



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

Sarah C. Gadd^{1,2} · Alexis Comber¹ · Mark S. Gilthorpe^{2,3,4} · Keiran Suchak^{1,2} · Alison J. Heppenstall^{1,2,3}

Received: 29 January 2020 / Accepted: 21 December 2020 / Published online: 26 July 2021
© The Author(s) 2021

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 including 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

JEL Classification C88

✉ Alison J. Heppenstall
a.j.heppenstall@leeds.ac.uk

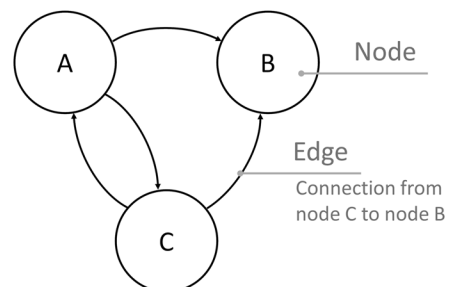
Extended author information available on the last page of the article

1 Introduction

Network or graph structures, like that shown in Fig. 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 are 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 maintenance while minimising disruption to passengers; or finding the most useful times and locations to add new services to reduce overcrowding and 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 et al. 2012). 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 et al. 2012). 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).

Fig. 1 Example of a simple, directed network graph with three vertices (A–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



Models based on Space–Time Autoregressive Integrated Moving Average (STARIMA) have been developed for predicting traffic conditions in road networks (Cheng et al. 2011, 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 (Ermagan 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 (Cheng et al. 2011, 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 (Cheng et al. 2011). LSTARIMA additionally accounts for spatial heterogeneity by allowing spatial weight matrices to differ across locations at a given time (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; Hwang 2000).

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 (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; Hwang 2000; 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 (Freeman 1989; Hawawini 1978; Rossana and Seater 1995; Wei 1981; Weiss 1984). High frequency variation is masked by aggregation due to averaging or summing multiple data measurements in one temporal unit (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 extend of the longest delay for each origin–destination pair and calculates a single function representing the average speed of journeys across the whole network.

2 Methods

2.1 Data

2.1.1 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 2009 London Underground Map, and data detailing the zones of the Underground stations, compiled from the 2010 London Underground Map (the 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.

2.1.2 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-h 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.

2.2 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 was 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 2020) 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 (see Brunsdon and Comber 2020).

2.2.1 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 Monte-Carlo Markov Chain (MCMC) estimation procedure with Gibbs sampling algorithm (Geman and 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 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 non-spatial 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 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 m 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 (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 1–20 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 Eq. 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; $basis_n(t)$ 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 α represents the coefficient for the autoregressive error structure estimated by the model.

$$\begin{aligned}
 \text{journeytime}_{i,\text{from},\text{to}} &= \beta_{0,\text{from},\text{to}} + \sum_{n=1}^{N_{\text{basis}}} \beta_{n,\text{from},\text{to}} \text{basis}_n(t_i) + e_{i,\text{from},\text{to}} \\
 \beta_{0,\text{from},\text{to}} &= \beta_0 + u_{0,\text{from}} + u_{0,\text{to}} + v_{0,\text{from}} + v_{0,\text{to}} \\
 \beta_{n,\text{from},\text{to}} &= \beta_n + u_{n,\text{from}} + u_{n,\text{to}} + v_{n,\text{from}} + v_{n,\text{to}} \\
 e_{i,\text{from},\text{to}} &= \alpha e_{i-1,\text{from},\text{to}} / (t_i - t_{i-1} + 1)
 \end{aligned} \tag{1}$$

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) (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 non-spatial 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 α 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 (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 20,000 iterations with 19,000 burn-in, leaving 4000 posterior samples to describe posterior distributions.

2.2.2 Extracting pattern features

To carry out the set aims, two pattern features were extracted: the time of maximum delay and the maximum delay. 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 2020). 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 (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.

2.2.3 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 cover distance d , at a single point in time, T , we used Eq. 2, where n indexes each origin–destination connection and t the time.

$$s = \frac{\sum_{i=1}^n (d_n/T_{n,t})}{n} \quad (2)$$

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.

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 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).

3 Results

Figure 2 shows a map of the London Underground network. The subsection used in simulations is shown in Fig. 3.

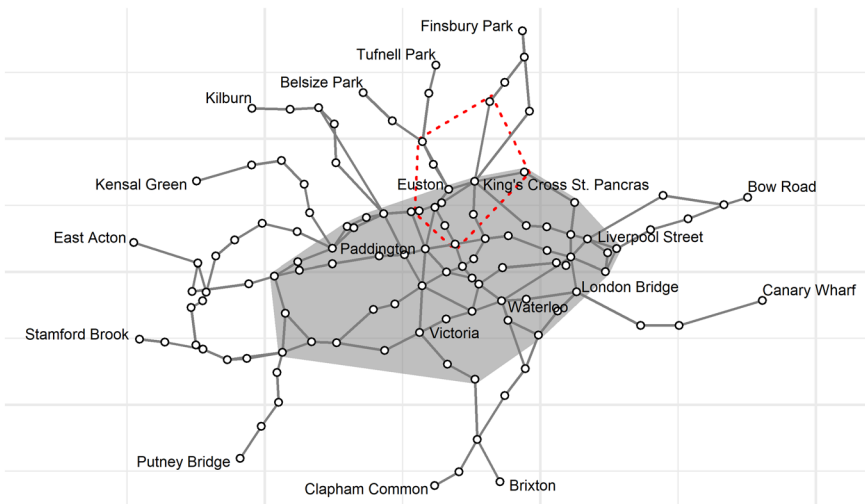


Fig. 2 Map of Zones 1 and 2 of the London Underground network. Zone 1 is shaded in grey. The area used to generate simulated data is outlined with a red dashed line

Fig. 3 Map of the section of the London Underground network used to generate simulated data



3.1 Real data

There were 27,085 journeys from the sample data that took place in the study period chosen. Table 1 summarises network properties for both the London Underground infrastructure (that is shown in Fig. 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

Table 1 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

Property	Value
<i>London Underground infrastructure</i>	
Number of vertices	118
Network centrality (closeness)	0.14
Edge density	0.02
Mean shortest path length	7.11
<i>Journey connections recorded</i>	
Number of vertices	118
Network centrality (closeness)	0.47
Edge density	0.48

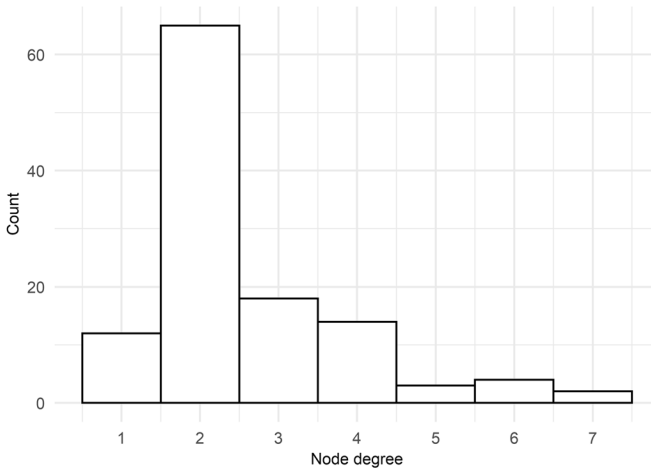


Fig. 4 Vertex degree distribution for the London Underground infrastructure network

is low, suggesting that the network is sparsely connected. Closeness, the centrality measure used in Table 1, represents how close (in terms of geodesic distance) each vertex is to others in the network (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 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 Fig. 4, where most of the vertices in the network have the same degree (two).

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 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 Fig. 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.

3.1.1 Summary of pattern features

Table 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

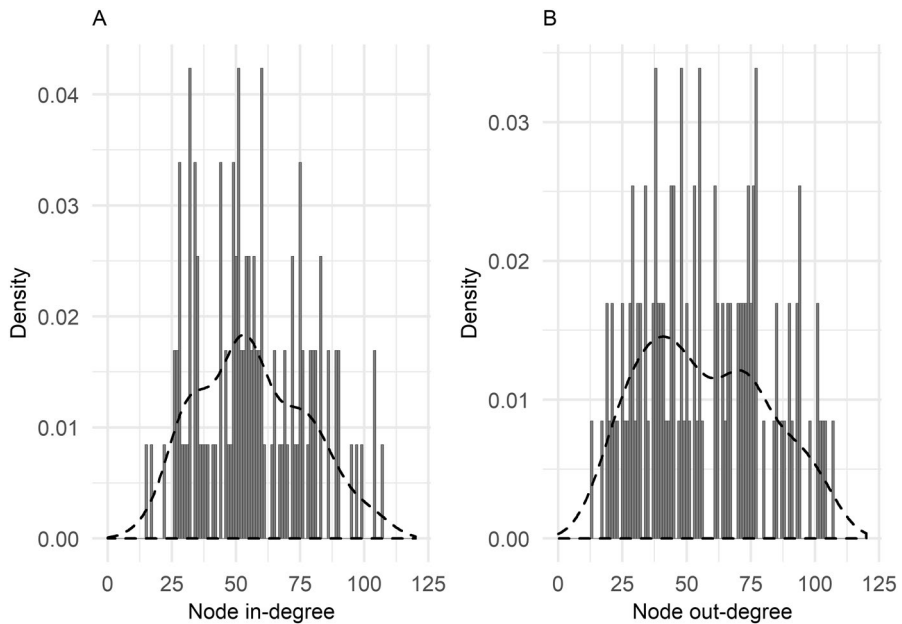


Fig. 5 Vertex degree distribution for the network formed of connections made between London Underground stations in the Oyster card data. **a** In-degree and **b** out-degree

be travelling home from 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 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.

3.1.2 Relationship with vertex centrality

Figure 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

Table 2 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

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

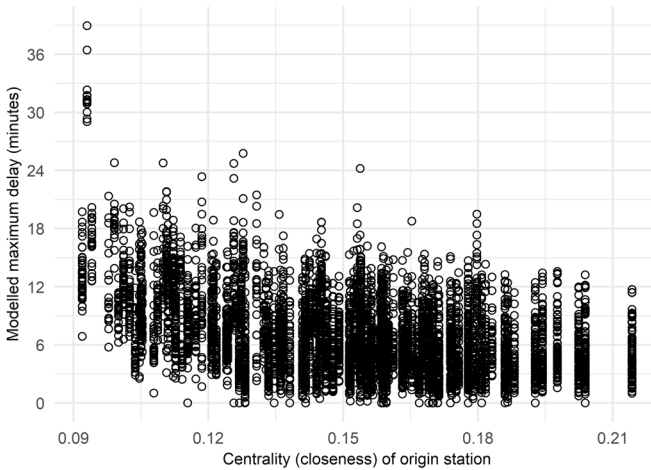


Fig. 6 Relationship between origin station centrality and maximum delay values for origin–destination pairs in the Oyster card data sample

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.

3.1.3 Continuous function of average speed

Figure 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 wide at the start and end of the time period. Throughout the time period, average journey speeds range from approximately 12.5 and 19 km 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 2.

3.2 Simulated data

Table 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 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 Fig. 8 which shows that all stations in this network have a degree between one and five.

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

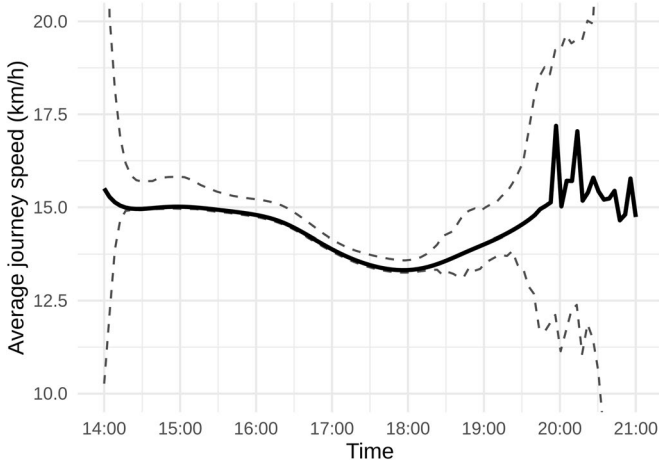


Fig. 7 Continuous temporal function of average journey speed across the whole network. Dashed lines in **a** 95% credible intervals. **b** The function without credible intervals for scaling purposes

Table 3 The network properties of the section of London Underground infrastructure used to generate simulated data

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

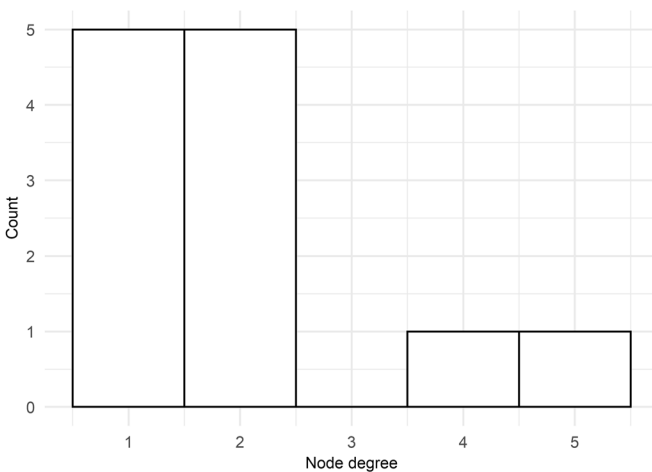


Fig. 8 Vertex degree distribution for the section of the London Underground infrastructure network used to generate simulated data

Table 4 The mean and standard deviation of simulated and modelled time of maximum delay and maximum delay

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

The differences between simulated and modelled values are also summarised, along with the width of the 95% credible intervals (CIs) associated with estimates

3.2.1 Summary of pattern features

Table 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 mean values very close to those simulated, with a difference of approximately three 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 (21.6 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 under 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 one hour 52 minutes, with larger variation between individuals (standard deviation of one hour 29 minutes).

3.2.2 Relationship with simulated values

Figures 9 and 10 show 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

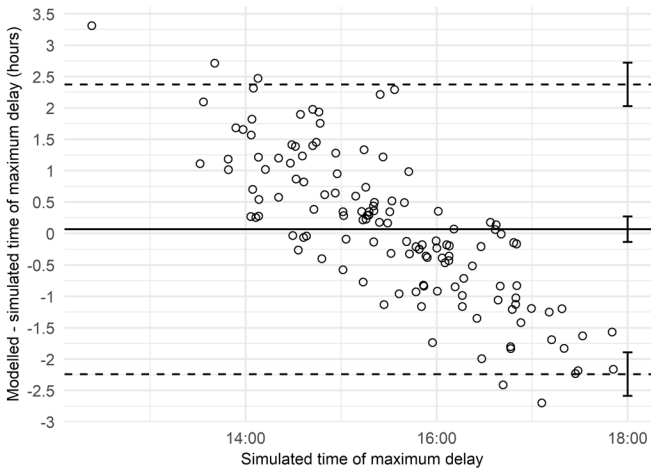


Fig. 9 Relationship between over or underestimation of time of maximum delay by models and the simulated time of maximum delay. Horizontal lines show mean bias (solid) and 95% limits of agreement (dashed) with 95% confidence intervals

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 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 $0.706-0.947$). 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.

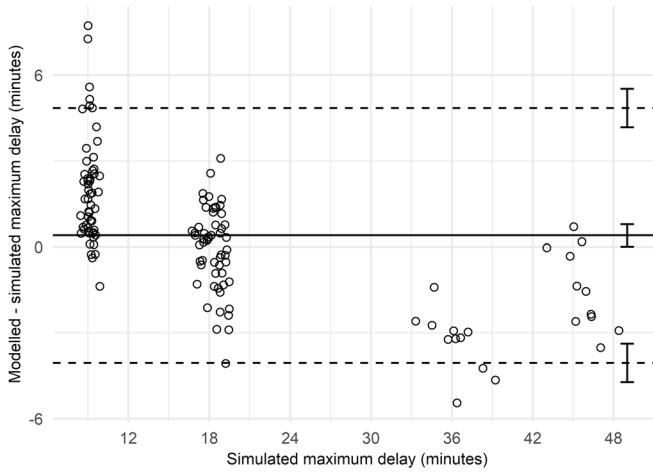


Fig. 10 Relationship between over or underestimation of maximum delay by models and the simulated maximum delay. Horizontal lines show mean bias (solid) and 95% limits of agreement (dashed) with 95% confidence intervals

3.2.3 Relationship with vertex degree

Figure 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 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

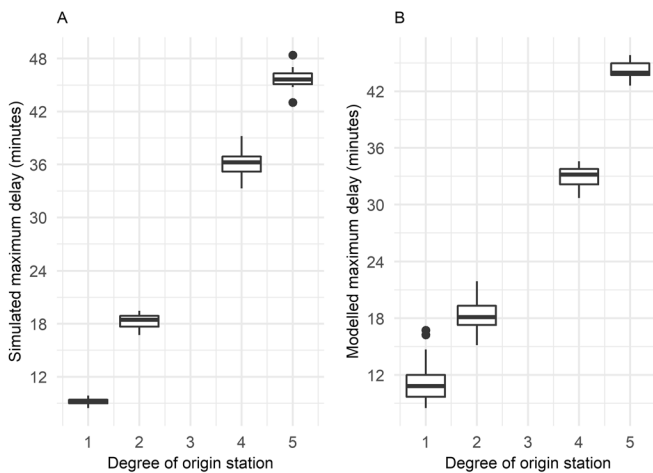
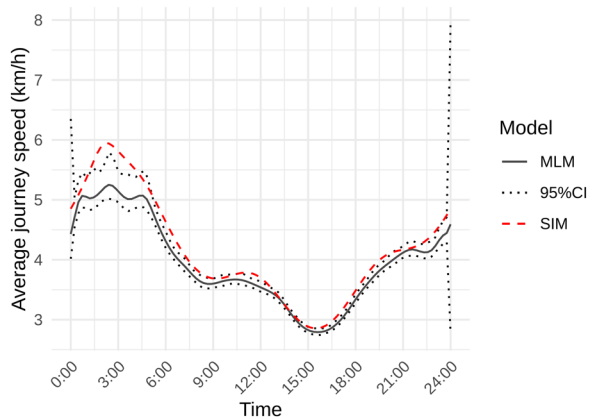


Fig. 11 Relationship between degree of origin station and the simulated (a) and modelled (b) maximum delay for origin–destination pairs in the simulated data

Fig. 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 (dotted black line) (color figure online)



maximum delay has a higher median value and larger spread than the simulated values.

3.2.4 Continuous function of average speed

Figure 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 width for this function is very 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 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).

4 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 continuous 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 days 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 Eq. 1) as the mean value of the non-spatial random effects (u_0 and u_n in Eq. 1), rather than fixing their mean to zero has been suggested to improve mixing (Browne 2004). However, results from models with this parameterisation 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 (Cheng et al. 2011, 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 (Freeman 1989; Hawawini 1978; Rossana and Seater 1995; 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 (Cheng et al. 2011, 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.

4.1 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 min. 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—under 4 min. 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 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 uncertainties 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 (Matthew 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 are in the city centre.

5 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 prevision 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.

Supplementary material The online version of this article (<https://doi.org/10.1007/s10109-020-00345-z>) contains supplementary material, which is available to authorized users.

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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References




- Aigner W, Miksch S, Müller W, Schumann H, Tominski C (2007) Visualizing time-oriented data—a systematic view. *Comput Graph* 31:401–409. <https://doi.org/10.1016/j.cag.2007.01.030>
- Angeloudis P, Fisk D (2006) Large subway systems as complex networks. *Phys A* 367:553–558
- Austwick MZ, O'Brien O, Strano E, Viana M (2013) The structure of spatial networks and communities in bicycle sharing systems. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0074685>
- Basag J, Kooperberg C (1995) On conditional and intrinsic autoregression. *Biometrika* 82:733–746. <https://doi.org/10.2307/2337341>

- Besag J, York J, Mollié A (1991) Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 43:1–20. <https://doi.org/10.1007/BF00116466>
- Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160. <https://doi.org/10.1177/096228029900800204>
- Blonder B, Wey TW, Dornhaus A, James R, Sih A (2012) Temporal dynamics and network analysis. *Methods Ecol Evol* 3:958–972. <https://doi.org/10.1111/j.2041-210X.2012.00236.x>
- Browne WJ (2004) An illustration of the use of reparameterisation methods for improving MCMC efficiency in crossed random effect models. *J Multilevel Model Newsl* 16:13–25
- Browne WJ, Draper D (2006) A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *J Bayesian analysis* 1:473–514
- Brunsdon C, Comber A (2020) Opening practice: supporting reproducibility and critical spatial data science. *J Geograph Syst.* <https://doi.org/10.1007/s10109-020-00334-2>
- Bürkner P-C (2017) brms: an R package for Bayesian multilevel models using Stan. *J Stat Softw* 80:28. <https://doi.org/10.18637/jss.v080.i01>
- Carpenter B et al (2017) Stan: a probabilistic programming language. *J Stat Softw* 76:32. <https://doi.org/10.18637/jss.v076.i01>
- Cheng T, Wang J, Haworth J, Heydecker B, Chow A (2011) Modelling dynamic space-time autocorrelations of urban transport network. In: *Proceedings of the 11th international conference on Geocomputation 2011*, pp 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 Geogr Anal* 46:75–97
- Cole TJ, Donaldson MDC, Ben-Shlomo Y (2010) SITAR—a useful instrument for growth curve analysis. *Int J Epidemiol* 39:1558–1566. <https://doi.org/10.1093/ije/dyq115>
- Comber A, Wulder M (2019) Considering spatiotemporal processes in big data analysis: insights from remote sensing of land cover and land use. *Trans GIS* 23:879–891. <https://doi.org/10.1111/tgis.12559>
- Congdon P (2014) *Applied Bayesian modelling*. Wiley, London
- de Haan-Rietdijk S, Voelkle MC, Keijsers L, Hamaker EL (2017) Discrete- vs. continuous-time modeling of unequally spaced experience sampling method data. *Front Psychol* 8:1849. <https://doi.org/10.3389/fpsyg.2017.01849>
- Dung VT, 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:e0173857. <https://doi.org/10.1371/journal.pone.0173857>
- Ermagun A, Levinson D (2018) An introduction to the network weight matrix. *Geogr Anal* 50:76–96. <https://doi.org/10.1111/gean.12134>
- Erős T, Schmera D, Schick RS (2011) Network thinking in riverscape conservation—a graph-based approach. *Biol Conserv* 144:184–192. <https://doi.org/10.1016/j.biocon.2010.08.013>
- Freeman LC (1978) Centrality in social networks conceptual clarification. *Social Netw* 1:215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Freeman JR (1989) Systematic sampling, temporal aggregation, and the study of political relationships. *Polit Anal* 1:61–98. <https://doi.org/10.1093/pan/1.1.61>
- Freni-Sterrantino A, Ventrucci M, Rue H (2018) A note on intrinsic conditional autoregressive models for disconnected graphs. *Spat Spatio-tempor Epidemiol* 26:25–34. <https://doi.org/10.1016/j.sste.2018.04.002>
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell PAMI* 6:721–741. <https://doi.org/10.1109/TPAMI.1984.4767596>
- Goldstein H (1994) *Multilevel cross-classified models*. *J Sociol Methods R* 22:364–375
- Goldstein H (2011) *Multilevel statistical models*, 4th edn. Wiley, Chichester
- Hawawini GA (1978) A note on temporal aggregation and serial correlation. *Econ Lett* 1:237–242. [https://doi.org/10.1016/0165-1765\(78\)90031-9](https://doi.org/10.1016/0165-1765(78)90031-9)
- Heck RH, Thomas SL (2015) *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*, 3rd edn. Taylor & Francis, London
- Holmes CC, Mallick BK (2003) Generalized nonlinear modeling with multivariate free-knot regression splines. *J Am Stat Assoc* 98:352–368. <https://doi.org/10.1198/016214503000143>
- Horvath S (2011) *Weighted network analysis: applications in genomics and systems biology*. Springer, New York

- Howe LD et al (2013) Linear spline multilevel models for summarising childhood growth trajectories: a guide to their application using examples from five birth cohorts. *Stat Methods Med Res* 25:1854–1874. <https://doi.org/10.1177/0962280213503925>
- Hwang S (2000) The effects of systematic sampling and temporal aggregation on discrete time long memory processes and their finite sample properties. *Econ Theory* 16:347–372. <https://doi.org/10.1017/S0266466600163030>
- 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. *Transp Res Part C Emerg Technol* 85:591–608. <https://doi.org/10.1016/j.trc.2017.10.016>
- Kim K, Sentürk D, Li R (2011) Recent history functional linear models for sparse longitudinal data. *J Stat Plan Inference* 141:1554–1566. <https://doi.org/10.1016/j.jspi.2010.11.003>
- Leigh C, Kandanaarachchi S, McGree JM, Hyndman RJ, Alsibai O, Mengersen K, Peterson EE (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. *Stat Med* 28:3049–3067
- Matthew S, Emily P, Ellena B (2017) Informative observation in health data: association of past level and trend with time to next measurement
- Opsahl T, Panzarasa P (2009) Clustering in weighted networks. *Soc Netw* 31:155–163. <https://doi.org/10.1016/j.socnet.2009.02.002>
- Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: generalizing degree and shortest paths. *Soc Netw* 32:245–251. <https://doi.org/10.1016/j.socnet.2010.03.006>
- Paccagnella O (2006) Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *J Eval Rev* 30:66–85
- Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M (2019) A review of spline function procedures in R. *BMC Med Res Methodol* 19:46. <https://doi.org/10.1186/s12874-019-0666-3>
- Peterson EE, Ver Hoef JM (2010) A mixed-model moving-average approach to geostatistical modeling in stream networks. *J Ecology* 91:644–651
- Pfeifer PE, Deutsch SJ (1980) A STARIMA model-building procedure with application to description and regional forecasting. *J Trans Inst Brit Geogr* 5(3):330–349
- Plummer M (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: 3rd international workshop on distributed statistical computing (DSC 2003); Vienna, Austria 124
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienne
- Ramsay JO, Silverman BW (1997) Functional data analysis. Springer, London
- Rossana RJ, Seater JJ (1995) Temporal aggregation and economic time series. *J Bus Econ Stat* 13:441–451. <https://doi.org/10.1080/07350015.1995.10524618>
- Scott J (1988) Social network analysis. *Sociology* 22:109–127. <https://doi.org/10.1177/0038038588022001007>
- Soetaert K, Herman P (2009) rootSolve: nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations. R-package version 1.6
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Ser B (Stat Methodol)* 64:583–639. <https://doi.org/10.1111/1467-9868.00353>
- Stimson JA, Carmines EG, Zeller RA (1978) Interpreting polynomial regression. *Sociol Methods Res* 6:515–524. <https://doi.org/10.1177/004912417800600405>
- Thomas A, Best N, Lunn D, Arnold R, Spiegelhalter D (2014) GeoBUGS user manual
- Ver Hoef JM, Peterson EE (2010) A moving average approach for spatial statistical models of stream networks. *J Am Stat Assoc* 105:6–18
- Wang W, Yan J (2020) splines2: regression spline functions and classes
- Wei WWS (1981) Effect of systematic sampling on arima models. *Commun Stat Theory Methods* 10:2389–2398. <https://doi.org/10.1080/03610928108828197>
- Weiss AA (1984) Systematic sampling and temporal aggregation in time series models. *J Econ* 26:271–281. [https://doi.org/10.1016/0304-4076\(84\)90022-8](https://doi.org/10.1016/0304-4076(84)90022-8)
- Xiao N (2016) Basic geometric operations. In: GIS algorithms: theory and applications for geographic information science & technology. SAGE Publications Ltd
- Yeh R, Nashed YSG, Peterka T, Tricoche X (2020) Fast automatic knot placement method for accurate B-spline curve fitting. *Comput Aided Des* 128:102905. <https://doi.org/10.1016/j.cad.2020.102905>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Sarah C. Gadd^{1,2}  · **Alexis Comber**¹  · **Mark S. Gilthorpe**^{2,3,4}  ·
Keiran Suchak^{1,2}  · **Alison J. Heppenstall**^{1,2,3} 

Sarah C. Gadd
mt15scg@leeds.ac.uk

- ¹ Centre for Spatial Analysis and Policy, School of Geography, University of Leeds, Leeds LS2 9JT, UK
- ² Leeds Institute for Data Analytics, University of Leeds, Leeds LS2 9JT, UK
- ³ The Alan Turing Institute, London NW1 2DB, UK
- ⁴ School of Medicine, University of Leeds, Leeds LS2 9JT, UK