamics of International Trade between China and League of Arab States	Economics	Description	Discrete network methods	Discrete	Years	Multi-country	Component	Component, Whole network	Trade between China and Arab States. Looking at density, centrality and internal subgroups. Aggregated into discrete periods.

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
X. L. Ma, Y. Li and P. Chen	2020	Identifying spatiotemporal traffic patterns in large-scale urban road networks using a modified nonnegative matrix factorization algorithm	Transport	Description	Discrete network methods	Discrete	Days	Partial city	Component	Whole network	Matrix factorisation produces one matrix with reduced dimensions containing temporal information about global traffic state, and another matrix describing spatial structure of traffic state. Clustering algorithm used to examine the matrices
J. Mateus-Anzola, A. Wiratsudakul, O. Rico-Chavez and R. Ojeda-Flores	2019	Simulation modeling of influenza transmission through backyard pig trade networks in a wildlife/livestock interface area	Epidemiology	Hypothetical	Stochastic simulation	Discrete	Months	Regional	Actor, Component	Component, Whole network	Pig influenza infections. Farms as nodes. Backyard pig trade. Different models looking at speed of spread. Visualised network at different time slices.
J. E. Mendoza-Cota and V. H. Torres-Preciado	2019	The impact of regional remittances on economic growth in Mexico: a dynamic space-time panel approach	Economics	Relationships	Statistical model	Discrete	Years	Single country	Component	Whole network	Regional remittances (money gifts) - impact on economic growth in Mexico. Spillover effects
T. M. Neeson, M. J. Wiley, S. A. Adlerstein and R. L. Riolo	2012	How river network structure and habitat availability shape the spatial dynamics of larval sea lampreys	Ecology	Hypothetical	Stochastic simulation	Discrete	Years	Unspecified	Actor, Whole network	Whole network	Large abstract river basin (49 reaches). Lamprey. How river system structure shapes sea lamprey larvae spatial dynamics - local habitat composition and river network shape. Models lamprey life cycle and movement through river system

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
P. A. Neto, T. L. Friesz and K. Han	2016	Electric Power Network Oligopoly as a Dynamic Stackelberg Game	Economics	Hypothetical	Stochastic simulation	Discrete	Hours	Multi-country	Component	Whole network	Electricity market based on western European market. Solution of dynamic stackelberg game (economic game). Consumers, market monitor, electricity generators. Network is electricity network with market for power at certain nodes. Nodes have behaviours.
J. Ou, J. Sun, Y. Zhu, H. Jin, Y. Liu, F. Zhang, J. Huang and X. Wang	2020	STP-TrellisNets: Spatial-Temporal Parallel TrellisNets for Metro Station Passenger Flow Prediction	Transport	Prediction	Machine learning	Discrete	Days	Citywide	Actor	Component	Predicting Metro station passenget flows. Origin destination flows
O. Parent and J. P. LeSage	2010	A spatial dynamic panel model with random effects applied to commuting times	Transport	Relationships	Statistical model	Discrete	Years	Single country	Component	Component	Spatial dynamic panel data model. Spatial spillover of increased highway capacity on travel times for neighbouring location. "Dynamic spatial lag panel model". Travel time follows state and local spending.
H. Peng, H. Wang, B. Du, M. Z. A. Bhuiyan, H. Ma, J. Liu, L. Wang, Z. Yang, L. Du, S. Wang and P. S. Yu	2020	Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting	Transport	Prediction	Machine learning	Discrete	Months	Citywide	Actor	Component	Traffic forecasting in network - subway, bus, taxi. Origin-destination data

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
L. Perez and S. Dragicevic	2009	An agent-based approach for modeling dynamics of contagious disease spread	Epidemiology	Hypothetical	Stochastic simulation	Discrete	Days	Partial city	Actor, Whole network	Actor	Propagation of communicable diseases using ABM and GIS. Measles epidemic in urban environment. Individuals as agents with transport network. Generic infection model with ABM. Network itself not temporal
G. L. W. Perry and F. Lee	2019	How does temporal variation in habitat connectivity influence metapopulation dynamics?	Ecology	Hypothetical	Stochastic simulation	Discrete	Abstract	Abstract	Component	Whole network	Network based metapopulastion model. Persistence of metapopulation.
Y. Qu, Y. Zhu, T. Zang, Y. Xu and J. Yu	2020	Modeling Local and Global Flow Aggregation for Traffic Flow Forecasting	Transport	Prediction	Machine learning	Discrete	Days	Citywide	Actor, Component	Component	Traffic flow forecasting with local and global components.
K. Ramachandran, B. Sikdar and Ieee	2007	On the stability of the malware free equilibrium in cell phones networks with spatial dynamics	Communications	Hypothetical	Stochastic simulation	Discrete	Unspecified	Unspecified	Component	Whole network	Malware propagation in mobile phone networks using 3 spread methods and movable nodes (phones). Infection model. Location paths/regions with variable infection rate.
R. Ryser, J. Haussler, M. Stark, U. Brose, B. C. Rall and C. Guill	2019	The biggest losers: habitat isolation deconsructs complex food webs from top to bottom	Ecology	Hypothetical	Stochastic simulation	Discrete	Abstract	Abstract	Actor, Component	Whole network	Food web network at each habitat patch connected in network. Food web is constant over all patches.

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
F. Semboloni	2000	The growth of an urban cluster into a dynamic self-modifying spatial pattern	Urban planning	Hypothetical	Stochastic simulation	Discrete	Abstract	Unspecified	Component	Whole network	Growth of urban cluster cells and road network. Looking at land use and urban morphology. Cell and road propagation occur seperately but incluence each other. Cell centroids randomly distributed, new nodes can change shape of cells and links between (shared edges). Shared edge=potential road. Road cannot be changed once built.
A. Srivastava and S. M. Salapaka	2020	Simultaneous Facility Location and Path Optimization in Static and Dynamic Networks	Network methods	Prediction, Description	Discrete network methods	Discrete	Hours	Abstract	Component	Component, Whole network	Path optimisation and planning location of long distance transport facilities (not necessarily haulage)
G. Sun, T. Tang, T. Q. Peng, R. Liang and Y. Wu	2017	SocialWave: Visual analysis of spatio-temporal diffusion of information on social media	Communications, Network methods, Sociology	Relationships, Description	Statistical model, Visualisation	Discrete	Unspecified	Unspecified	Component	Whole network	Visualise spatio-temporal diffusion of info on social media. Dynamic social gravity model, distance, language and cultural similarity. "Temporal visualisation" not totally clear. Tended to be cross sectional networks and single properties over time.
B. Sun, D. Zhao, X. Shi and Y. He	2021	Modeling Global Spatial-Temporal Graph Attention Network for Traffic Prediction	Transport	Prediction	Machine learning	Discrete	Months	Partial city, Regional	Component	Component	Traffic prediction.
K. Tian, J. Guo, K. Ye and C. Z. Xu	2020	ST-MGAT: Spatial-Temporal Multi-Head Graph Attention Networks for Traffic Forecasting	Transport	Prediction	Machine learning	Discrete	Months	Partial city	Component	Component	Traffic forecasting

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
X. B. Wang, J. P. Ge, W. D. Wei, H. S. Li, C. Wu and G. Zhu	2016	Spatial Dynamics of the Communities and the Role of Major Countries in the International Rare Earths Trade: A Complex Network Analysis	Economics	Description	Discrete network methods	Discrete	Years	Multi-country	Component	Whole network	Rare earth trade. Communities - dynamics of these. Uses unweighted and weighted networks. Annual data used for community detection.
I. Wanzenböck and P. Piribauer	2018	R&D networks and regional knowledge production in Europe: Evidence from a space-time model	Geography	Relationships	Discrete network methods, Statistical model	Discrete	Years	Multi-country	Component	Component	Regional knowledge production follows embeddedness (like centrality) and local characteristics in Dynamic spatial Durbin model. Annual data. First with no spatio-temporal element then with. Grant connections ->network ->regional net properties ->space-time model. Spatially aggregated into regions too.
V. J. Wei, R. C. W. Wong and C. Long	2020	Architecture-Intact Oracle for Fastest Path and Time Queries on Dynamic Spatial Networks	Network methods	Description	Discrete network methods	Discrete	Months	Regional	Component	Component, Whole network	Identifying fastest path in network with changing weights over time.
M. Xu and Z. Y. Gao	2009	Chaos in a dynamic model of urban transportation network flow based on user equilibrium states	Transport	Hypothetical	Stochastic simulation	Discrete	Abstract	Unspecified	Component	Whole network	Model to analyse network flow and estimate number of trips based on dynamical gravity mode. Estimated trip cost using "user equilibrium network assignment model". Variations in O-D flow. Does chaos exist?
B. Yang, Y. Kang, H. Li, Y. Zhang, Y. Yang and L. Zhang	2020	Spatio-temporal expand-and-squeeze networks for crowd flow prediction in metropolis	Transport	Prediction	Machine learning	Discrete	Months	Citywide	Actor	Component	Traffic flow prediction (pedestrian)

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
H. X. Yao, X. F. Tang, H. Wei, G. J. Zheng, Z. H. Li and Aaai	2019	Revisiting Spatial-Temporal Similarity: A Deep Learning Framework for Traffic Prediction	Transport	Prediction	Machine learning	Discrete	Months	Citywide	Component	Component	Traffic prediction with dynamic spatial dependence. Temporal not strictly periodic. Novel space-time dynamic network. Taxi/bike trips. Aggregated spatially to 1km and 30 mins.
H. C. Yu, Z. X. Fang, F. Lu, A. T. Murray, Z. Y. Zhao, Y. Xu and X. P. Yang	2019	Massive Automatic Identification System Sensor Trajectory Data-Based Multi-Layer Linkage Network Dynamics of Maritime Transport along 21st-Century Maritime Silk Road	Transport	Description	Discrete network methods	Discrete	Years	Multi-country	Actor, Component	Component, Whole network	Described dynamics of global trade in silk
W. H. Yu, M. L. Guan and Z. L. Chen	2019	Analyzing Spatial Community Pattern of Network Traffic Flow and Its Variations across Time Based on Taxi GPS Trajectories	Transport	Description	Discrete network methods	Discrete	Hours	Partial city	Actor, Component	Whole network	Communities based on Taxi GPS data at different time points/windows
M. A. Zambrano-Monserrate, M. A. Ruano, V. Ormeño-Candelario and D. A. Sanchez-Loor	2020	Global ecological footprint and spatial dependence between countries	Ecology	Relationships	Statistical model	Discrete	Years	Multi-country	Component	Whole network	Spillover of ecological footprint between countries. Dynamic spatial Durbin Model
C. Zhang, J. J. Q. Yu and Y. Liu	2019	Spatial-Temporal Graph Attention Networks: A Deep Learning Approach for Traffic Forecasting	Transport	Prediction	Machine learning	Discrete	Months	Regional	Component	Component	Traffic flow forecasting

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
W. Zhang, Z. Tian, G. Zhang and G. Dong	2019	Spatial-temporal characteristics of green travel behavior based on vector perspective	Transport	Description	Discrete network methods	Discrete	Days	Citywide	Actor	Component, Whole network	Tiny discrete intervals. Vector coarse graining of network - looking at detailed graph at a coarser/higher level. Extract dominant directional component for inflow and outflow of node (if a node has one weighted edge with w>= half weighted degree than this is the DDC). Coarse graining keeps all DDCs and weights them at 1 and calculated directional degree - tells us on average how unidirectional flow is for each node on average. Also include way of defining temporal window size to use.
X. Z. Zhang, G. Chen, J. C. Wang, M. C. Li and L. Cheng	2019	A GIS-Based Spatial-Temporal Autoregressive Model for Forecasting Marine Traffic Volume of a Shipping Network	Transport	Prediction	Statistical model	Discrete	Years	Multi-country	Actor, Component	Component	Traffic forecasting
Z. C. Zhang, M. Li, X. Lin, Y. H. Wang and F. He	2019	Multistep speed prediction on traffic networks: A deep learning approach considering spatio-temporal dependencies	Transport	Prediction	Machine learning	Discrete	Abstract	Partial city	Actor, Component	Component	This one has a 3D graph of speed over time on different road links. Unspecified time period of data.
H. Zhang, L. Zhang, F. Che, J. Jia and B. Shi	2020	Revealing Urban Traffic Demand by Constructing Dynamic Networks with Taxi Trajectory Data	Transport	Description	Discrete network methods	Discrete	Hours	Citywide	Actor, Component	Whole network	Networks based on flows between 2km square areas constructed for 2 hour time windows.

Author	Year	Title	Subject Area	Aim	Method	Time specification	Time period	Spatial Level	Input Network level	Output Network Level	Notes
X. Y. Zhao and Y. X. Pu	2018	The Complex System of Interprovincial Migration Flows in China, 1985-2015: A Spatial Dynamic Panel Modeling Approach	Geography	Relationships	Statistical model	Discrete	Years	Single country	Component	Whole network	Migration in China - impacts of regional characteristics. Spatial dynamic panel model for OD flows, based on gravity model O-D characteristics, distance decay on current flows and time dependence on self and neighbouring regions. Annual data. Origin anddestination effects.
N. Zhao, Z. Ye, Y. Pei, Y. C. Liang and D. Niyato	2020	Spatial-Temporal Attention-Convolution Network for Citywide Cellular Traffic Prediction	Communications	Prediction	Machine learning	Discrete	Months	Citywide	Actor	Component, Whole network	Telephone traffic prediction
L. Zheng, J. Yang, L. Chen, D. Sun and W. Liu	2020	Dynamic spatial-temporal feature optimization with ERI big data for Short-term traffic flow prediction	Transport	Prediction	Machine learning	Discrete	Days	Citywide	Actor, Component	Component	Traffic flow prediction
F. Zhou, Q. Yang, K. Zhang, G. Trajcevski, T. Zhong and A. Khokhar	2020	Reinforced Spatiotemporal Attentive Graph Neural Networks for Traffic Forecasting	Transport	Prediction	Machine learning	Discrete	Months	Regional	Component	Component	Machine learning traffic prediction

Table A.1: Table summarising reviewed papers in Chapter 2

Appendix B

Material associated with the development of the longitudinal simulation tool

Listing B.1: R code to generate three-phase growth of several units of observation

```
#Some "hidden markov chains" to simulate three growth stages of fifteen imaginary things over 30 time units.
#The HMC can move from "INITIAL SLOW GROWTH"->"GROWTH SPURT"->"STEADY STATE" (always starting at "ISG").
#The probability of these transitions increases with time (and those of looping back decrease).
#The current state determines the distribution from which "growth during time period" is sampled.
#This is added to size at the previous time period.
#NB/ In STEADY STATE", "growth during time period" can be negative.

library(ggplot2)
maxt<-30 #set timeframe
obs<-15 #set number of observation units
alldata<-data.frame(cbind(numeric(), numeric(), numeric(), numeric())) #empty data frame for results


for (i in 1:obs){
  #set up vectors to record individual ob unit results
  v<-numeric(length=maxt)
  inc<-numeric(length=maxt)
  cumul<-numeric(length=maxt+1)
  v[1]<-1
  inc[1]<-rlnorm(1,0,1)
  cumul[1]<-0
  for (t in 2:maxt){
    #set cumulative value from last cumulative value and last growth amount
    cumul[t]<-inc[t-1]+cumul[t-1]
    #set transition probabilities that depend on time using logistic curve

    p1<-1/(1+exp(-0.5*(t-13)))


    p2<-1/(1+exp(-1*(t-18)))
    #set new state based on transition probability and current state
    if (v[t-1]==1){
      v[t]<-rbinom(1, 1, p1)+1
    }
    if (v[t-1]==2){
```

```
      v[t]<-rbinom(1,1,p2)+2
    }
    if (v[t-1]==3){
      v[t]<-3
    }
    #sample growth amount from distro dependent on current state
    if (v[t]==1){
      inc[t]<-rlnorm(1,0.5,1)
    }
    if (v[t]==2){
        inc[t]<-rlnorm(1,2.5,1)
     }
    if (v[t]==3){
        inc[t]<-rnorm(1,0,5)
    }
  }
  #set final growth value
  cumul[maxt+1]<-inc[maxt]+cumul[maxt]
  if (cumul[maxt+1]<0) {
    cumul[maxt+1]<-0 #can't be smaller than zero
  }
  addon<-cbind(c(0:maxt), c(rep(i, maxt+1)), cumul, c(1,v))
  #add on to data for other ob units
  alldata<-data.frame(rbind(alldata, addon))

}

#plot
colnames(alldata)<-c("time", "id", "size", "stage")
#graph shows individual growth measurements over time with colouring according to growth "stage" simulated
ggplot(data=alldata, aes(x=time, y=size, group=id, colour=as.factor(stage)))+
  geom_line()+
  geom_point()+
  scale_colour_discrete(name="Growth Stage", labels=c("Initial slow growth", "Growth spurt", "Steady state"))
```

Listing B.2: R code to generate temporal observations for units of observation with a set average value

```
    #generates patterns with desired average over time and no big changes

maxt<-30 #set timeframe
obs<-10 #set number of observation units
alldata<-data.frame(cbind(numeric(), numeric(), numeric(), numeric(), numeric(), numeric())) #empty data frame for results
library(ggplot2)
for (i in 1:obs) {
  change<-numeric(length=maxt) #set empty vectors
    value<-numeric(length=maxt+1)
  average<-rnorm(1,20,7) #set desired average
  value[1]<-rnorm(1,average,3) #set first value on either side of desired average

  for (t in 1:maxt){

    if(value[t] > average) {
      change[t]<-(-1)*rlnorm(1,log(1+value[t]-average),log(2)) #if above desired average, pick negative value of change
      }

    if(value[t] < average) {
      change[t]<-rlnorm(1,log(1+average-value[t]),log(2)) #if below desired average pick positive value of change
      }
    #if value==desired average, simulate change value
    if(value[t]==average) {
      change[t]<-rnorm(1,0,3) #if = desired average, pick value of change in either direction
      }
    value[t+1]<-value[t]+change[t] #add change to previous value to get new value
  }
  addon<-cbind(c(0:maxt), c(rep(i, maxt+1)), value, c(NA, change), c(average, rep(NA, maxt)), c(mean(value), rep(NA, maxt))
    )
  alldata<-data.frame(rbind(alldata, addon)) #collect data and combine with other ob units
}
#plot and display
colnames(alldata)<-c("time", "id", "value", "change", "average.input", "average.output")
ggplot(data=alldata, aes(x=time, group=id, colour=as.factor(id)))+
  geom_line(aes(y=value))+
  geom_point(aes(y=value))+
  theme(legend.position="none")
```

B.1 Simulation function details

B.1.1 Required libraries

The function simulate.trajectory requires the libraries `ggplot2`, `gridExtra`, `Matrix`, `matrixcalc`, `MASS` and dependencies.

B.1.2 Arguments

Argument	Form	Default	Description
`n.obs`	numeric, length=1	`1`	Number of trajectories to generate (number of observation units
`max.time`	numeric, length=1	`10`	Maximum time of window to simulate data in (minimum is always zero)

Argument	Form	Default	Description
`record.times`	numeric, length≥1	`1:n.measures`	Two options. For wave measurements, specify the times at which each wave is centred. Recording times will vary around the specified times. For irregular measurements, specify the number of records collected and a uniform distribution will be used to choose this many recording times for each trajectory.
`record.times.vary`	numeric, length=1	`1`	Only required for wave measurements. Standard deviation for the variation of actual measurement times around those specified in `record.times`.

Argument	Form	Default	Description
`part.1.average`	character, 3≤length≤4	no default	Population distribution of the average value of the first trajectory section. Individual averages will be sampled from this distribution. Takes the form `distribution, parameter1, parameter2`. Two of `part.1.average`, `part.2.average` and `value.step.change` must be specified. If all three are specified, `part.2.average` is ignored.
`part.2.average`	character, 3≤length≤4	no default	Population distribution of the average value of the second trajectory section. Individual averages will be sampled from this distribution. Takes the form `distribution, parameter1, parameter2`. Two of `part.1.average`, `part.2.average` and `value.step.change` must be specified. If all three are specified, `part.2.average` is ignored.

Argument	Form	Default	Description
`part.1.slope`	character, length=3	`c("rnorm", "0", "1", "1")`	Population distribution of the slope parameter for the first trajectory section. Individual averages will be sampled from this distribution. Takes the form distribution, `parameter1, parameter2, multiplier`. The multiplier defaults to 1 and can be specified in terms of `t` (for time) if a non-linear pattern is required, for example, specifying `t` will give a quadratic curve and `t^2` cubic.
`part.2.slope`	character, length=3	`c("rnorm", "0", "1", "1")`	Population distribution of the slope parameter for the second trajectory section. Individual averages will be sampled from this distribution. Takes the form distribution, `parameter1, parameter2, multiplier`. The multiplier defaults to 1 and can be specified in terms of `t` (for time) if a non-linear pattern is required, for example, specifying `t` will give a quadratic curve and `t^2` cubic.

Argument	Form	Default	Description
`time.step.change`	character, length=3	`c("rnorm", "0", "1", "1")`	Population distribution of the time at which the step change between trajectory sections occurs. Individual averages will be sampled from this distribution. Takes the form `distribution, parameter1, parameter2`.
`value.step.change`	character, length=3	no default	Population distribution of the average value of the step change between trajectory sections. Individual averages will be sampled from this distribution. Takes the form `distribution, parameter1, parameter2`. Two of `part.1.average`, `part.2.average` and `value.step.change` must be specified. If all three are specified, `part.2.average` is ignored.

Argument	Form	Default	Description
`sd.error`	character, length=3	`sd.error=c("rlnorm","0.5","0.5")`	Population distribution of the standard deviation of total random error terms for each individual. Individual averages will be sampled from this distribution. Takes the form `distribution, parameter1, parameter2`. The absolute value of any negative values will be used.
`between.cor`	numeric, length=1	`0`	Correlation of error terms between individuals at the same time point. Must be between -1 and 1.
`within.cor`	numeric, length=1	`0`	Correlation of error terms within individuals at adjacent time points. An order 1 autocorrelation structure is specified within individuals. Must be between -1 and 1.
`dayreps`	numeric, length=1	`1`	Routes can be measured over the time specified on multiple occasions (for example, different days). This specifies the number of these occasions.

Argument	Form	Default	Description
`dayreps.cor`	numeric, length=1	`0`	Error correlation between data measured at the same time on the same route on consecutive days.
`fixed.spacegroups`	Matrix or data frame with 3 or 6 columns	`NULL`	Specify this if you wish the observations to be in pre-specified spatial groups. Matrix/data frame columns specify: Group 1 ID, Group 1 X coordinate, Group 1 Y coordinate, Group 2 ID, Group 2 X coordinate, Group 2 Y coordinate. If two groups (6 columns) are specified, origin-destination data will be produced, if 3 columns are specified it will not (this will override the specification of `origin.dest`). The length of the matrix/data frame must be equal to n.obs.
`fixed.effect.part.1.slope`	Numeric, length=n.obs	`NULL`	A vector of values specifying a fixed amount by which `part.1.slope` will be multiplied for each observation unit.
`spacegroups`	numeric, length=1	no default	Number of spatial groups or clusters of individuals to be included in the dataset.

Argument	Form	Default	Description
`spacegroups.size`	numeric, length=1	no default	The average deviation of the x and y coordinates for an individual in a spatial group from its centre (the distribution of points follows a bivariate normal distribution, this is the standard deviation of both distributions that comprise this).
`origin.dest`	logical, length=1	`FALSE`	If `TRUE`, data will be associated with a 'route' between two spatial points.
`area.x`	numeric, length=1	no default	Upper bound for x-coordinates of the centre of each spatial group. These will be sampled from a uniform distribution spanning from zero to this value.
`area.y`	numeric, length=1	no default	Upper bound for y-coordinates of the centre of each spatial group. These will be sampled from a uniform distribution spanning from zero to this value.

Argument	Form	Default	Description
`sd.ratio`	numeric, length=1	`0.5`	Used to specify the amount of random variation in parameters within spatial groups. The standard deviation for this spatial-group-related variation is specified as a fraction of the random variation already introduced to the parameters (the third argument for each of `part.1.average`, `part.1.slope` etc.) If spatial groups are specified, the random variation introduced in these arguments is between spatial group variation.

Argument	Form	Default	Description
`sd.ratio.day`	numeric, length=1	`0.5`	Used to specify the amount of random variation in parameters between each dayrep specified. The standard deviation for this date-related variation is specified as a fraction of the random variation already introduced to the parameters (the third argument for each of `part.1.average`, `part.1.slope` etc.) If `sd.ratio.day` is set to zero there will be no random variation in parameters or horizontal offset within spatial groups for different days.
`print.plot`	logical, length=1	`TRUE`	Whether the summary plot is printed.
`offset.amount`	numeric, length=1	`0`	For specifying a horizontal offset for each spatial group. Each group will be offset by a value sampled from a normal distribution with mean 0 and standard deviation `offset.amount` with some random within-group variation around this value.

Table B.1: Function arguments for simulation tool

B.1.3 Output

The function outputs a list containing a maximum of seven elements.

Element	Description
`data`	Data frame containing simulated data in long format
`simulation.summary`	Data frame containing variables showing individual input parameters, values expected from a simulation with no error (given discrete measurement times), and actual parameters from simulated data.
`plot`	Graphical object containing spaghetti plot of trajectories and differences of average values and step change values from inputs and expected values. Can be displayed using plot() function.
`section1.function`	Individual functions specified for first trajectory section (including multiplier)
`section2.function`	Individual functions specified for second trajectory section (including multiplier)
`coordinates.info`	Contains two parts. The first contains central coordinates for each spatial group. The second contains individual level coordinates and spatial group membership.
`group.info`	Data frame containing information about the average parameter values in spatial and day groups.
`distmat`	Matrix or list of two matrices detailing the distances between observation units or observation unit origins and destinations.

Table B.2: Elements of list output of simulation tool function

B.1.4 Warning messages

Common warning messages include the following:

```
In max(recordtime[which(recordtime[, i] <= prejumpt[i]), i]) : no non-missing arguments to max; returning -Inf
In min(recordtime[which(recordtime[, i] >= postjumpt[i]), i]) : no non-missing arguments to max; returning Inf
Removed 1 rows containing non-finite values (stat_density).
```

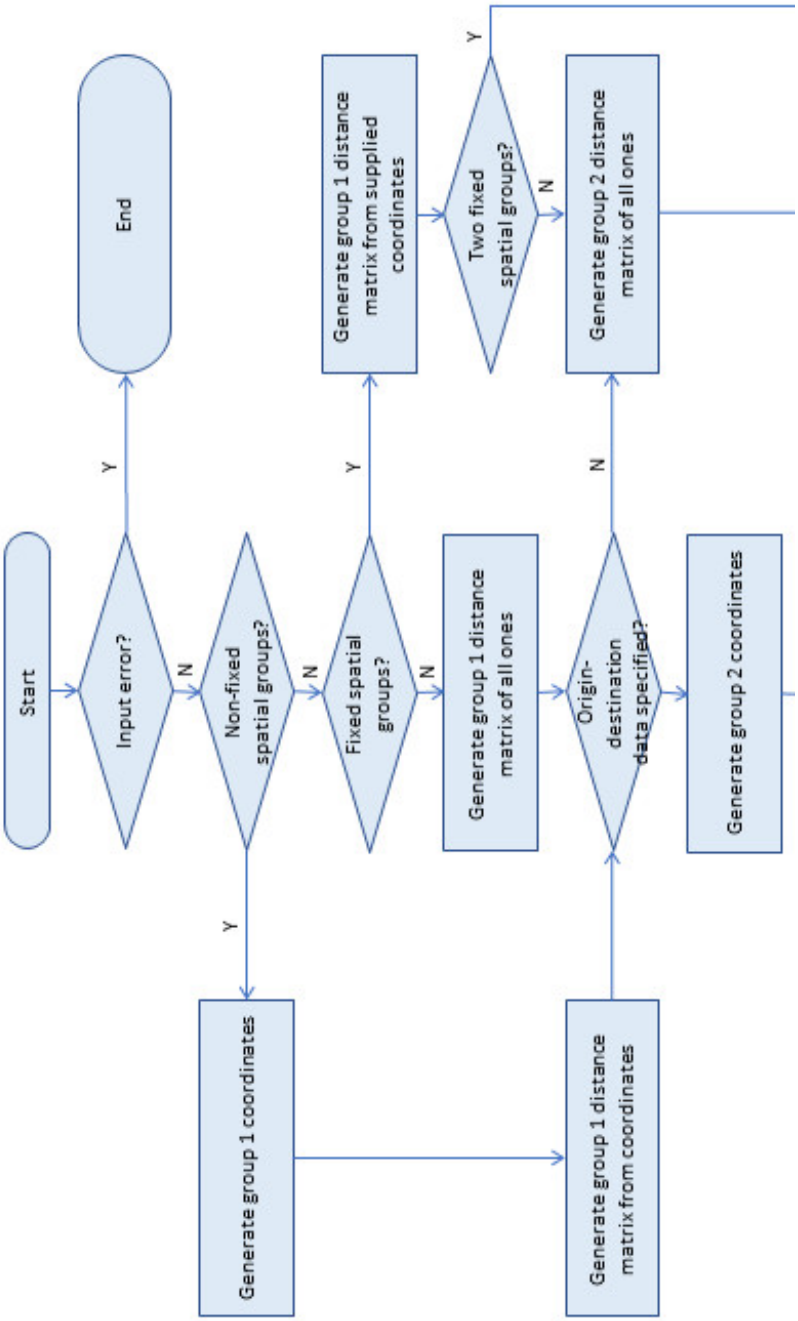
These occur if wave or irregular sampling is specified and an individual has no recording times either before or after the step change. This means some summary values cannot be calculated but does not cause any other issues.

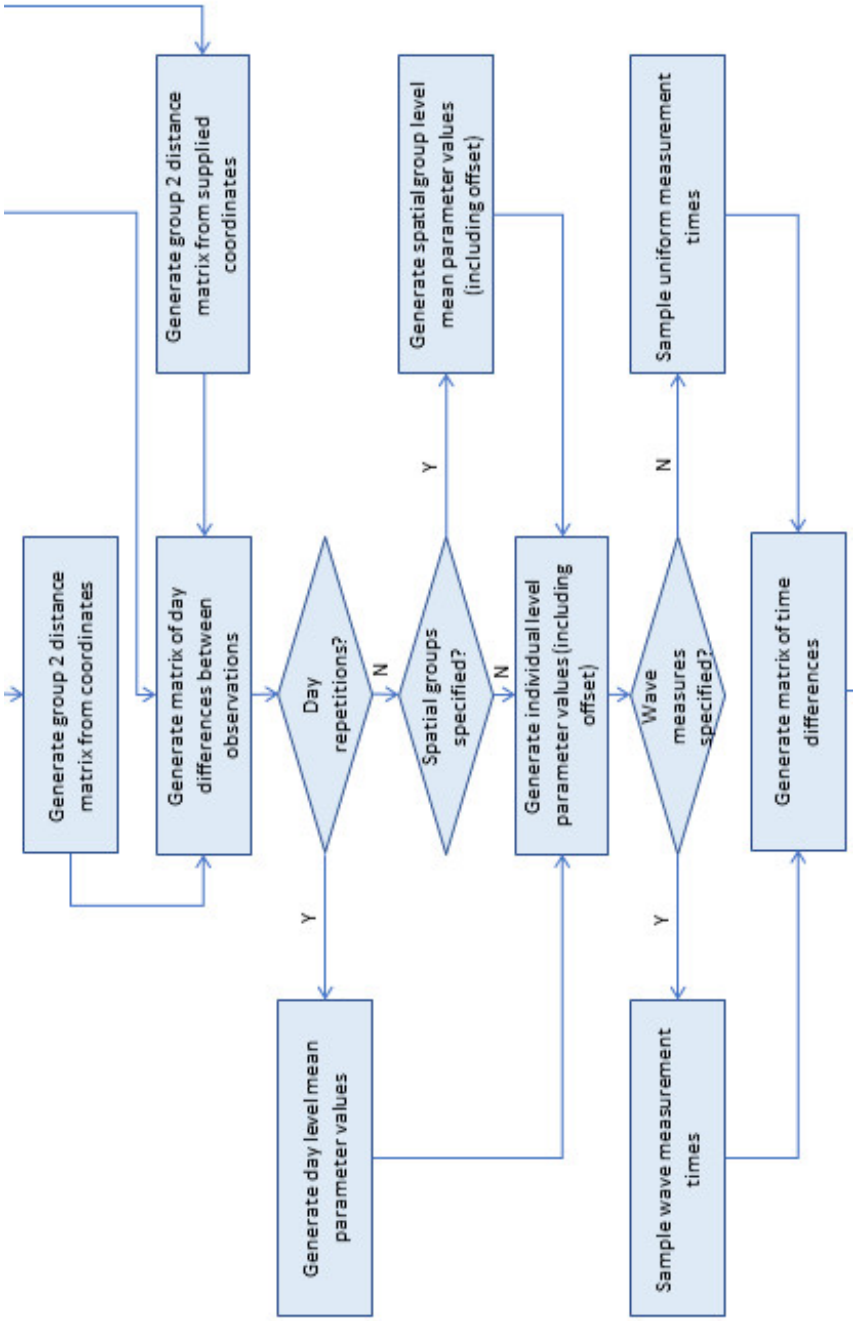
Listing B.3: Code used to simulate example data

```
B <- simulate_trajectory(n.obs=50 , max.time=30, record.times=10, part.1.average=c("rnorm","100", "3"), part.1.slope=c("
    rlnorm", "1", "0.1", "log(t)/t") , part.2.slope=c("rnorm", "1", "0.5", "cos(t)") , time.step.change=c("runif","10", "20"
    ) , value.step.change=c("rnorm","-100", "3"), sd.error=c("rlnorm","1", "0.5"), between.cor=0.3, within.cor=0.5, dayreps
    =1, spacegroups=2, spacegroup.size=1, area.x=30, area.y=30, sd.ratio=0.5, offset.amount=0)
```

Appendix C

Large scale flowchart showing the process underlying the simulation function in Chapter 5


Start
Input error?
Y
End
N
Non-fixed spatial groups?
Y
Generate group 1 coordinates
N
Fixed spatial groups?
Y
Generate group 1 distance matrix from supplied coordinates
N
Two fixed spatial groups?
Y
Generate group 2 distance matrix of all ones
N
Generate group 1 distance matrix of all ones
Origin-destination data specified?
N
Generate group 1 distance matrix from coordinates
Y
Generate group 2 coordinates


Generate group 2 distance matrix from coordinates
Generate group 2 distance matrix from supplied coordinates
Generate matrix of day differences between observations
Day repetitions?
Y
N
Generate day level mean parameter values
Spatial groups specified?
Y
N
Generate spatial group level mean parameter values (including offset)
Generate individual level parameter values (including offset)
Wave measures specified?
N
Y
Sample uniform measurement times
Sample wave measurement times
Generate matrix of time differences

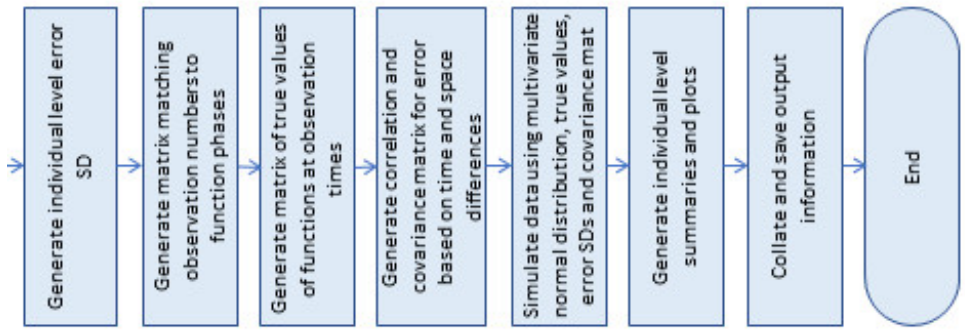


Figure C.1: Large size flowchart in from Figure 5.7

Appendix D

London Underground Maps

September-December 2009

Figure D.1: Front cover of London Underground Map September 2009

This image has been removed by the author of this thesis for copyright reasons

Figure D.2: Inside of London Underground Map September 2009 - note that fare zones are missing

This image has been removed by the author of this thesis for copyright reasons

Figure D.3: Front cover of London Underground Map December 2009

This image has been removed by the author of this thesis for copyright reasons

Figure D.4: Inside of London Underground Map December 2009 - this was published soon after the September 2009 map with the River Thames (which had been removed) and fare zones reinstated.

Appendix E

Supplementary figures and tables for Chapter 7

Initial tau value	**0.0025**	**0.01**	**0.04**	**1**	**4**	**100**
Time of maximum delay (hours)						
Mean Estimate	15:35	15:35	15:35	15:35	15:35	15:35
SD Estimates	0.692	0.69	0.694	0.696	0.694	0.695
Mean CI width	2.142	2.1	2.148	2.134	2.117	2.104
SD CI width	1.813	1.728	1.787	1.767	1.733	1.731
Maximum delay (minutes)						
Mean Estimate	18.654	18.654	18.682	18.682	18.68	18.67
SD Estimates	9.909	9.931	9.916	9.911	9.9	9.914
Mean CI width	6.307	6.313	6.295	6.289	6.298	6.297
SD CI width	0.461	0.464	0.455	0.447	0.487	0.492

Table E.1: Table summarising estimates from simulated data of time of maximum delay and maximum delay for different initial values of tau.

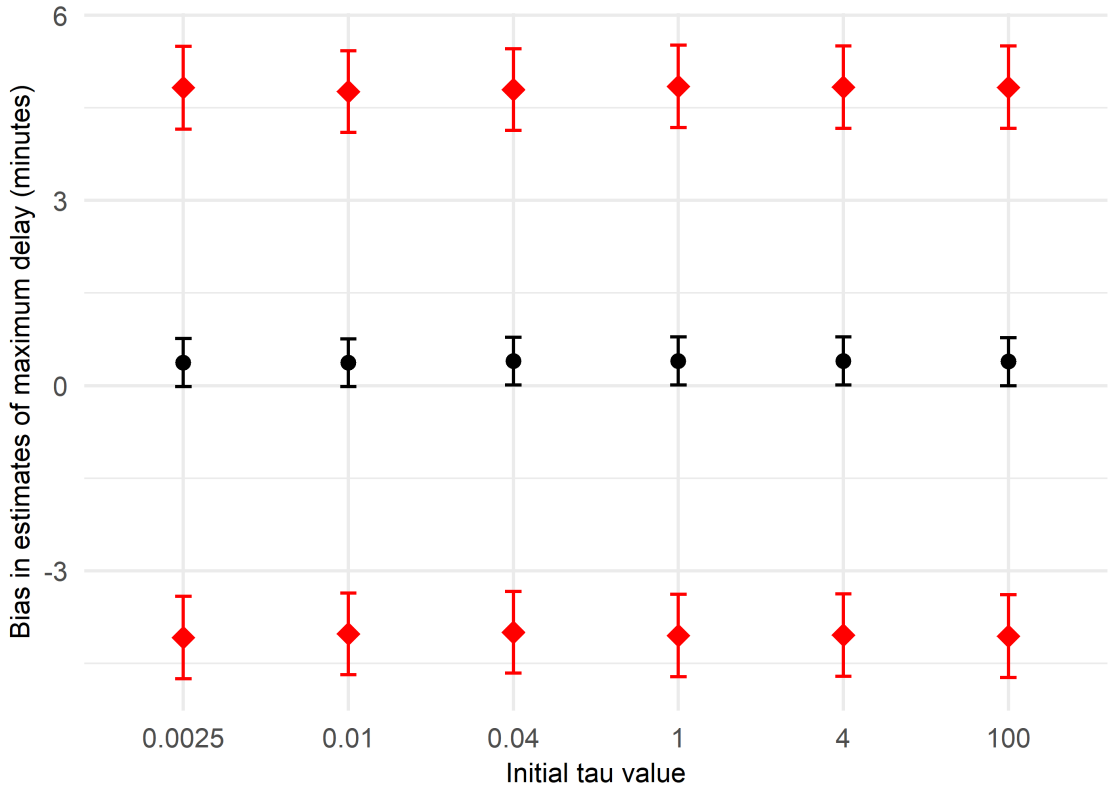


Figure E.1: Mean bias (black circle) and limits of agreement (red diamond) in estimates of maximum delay from simulated data according to initial value of tau, with 95% CI.

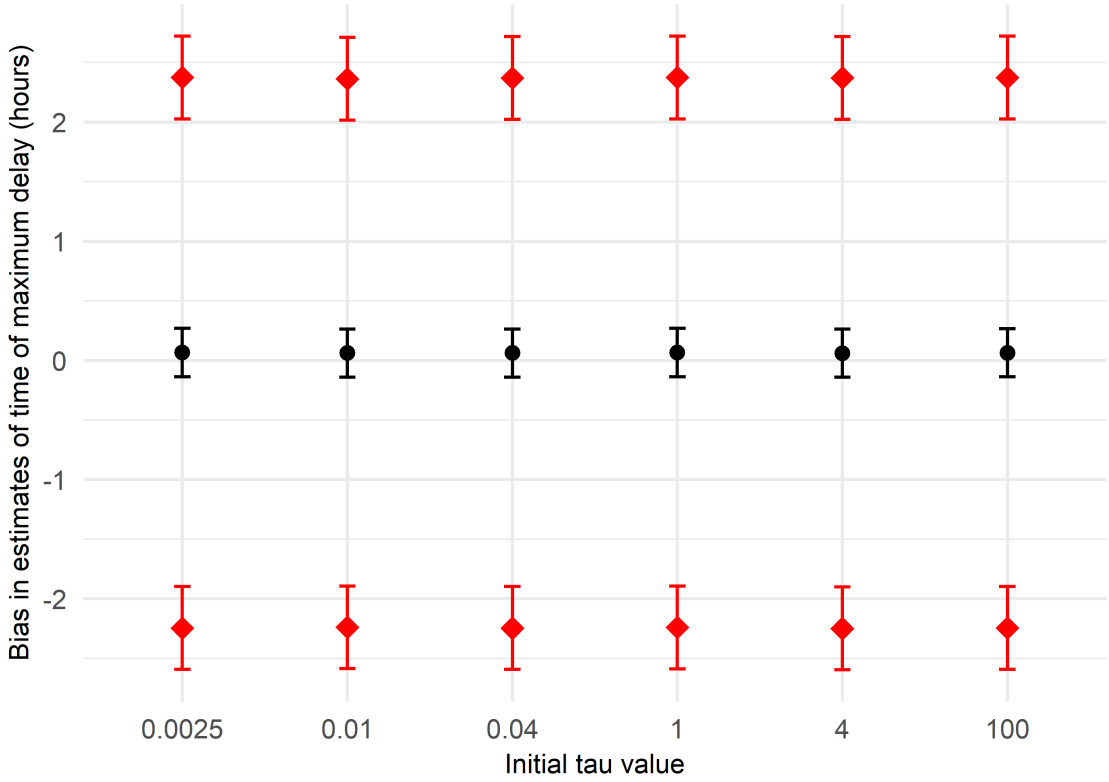


Figure E.2: Mean bias (black circle) and limits of agreement (red diamond) in estimates of time of maximum delay from simulated data according to initial value of tau, with 95% CI.

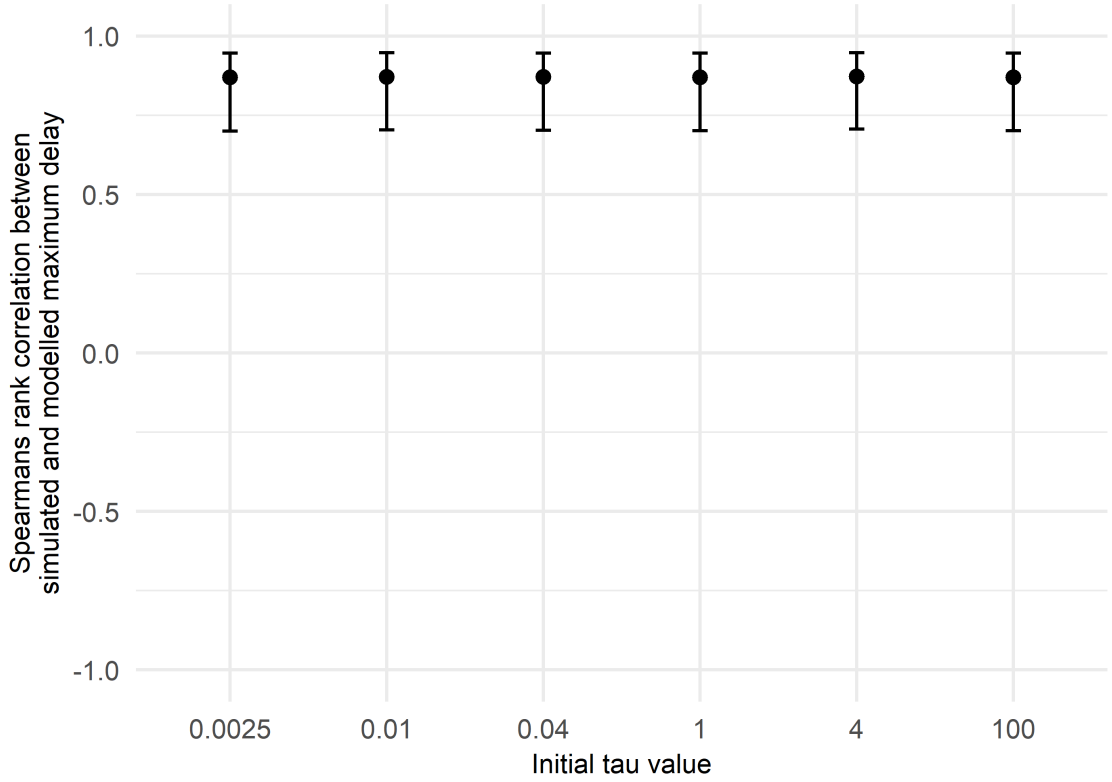


Figure E.3: Spearman's rank correlation coefficient and 95% confidence interval for relationship between simulated and modelled values of maximum delay from simulated data according to initial value of tau.

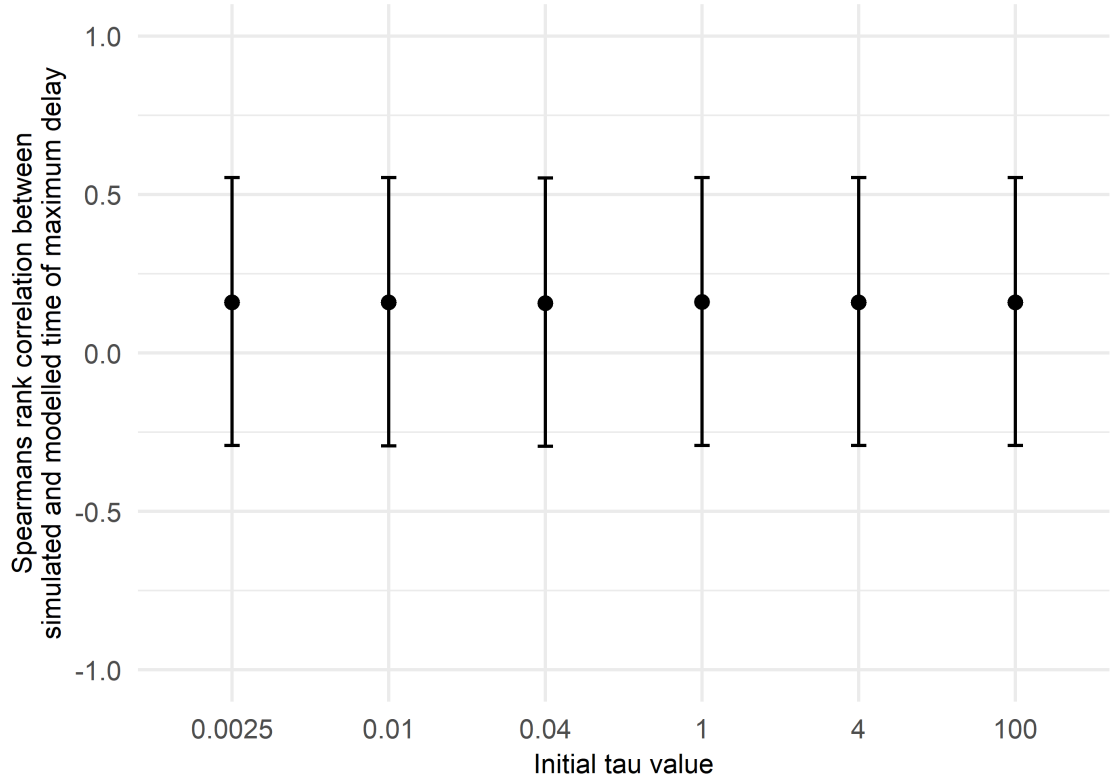


Figure E.4: Spearman's rank correlation coefficient and 95% confidence interval for relationship between simulated and modelled times of maximum delay from simulated data according to initial value of tau.

Appendix F

Supplementary figure for Chapter 8

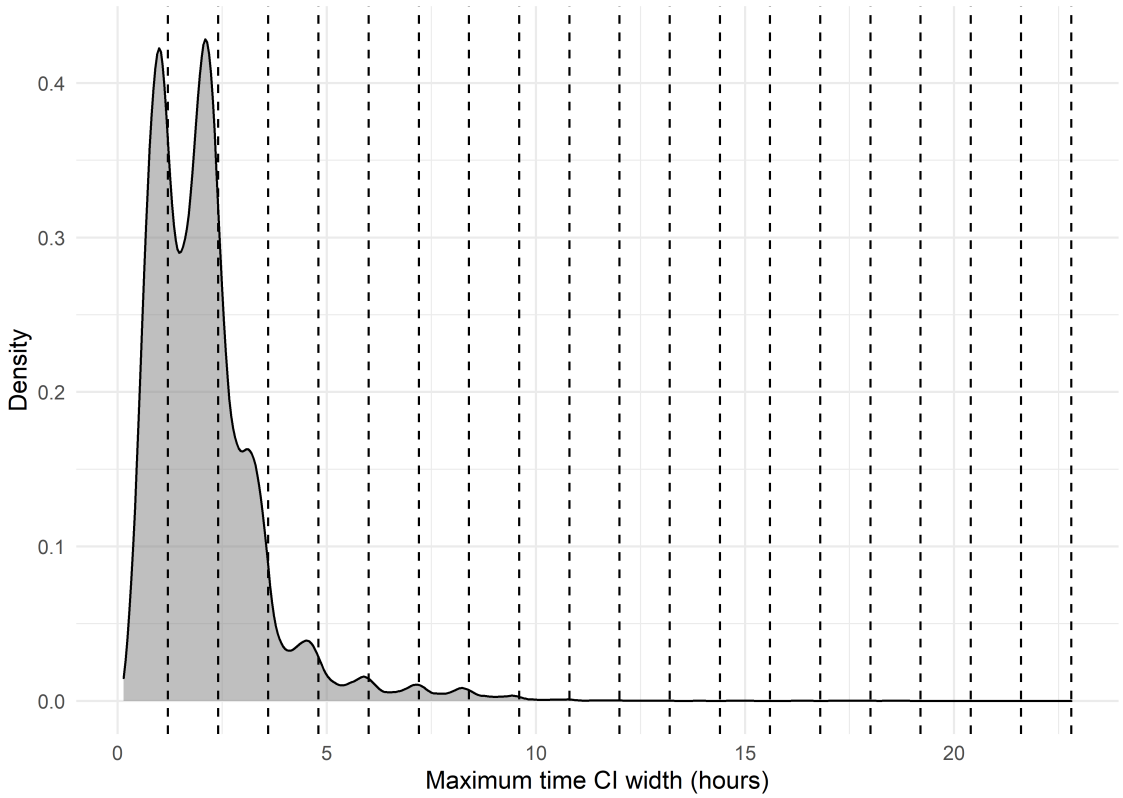


Figure F.1: Density plot showing 95% credible interval widths for estimates of maximum time, with the distances between knot points marked as vertical dashed lines.

Appendix G

Plots of posterior distributions of pattern features estimated for one simulated dataset in Chapter 8

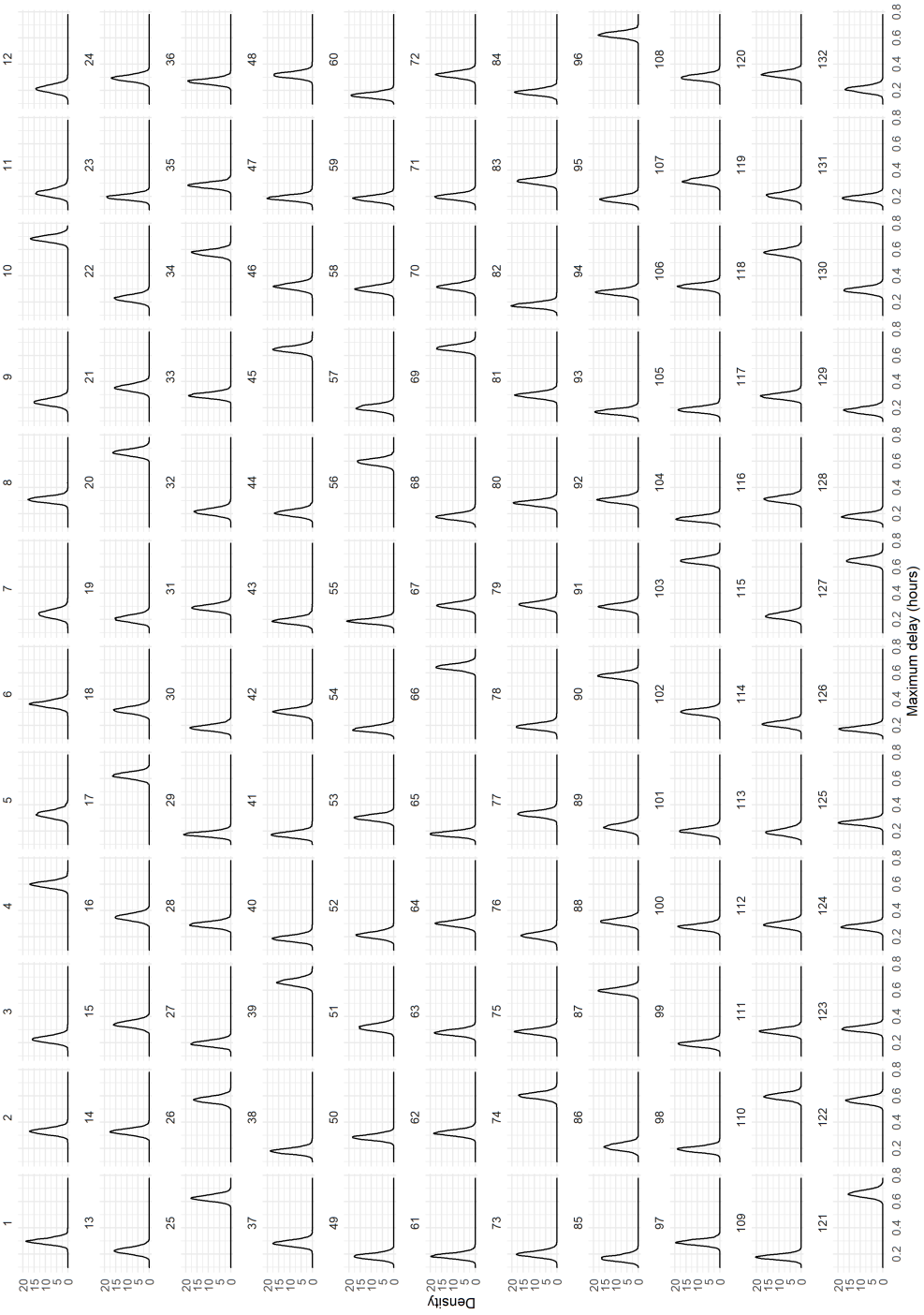


Figure G.1: Posterior distributions of maximum delay estimated for each edge in one simulated dataset

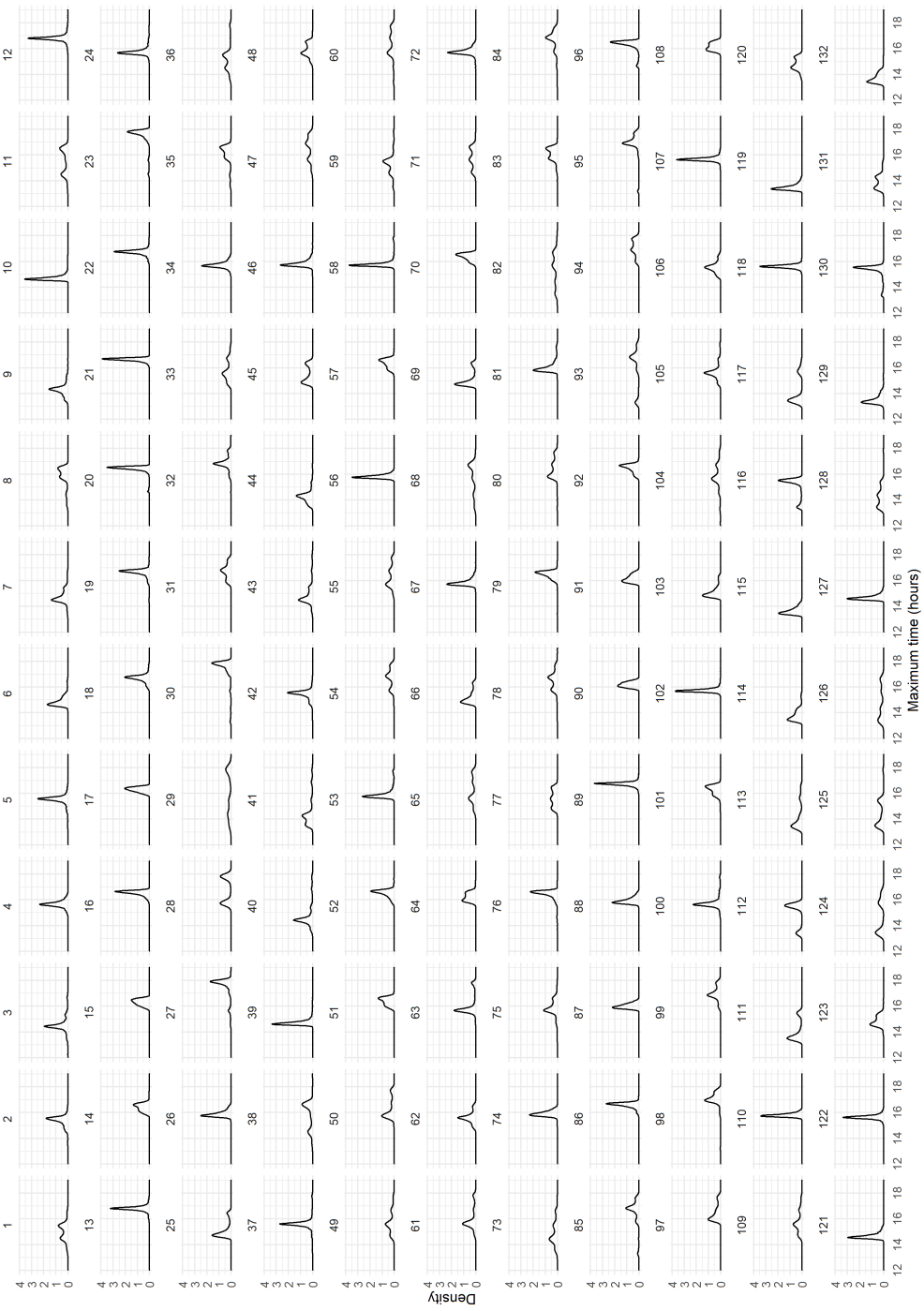


Figure G.2: Posterior distributions of time of maximum delay estimated for each edge in one simulated dataset

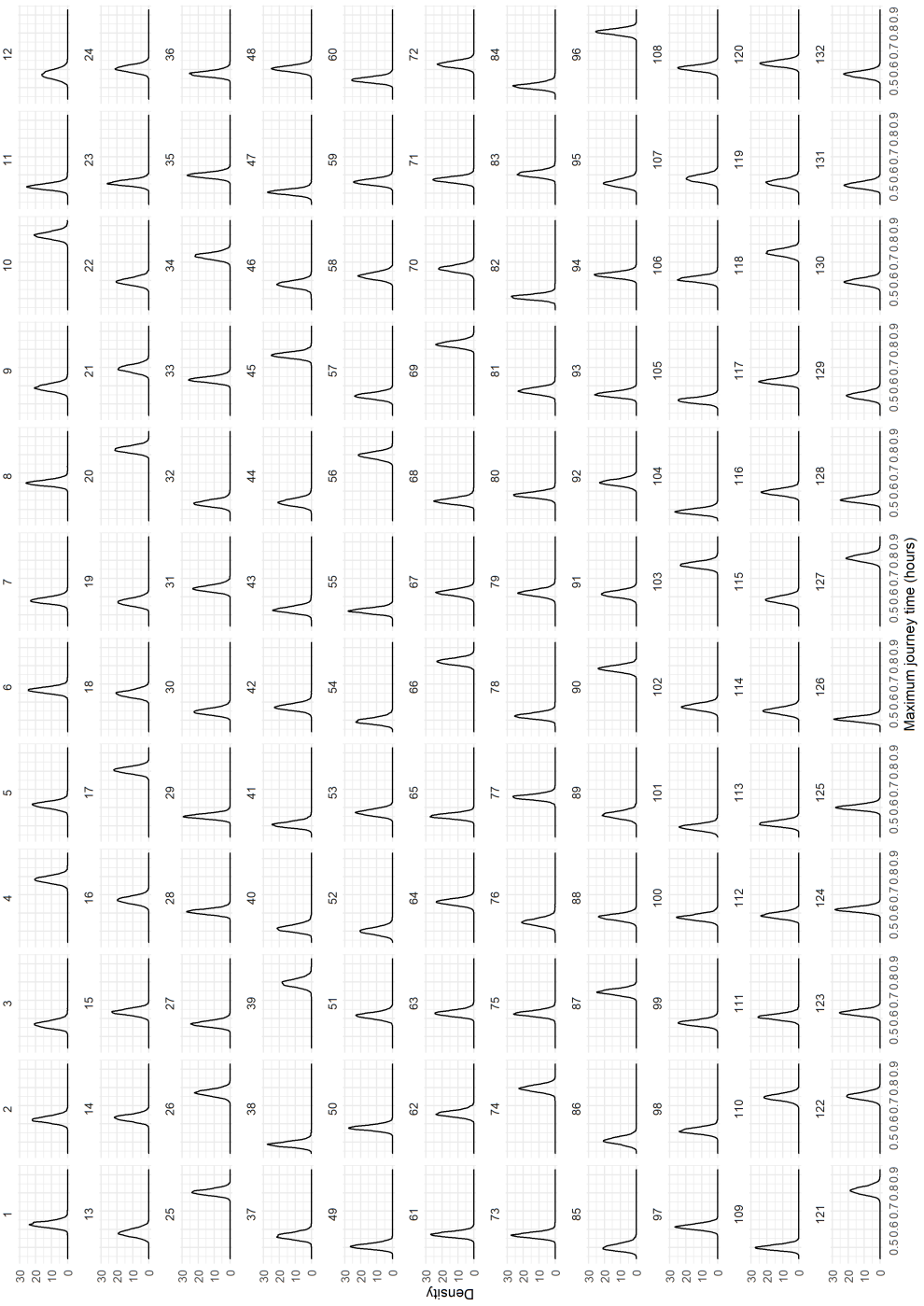


Figure G.3: Posterior distributions of maximum journey length estimated for each edge in one simulated dataset