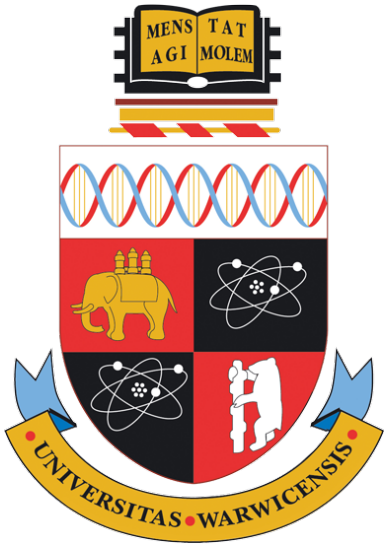
WARWICK
THE UNIVERSITY OF WARWICK



Data science perspectives on problems in intelligent transportation systems and mobility

by

Kieran Kalair

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Mathematics for Real-World Systems CDT

March 2021

Contents

List of Tables

List of Figures

List of Algorithms

Acknowledgments

Firstly, I thank my supervisor Colm Connaughton for guidance over the course of my project. Secondly, I am grateful to the many industry experts I have had discussions with throughout the project. These include Steve Hilditch and his colleges from Thales UK, in-particular for sharing expertise on the operation of NTIS and UK transportation systems more generally. Additionally Anna Watt and her colleges from Transport for the West Midlands and Harita Joshi and her colleges from Jaguar Land Rover. Further, I am grateful to the collaborators I have worked with throughout different stands of work. The first of these is Pierfrancesco Alaimo Di Loro for discussions on modelling and validation aspects of point processes. The second are Harita Joshi, Arun Raveendran and Andrew Fairgrieve, for extensive discussions on autonomous driving.

Finally, thanks to those within our CDT, particularly to Ayman Boustati and Alvaro Cabrejas Egea for helpful insight on data-acquisition and pre-processing, and for general discussions with Bhavan Chahal, Kutlwana Bashe, Steve Bennett, Andrew Hilditch, Charlotte Roman, Harvey Devereux and Arthur King that provided useful insight into different views on problems.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.
Parts of this thesis have been published by the author or are currently under review.

(1) Chapter 4 and its appendix has been published as:
K. Kalair and C. Connaughton. Anomaly detection and classification in traffic flow data from fluctuations in the flow-density relationship. *Transportation Research Part C: Emerging Technologies*, 127:103178, 2021

(2) Chapter 5 and its appendix has been published as:
K. Kalair, C. Connaughton, and P. Alaimo Di Loro. A non-parametric Hawkes process model of primary and secondary accidents on a UK smart motorway. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70(1):80–97, 2021

(3) Chapter 6 and its appendix has been published as:
K. Kalair and C. Connaughton. Dynamic and interpretable hazard-based models of traffic incident durations. *Frontiers in Future Transportation*, 2:15, 2021

(4) Jaguar Land Rover are pursuing patents relating to chapter 7.

(5) Parts of the data processing described in chapter 3 are discussed in:
K. Kalair and C. Connaughton. Large scale performance assessment of the Lighthill-Whitham-Richards model on a smart motorway. In *2018 21st*

International Conference on Intelligent Transportation Systems (ITSC), pages 3718–3723, Maui, HI, November 2018. IEEE

Abstract

This thesis is a body of work applying data science and mathematical modelling to problems in intelligent transportation systems. Utilising data collected from the M25 London orbital, four problems relevant to both industry and academia are considered. In chapter 4 we develop a novel methodology for anomaly detection on road networks. We determine a data-driven region of typical behaviour in the flow-density plane, tracking fluctuations from this to identify anomalies in real time. We find this offers generally comparable performance to existing methods, but is clearly superior when the distribution of speeds conditioned on time of week is bimodal.

In chapter 5, we quantify the prevalence of primary and secondary traffic incidents in our data using a novel self-exciting point process. The self-excitation component suggests 6-7% of incidents are most likely secondary, occurring temporally and spatially in the wake of other incidents. Our modelling further identifies two spatial hotspots and captures commuting patterns in the UK. We are able to apply out-of-sample validation and show the model is statistically defensible.

Chapter 6 explores dynamic prediction of incident durations. We find non-parametric neural network models offer strong performance compared to a range of alternative candidates, achieving errors below current industry targets. By exploring feature importance, we find time series prove informative for predictions on short horizons whereas time of day and location do so at longer horizons.

We explore an emergent behaviour path planning model in the context of autonomous vehicles in chapter 7. This was developed in conjunction with engineers from Jaguar Land Rover and incorporates practical constraints real-world vehicles must satisfy. Formulating an optimization problem incorporating comfort, safety and progress, we show dynamically solving this results in emergent complex driving behaviours: vehicle following, passing and overtaking. Safety is based on a distributional prediction of drivers behaviours, with its variance indirectly defining properties of the emergent behaviours.

Our findings throughout this work offer models and methodologies that can be used to improve the management and better understand the behaviour of existing transportation infrastructure, as well as the development of future technologies.

Sponsorships and Grants

This work was supported by the EPSRC and MRC Centre for Doctoral Training in Mathematics for Real World Systems under grant number EP/L015374/1.

Acronyms

AFT Accelerated failure time.

AIC Akaike information criterion.

AMISE Asymptotic mean integrated square error.

ANPR Automatic number plate recognition.

APE Absolute percentage error.

AUROC Area under the receiver operator curve.

CDF Cumulative density function.

CNN Convolutional neural network.

DFP Deviation from profile.

DFTB Deviation from typical behaviour.

DR Detection rate.

DTW Dynamic time warping.

FAR False alarm rate.

GPS Global positioning system.

HAC Hierarchical agglomerative clustering.

IP Interior point.

IQR Inter-quartile range.

KDE Kernel density estimation.

LIDAR Light detection and ranging.

LIMA Local indicators of mobility association.

LIME Local interpretable model-agnostic explanations.

LN Log-normal.

LSTM Long-short term memory.

LUD Lower uncongested data bound.

MAD Median absolute deviation.

MAE Mean absolute error.

MAPE Mean absolute percentage error.

MIDAS Motorway incident detection and automatic signalling.

MISE Mean integrated square error.

MPC Model predictive control.

MPE Mean percentage error.

MPH Miles per hour.

MTTD Mean time to detect.

NN Neural network.

NP Non-parametric.

NTIS National traffic information service.

PDF Probability density function.

PE Percentage error.

PI Performance index.

PL Partial likelihood.

QQ Quantile-Quantile.

RGB Red green blue.

RMSE Root mean square error.

RRT Rapidly-exploring random tree.

RSF Random survival forest.

RTN Return to normal.

SHAP Shapley additive explanations.

SMS Space mean speed.

SND Standard normal deviate.

SRN Strategic road network.

SW Sliding window.

TMS Time mean speed.

TSF Time series feature.

V2X Vehicle to everything.

XML Extensible markup language.

Chapter 1

Introduction

Efficient and reliable transportation infrastructure is a fundamental component of an advanced economy [5]. Road infrastructure in-particular accounts for 90% of passenger journeys and 70% of freight transport in the UK [6]. As a result, effective management of the country's road network is of significant practical concern. In many countries, including the UK, there is limited scope to build additional infrastructure to cope with the demand for road traffic. Instead, the focus is on approaches referred to as 'intelligent mobility' and 'intelligent transportation systems'. These broadly describe efforts to improve the usage of existing infrastructure by combining real-time data, mathematical or statistical models and communication of information to users. These efforts can be both network-wide or on the scale of individual commuters and their vehicles. We will consider the network-wide problems to fall under the umbrella of intelligent transpiration systems, and those concerning individuals and their vehicles to fall under intelligent mobility. An appropriate test-bed for the network-wide approaches is the Strategic Road Network (SRN), which constitutes around 4,400 miles of motorways and major trunk roads across England [7]. The SRN carries 30% of all traffic in the country, with 4 million vehicles using it each day and 1 billion tonnes of freight being transported across it each year [8]. Whilst a large component of transportation infrastructure in England, congestion remains a significant problem on the network, with [8] suggesting 75% of businesses consider tackling congestion on the SRN to be important or critical to their business.

There is little chance that in the short term, the SRN will see significant infrastructure changes, however data describing the traffic state across the network is already being collected and made available for analysis through the National Traffic Information Service (NTIS)[1] [9]. There is therefore significant scope to address a range of problems faced by commuters, traffic operators and

[1]Technical details of the NTIS data feeds are available at `http://www.trafficengland.com/services-info`.

traffic management centres using this data. Some of these problems relate to incidents on the network, for example detecting where incidents are occurring and altering operators of their existence, or making informed predictions of how long a part of the network will be impacted by an incident. Others include identifying locations for infrastructure improvement, and more future facing tasks on the microscopic scale considering challenges in the application of autonomous vehicles. Fundamentally, one of the major challenges in intelligent mobility and intelligent transportation systems is to understand how best to utilise the available data, in conjunction with statistical, mathematical and engineering methods to address these types of problems, whilst ensuring solutions are practically implementable and robust.

We focus on four problems relevant to both industry and academia throughout this thesis. Literature relevant to each of these problems and that has influenced our modelling is reviewed in chapter 2. We then detail the data from the SRN that we extract and analyse in chapter 3. In chapter 4, we consider how to detect incidents that occur on the network and alert operators in a timely manner. To do so, a consideration of what constitutes 'normal' or baseline behaviour is required. We explore what natural definitions arise from an analysis of the data, and how one can use these in real-time to detect incidents as they occur and develop. The agreement between incidents detected by the proposed methodology and those registered in the data is investigated, providing validation for the practical application of the method.

In chapter 5, we consider how one might model the spatio-temporal variation in the occurrence of incidents on the network, and evidence for different incident generation mechanisms in the data. We apply a model that distinguishes between primary and secondary incidents, with primary incidents occurring at some 'background' rate that varies in space and time, and secondary incidents occurring after an initial incident, at short time and length scales relative to the background. We show such a model is statistically valid on out-of-sample data, and investigate how the components of the model vary spatially across the domain, temporally across seasons, and when the data is restricted to incidents that result in large speed drops on the network.

The third problem, addressed in chapter 6, is predicting the duration of incidents on the network. By duration, we mean the time between an incident starting and the traffic behaviour at the incident location returning to some baseline. We specifically consider how one can dynamically update duration predictions using real-time data taken from the incident location, reducing uncertainty and improving prediction performance as more data becomes available. We ensure the models are interpretable and explore what factors influence predictions at different time horizons.

In chapter 7, we look further into the future and to a more microscopic scale,

considering how paths through uncertain environments can be generated in the context of autonomous vehicles, whilst meeting practical, comfort and safety requirements. An optimisation problem is formulated accounting for these requirements, from which a range of natural driving behaviours are emergent. Such an approach avoids the specification of explicit rules defining driving behaviours, rather they emerge as a balance between the costs associated with performing certain actions and occupying certain states, along with the uncertainty in other drivers behaviours.

Whilst the problems we tackle are all challenges in the domains of intelligent transportation systems and intelligent mobility, the methodological connections between them run deeper. We repeatedly consider distributional predictions to account for various forms of uncertainty, and often utilise non-parametric statistical methods to capture variation inherent to the data. We utilise various tools to retain interpretability and understanding of our methods, through either the construction of the methods themselves, limiting freedom of certain model components, or applying methods from machine learning to interpret 'black box' models. Further, our work provides both practical tools and insight for traffic management centres and operators, identifying areas for improvement to existing infrastructure and tools to help better manage situations or develop future methodologies. This research has been guided by interactions with industry experts from Thales, Jaguar Land Rover and Transport for the West Midlands to ensure our analysis and recommendations are of practical use and address pressing industry concerns.

Chapter 2

Background

In this chapter, we review literature relevant to each of the challenges we consider. We also offer an overview of some of the relevant mathematical and statistical methodologies we apply throughout our work. Before we begin, we note that different areas of the literature may tend to use different terms, either 'events' or 'incidents' to describe situations that occur on the network, for example crashes, breakdowns and abnormal traffic jams. We will refer to such situations as incidents throughout this work.

2.1 Automatic incident detection

Multiple approaches have been proposed for the task of anomaly detection as applied to traffic incident detection. They broadly fall into two classes. The first class uses data from individual vehicles. Examples include the use of data from automatic vehicle identification systems as in [10], the identification of single vehicles using cameras in [11] or the use of global positioning system (GPS) and social media data from navigation applications such as Waze in [12]. The second class uses time series data from embedded loop sensors, usually aggregated measures of traffic. Typically, these include one or more of vehicle counts, occupancy or density and speed. These methods do not identify individual vehicles and are often statistical in nature. We explore methods of the second class in chapter 4.

Various approaches that utilise data derived from loop sensors have been considered throughout the literature. Firstly, basic pattern matching is an approach which has proven to be successful in some contexts. The idea is to determine some features or thresholds in time series data that indicate incident and non-incident windows. Early and well known examples of this are the California algorithms and their variants, discussed in [13, 14]. These algorithms compare occupancy values at adjacent sensors, looking for when the pair-wise comparison metrics pass some thresholds. As a specific example, for

two sensors i and $i+1$, where $i+1$ is downstream of i, the California Algorithm #7 computes the occupancy difference at time t between sensor i and $i+1$, as well as the relative occupancy difference between the two. Finally, it compares the downstream occupancy to some threshold, and outputs an incident free, tentative incident, incident occurred or incident continuing flag. Modifications are made to the logic for other variants. To apply this in practice, one needs labelled time series data at the loop level, with labels specifying when and where incidents occurred. The model thresholds can then calibrated based on this data, and the algorithm applied in real-time.

A second popular approach is the so-called McMaster algorithm, based on catastrophe theory, discussed in [15], and further developed in [16–18]. The initial observation for the approach is that in uncongested regimes, flow-occupancy data typically fits a linear relationship with little scatter. From this uncongested data, one can construct a lower bound of flow-occupancy data through some parametric form, and also define critical occupancies and flows. Together, these segment the flow-occupancy diagram into sections that are then considered to represent system states, either uncongested, bottleneck flow or congested. If a sensor is considered to be congested, then the next sensor downstream is examined, determining if this is also congested, or appears uncongested. If the downstream detector is uncongested, then there is likely an incident between the two examined detectors, otherwise there is likely a problem somewhere further downstream that has impacted further upstream. Such an approach requires calibration of particular thresholds, which may differ from station to station and also requires fitting some parametric form of the lower flow-occupancy bound. Subsequent work on the McMaster algorithm has explored how to correctly calibrate these parameters, with [19] considering a particle-swarm approach. Our work in chapter 4 also considers a segmentation of data into different regions, however we consider a non-parametric approach to avoid introducing a large number of parameters to fit, allowing the data to define the regions itself.

A third important approach is the standard normal deviate (SND) methodology, an early use being in [20]. These methods attempt to construct some average value μ_{SND} for a traffic variable on a particular time interval, and some measure of variation σ_{SND}. Then, one can question if new data, for example speed, is below some threshold value of $\mu_{\text{SND}} - c\,\sigma_{\text{SND}}$, determining the parameter c to best fit available data. Of course, if one simply uses a mean and standard deviation for this, there is a risk that outliers may significantly influence these values. To account for this, one can pre-filter a dataset to remove any known incident periods before computation as in [21], or use robust summary statistics as in [22]. Extensions of this idea include incorporation of spatial information, seen in [23], where the total delay time caused by traffic

accidents is examined by considering the propagation backwards in space and forwards in time of traffic queues. Further, [24] considers robust construction of noise thresholds and incorporates smoothing, first determining a robust threshold for each particular interval of a day, then smoothing these using a bilateral filter to incorporate spatial-temporal correlations, assuming nearby windows of space and time should have reasonably similar thresholds. The methodology we develop in chapter 4 is similar in some sense to a robust SND methodology, however we detail further in that chapter how we construct a robust threshold for normality from a joint representation of flow-density data, and investigate how deviations from this correspond to incidents on the network.

Closely related to the SND family of algorithms are various approaches based on change-point detection. Such ideas are considered in [25], where the authors use a Markov structured model to predict the transitions between a congested and free flow state. Additionally, [26] based their methodology on forecasting with ARIMA models, classifying an incident period as one where the data lay outside the confidence bounds of their forecasting model. Approaches such as this one assume that the confidence limits provided by a forecasting model are indeed representative of the true uncertainty present in the data, and as a result require extensive time series validation. The question of what is an optimal model to forecast traffic states, in the wide range of potential scenarios is also a complex one. These methods, along with the SND methods, have the advantage that they can be applied without any changes to both the individual loop level, or at the aggregated link level.

However, calibrating any form of incident detection method is generally difficult because of data quality problems [27]. These problems include incidents that are present in the data but were not flagged and labelled by operators, delays in the reporting of incident occurrence times, and lack of location information. These can present particular difficulty for methods that rely on pre-filtering data and removing incidents to determine normal behaviour, but still result in difficulties more generally for determining the detection rate, false alarm rate and detection time for methodologies. To combat this, some authors use simulated data, as in [28, 29], however how close this simulated data is to that generated from a real-world location of interest may vary in practice.

Recently, alternative methodologies have emerged for clustering of data into typical and anomalous states to discover non-recurrent congestion as in [30], where higher than expected journey times are distinguished from typical ones using expected journey times scaled by a congestion factor. Further, clustering of vehicle trajectories is performed in [31], creating a probabilistic tree-like clustered structure that can used to identify anomalous incidents. Additionally

[32] use a Gaussian Mixture Hidden Markov Model to extract features and classify traffic states from video data into categories of: empty, open flow, mild congestion, heavy congestion and stopped.

Finally we mention that machine learning and deep learning methodologies have seen significant developments in their application to incident detection in recent years. Some of these use traditional datasets but new methodologies, for example [33] which incorporate traffic data along with weather information and spatial structure into a convolutional-LSTM model to predict the time and location of traffic accidents. Additional deep learning approaches include [34], which uses a convolutional neural network for anomaly detection on the Luxembourg road network and [35] which again uses a convolutional neural network but for incident detection on urban roads in London. For such methods, computational intensity and data quality issues continue to act as barriers.

2.2 Spatio-temporal incident modelling

There is significant literature focusing on analysis and modelling of traffic incidents from a spatio-temporal perspective. The goal of work in this area may be to offer descriptive analysis of incident data to better understand existing traffic patterns, or to develop models for forecasting incident occurrence times and locations. It is observed throughout the literature that incidents along a road network do not occur uniformly, and instead there is some spatial auto-correlation present in the data. A discussion of the evidence for this is given in [36]. Within, the authors discuss previously applied descriptive analysis methods, including K-function analysis and comparison to complete spatially random patterns, and how they showed clear evidence of spatial correlation among incident locations. They then further this work by incorporating spatial components into an auto-regressive model, finding it improved upon models disregarding the spatial correlation in the data. Additional analysis is provided in [37], where the authors focus on an urban network in China and analyse the evolution of incident hot-spots through time. They develop various network spatial analysis methods, extending kernel density estimation on networks, cross K-function analysis, Moran's I and Local Indicators of Mobility Associations (LIMA) and offer various exploratory analysis of the dataset with these methods. They specifically found that network kernel density analysis revealed differing spatial structure of incidents during the day and night whereas cross K-function analysis showed strong association with collisions localised near hotels and low association near sports and recreation centres. Their use of Moran's I considered increases and decreases in incidents on weekdays compared to weekends, finding large commercial areas showed large increases on weekends, and finally their use of LIMA showed that local areas in the network often

showed coherent increasing and decreasing incident numbers.

Further spatial-temporal analysis is completed in [38], where the authors used Kulldorff's space-time scan statistics to determine statistically significant clusters of traffic accidents across the entire UK in 2016. They found two significant clusters, both in the north of the country, but conceded that they do not explicitly account for the network structure in their analysis. Whilst we also consider data from the UK, our work in chapter 5 focuses on a different time-period and on a much smaller region, a single motorway rather than the entire country. We also construct a predictive model of traffic incidents based on point-process, not just discovering locations of statistically significant clusters. Of course, through constructing a model, we can infer spatially and temporally where incidents are most common, but the approach we take later in this work is fundamentally different from [38].

A point of particular interest to our work arises from the modelling of traffic incidents in [39]. Here, logistic regression and random forest models are used to predict the likelihood of incident occurrences. In-particular, the models incorporate a significant 'cascading effect' variable, in which the presence of an incident showed significant influence on the likelihood of another nearby in space and time. Although a different type of model and formulation to the one we explore in chapter 5, there is clearly a sense that cascading effects are a real component in some traffic data that one may want to incorporate into a model. It is unclear what time and length scales are associated with this effect however, and if smart motorway features may remove this effect from the data.

Based on this analysis, it is natural to consider applying spatio-temporal point-process models to traffic incidents. Recent work considering these models on linear networks is given in [40], where some specific comments are made in regards to applying them to traffic incident modelling. In-particular, the authors consider road-networks in Huston, Medellin and Eastbourne and question what features the data show. They find statically significant evidence that incidents on the network did not follow a uniform spatial-temporal Poisson process and that tests indicated favouring clustering of data in space and time. Further, they show clear structure in the incidents data relating to time of day, finding a peak at the evening rush-hour in all datasets. However they focus on urban road networks, whereas we focus on highways throughout this work, in-particular smart motorways. We choose to collect data on a large ring road, detailed in chapter 3, and doing so simplifies the structure of our domain from a network to a continuous space. Indeed, we will use this, and with it a standard distance measure in space throughout the work in chapter 5, rather than a graph distance as in [40]. Our choice of dataset simplifies the fitting procedure of a point-process model from a network space onto continuous space, whilst also offering some insight into a much discussed topic of smart

motorway safety.

Readers interested in the discussed spatial-temporal analysis methodologies that are not explored further in this work can find details of K-function analysis in [41], Kulldorff's space-time scan statistic in [42] and LIMA in [43].

2.3 Incident duration prediction

As well as exploring methods to detect incidents on a network and model their spatio-temporal properties, there is significant work on predicting the duration of incidents. Whilst this is a large body of existing work, many fundamental challenges remain that are both of practical interest to traffic management centres and remain active areas of research in an academic sense. A review of existing work in this area is found in [44], where six future challenges for incident duration prediction are listed. These are: combining multiple data-sources, time sequential prediction models, outlier prediction, improvement of prediction methods through machine learning or alternative frameworks, combining recovery times and accounting for unobserved factors. We base our work in chapter 6 on addressing some of these challenges, and review relevant literature below that covers other aspects that influence our modelling.

Before any methodologies are considered, it is first important to define exactly what is being modelled in existing works focusing on incident duration analysis and prediction. A traffic incident is considered to have four different time-phases: the time taken to detect and report an incident, the time to dispatch an operator to the scene, the travel time of the operator to the scene, and finally the time to clear an incident. Such a framework is described in [44]. Often authors cannot estimate reporting time, and so an incident is considered to have 'started' when it is reported by an operator. Typically, an incident is considered to have ended when it is closed by an operator or management centre, and many works focus on modelling the duration of an incident as measured by the difference between these two times. However, [44] highlights that a future challenge for works in this area is to also model the recovery times of incidents, that is to include the time until the road returns to some normal behaviour in the modelling. We discuss this further and how we account for it in chapter 6.

Whatever explicit definition of duration is used, there is an enormous amount of work on developing models to explain and predict this. For many works, an initial step in modelling is to determine an appropriate distributional form to represent the measured durations. These are typically heavy tailed and empirically show significant variation. Examples of this include modelling the distribution of incident durations as log-normal in [45, 46], log-logistic in [47, 48], Weibull in [49, 50] and generalised F in [51]. In the later, it is

noted that the generalised F distribution can be equivalent to many other distributional forms for particular parameter choices, including the exponential, Weibull, log-normal, log-logistic and gamma distributions. Hence, it offers more freedom than choosing any single one of these forms. Indeed, the authors state that the increased flexibility it offers allows it to fit the data better. Even further flexibility in the distributional choice is given in [52], where it is shown that modelling the distribution as a mixture, that is the sum of multiple components, may improve model performance. Specifically, they consider a model with a single log-logistic distribution, and another model with a 2-component log-logistic mixture distribution, achieving mean absolute percentage errors of 49% and 39% respectively, clearly showing promise in this mixture approach. Finally, [53] had difficulty finding statistically defensible distributional fits to their data, although it should be noted that different definitions of incident durations will likely impact this.

Using some common probability distribution is appealing in some sense as it limits the freedom of a model and can be easier to fit to data. However, it is clear that authors are exploring more complex distributional forms to better model the data and seeing better results when they do so, as in [51] and [52]. Mixture distributions are a naturally appealing form, as we assume the data is generated by multiple sub-populations, and can have different effects of covariates for different populations. Recent applications of survival analysis in healthcare [54] have removed distributional assumptions entirely, and instead formulate models that output distributions with no closed form. This is done by treating the output space as discrete, and treating the model output as a probability mass function (PMF) defined over it, allowing for construction of a fully non-parametric estimate. Such an approach offers even more freedom, and as we see more complex distributions used in the traffic literature to provide similar freedom, one could question if removing the distribution assumption entirely can improve model performance. We incorporate this section of the literature and idea of increasing freedom into our work in chapter 6 by considering models with simple log-normal and Weibull distributional assumptions, and then more complex mixture and non-parametric methods. One of the main questions in that section of our work is does using a non-parametric distribution that has shown success in healthcare have any benefits in the domain of traffic incident duration analysis.

An important point to consider as increasingly complex distributions are used in modelling is overfitting, describing models that fit training data too closley and do not generalise to unseen data. The more parameters a model has, and more complexity built into it, the higher the risk of overfitting. There are a number of approaches taken to avoid this in the literature, and we are mindful to apply them in our work. The first is to judge model performance on

unseen data to determine if, after training, the model is able to generalise and retain performance on new data. This is done throughout chapters 4, 5 and 6. When we specifically consider prediction of incident durations in chapter 6, we not only assess out-of-sample performance, but we also apply variable selection using information criterion for classic models, and elastic-net regularisation for our neural network models. This form of regularisation ensures weight values in the neural networks do not grow to large, as this is typically an indication of overfitting. Further, elastic-net regularisation encourages some weights in the network to have values of 0, thereby reducing the complexity of the network. Dropout is used during training of the neural networks, meaning values for nodes are randomly set to 0 with some probability, ensuring the network does not become too dependent on specific nodes. Finally, training for the neural networks is halted when performance on some unseen dataset begins to decrease, a practice known as early stopping. All of these techniques highlight the fact that whilst we do consider models with significant complexity and freedom, we are also mindful to ensure they generalise to unseen data.

After a distribution is determined, many works apply methods from survival analysis, with a common choice being the accelerated failure time (AFT) model. Example applications of this are given in [55–57]. Such a model assumes that each covariate either accelerates or decelerates the life-time of a particular individual. In practice, this model is equivalent to performing a regression on the log of the durations with some assumed form of noise. These models are widely used, and offer an interpretable means of investigating what factors strongly or weakly influence incident duration. However, it can be difficult to incorporate time series features into them. Whilst it is possible to model time-varying effects of covariates, for example in [58], it is more complex to derive optimal features from a time series that are also interpretable. As well as producing interpretable outputs and relationships between variables, AFT models can incorporate the various distributional forms for incident durations suggested in the literature. The alternative and well known classical survival model some apply is a Cox regression model [59]. Such a model assumes a baseline hazard function for the population, describing the instantaneous rate of incidents, from which survival probabilities can be calculated. Covariate vectors for individuals shift this baseline hazard allowing for individualised predictions. Applications of this to transportation problems are given in [60–62]. We consider both of these models and offer more detail on them in chapter 6.

Whilst these two methods are widely used, a number of alternative methodologies exist. One such method is a sequential regression approach, an early example being [63]. Here the authors identify that more information describing an incident will become available over time, and hence consider a series of models to make sequential predictions. Their sequential information is descriptive

and included information such as damage to the road and the response time of rescue vehicles. An alternative type of sequential information one could consider is the time series provided by a sensor network, for example those recorded along the SRN, and we investigate their use in sequential prediction in chapter 6. Truncated regression approaches are discussed in [47], where there is a specific effort to model 'cascading' incidents (referred to as primary and secondary incidents in some literature). The thought here is that incidents that occur nearby in time and space to another would lead to a significantly longer clearance time for the road segment, and hence this should be accounted for in modelling. Extensive analysis of primary and secondary incidents in our dataset can be found in [2], which constitutes a large part of chapter 5, and as a result of both this and [47], we consider the inclusion of a cascade variable in our modelling in chapter 6, allowing this to influence duration predictions.

Further regression approaches are explored in [64, 65], and switching regression models are used in [66]. Note that in [65], the authors first cluster the incident data, then use this clustering as additional features for a model, further suggesting that there is some element of sub-population structure in the data. It would be interesting to consider how one might apply clustering in a dynamic setting, taking measurements up to some prediction time and identifying sub-groups in the data based on this. A final relevant regression based work is [67], where quantile regression is used to model incident durations. This is a natural choice, as there is a clear skew in the empirically observed duration distributions, and if one does not want to assume a particular distributional form, they can instead model properties of the distribution, in this case quantiles.

Alternative modelling approaches applied to this problem include those based on trees or ensembles of them. Tree based models are discussed in [53], where the authors compare models that assume particular incident duration distributions, a k-nearest-neighbour approach and a classification tree method based on predicting 'short', 'medium' and 'long' incidents. They conclude that no model provides accurate enough results on their dataset to warrant industrial implementation, but find the classification tree is the preferred model of those considered. Classification trees are further considered in [68], where a specific advantage of their method is that only the branches corresponding to available information on a per-incident basis are used for analysis, offering a natural way to deal with missing data. The authors specifically state that they do not use a fixed tree, instead they build a sequence of models, each built using a different subset of features. The predicted duration for and incident is then generated by passing the information it provides down the tree fit to the corresponding subset of features it provides. Doing this is reasonable if there are a small number of feature subsets that each event can possess, however

quickly becomes impractical otherwise.

Further, [69] consider a regression tree approach, where the terminal nodes of each tree are themselves multivariate linear models. Such an approach avoids binning of incidents into pre-defined categories, and achieved 42.70% mean absolute percentage error, better than the compared reference models. From an interdisciplinary setting, alternative tree methods have been considered, namely one known as 'random survival forests' (RSF) [70] as an extension to random forests to a survival analysis setting. In such a framework, the terminal nodes of each tree specify cumulative hazard functions for all data-points that fall into that node, and these hazards are combined across many trees to determine an ensemble hazard. There is no defined distributional assumption in such a model, again leaning towards the side of freedom in allowing the data to construct its own hazard function estimate rather than parametrising an estimated form. Applications of such a method are seen in [71, 72] and a review is given in [73].

Of course, the recent explosion in machine learning research has lead to a number of authors considering its application to incident duration modelling. Some of these include support vector regression, as in [74], and k-nearest-neighbour clustering methods as in [75]. Another rapidly developing branch of machine learning is neural networks, which have been used extensively in incident duration prediction and form the basis for some of our work in chapter 6. Examples of this include [76–78]. Each of these applies feed-forward neural networks to determine estimates of incident durations, and particularly in [78] sequential prediction is considered, using two models. The first takes standard inputs such as time of day and details on location and incident type, whereas the second takes these along with detector readings from upstream and downstream locations after the incident had occurred. These are input into feed-forward neural networks, and used to generate point predictions. Additional neural network applications are given in [79], where their performance is compared to that of linear regressions, decision trees, support vector machines and k-nearest-neighbour methods. The authors find that different models have optimal performance at different incident durations, suggesting there is still much to improve on feed-forward networks. A final point to note is that neural networks have been applied to survival analysis problems in other disciplines, for example healthcare, a number of times. Examples of this include [80] which develops a Cox model, replacing a linear regression component with a neural network output, [54] which as discussed removes any distributional assumptions, and [81] which uses a sliding window mechanism and temporal convolutions for dynamic predictions. We consider if the later two are useful when applied to traffic incidents in chapter 6. Specifically, using [81] offers an automated way to engineer features from data provided by the sensor network,

whilst being able to model a parametric or non-parametric output.

Whilst we have discussed a number of different methodological approaches, the features used by these models, regardless of approach, appear quite consistent across different works. In [64], the authors state that by using number of lanes affected, number of vehicles involved, truck involvement, time of day, police response time and weather condition, one can explain 81% of variation in incident durations. An overview of various feature categories and other important factors is given in [44], identifying incident characteristics, environmental conditions, temporal characteristics, road characteristics, traffic flow measurements, operator reactions and vehicle characteristics as important factors when modelling incident durations.

Our work throughout this thesis focuses on the use of sensor data for a number of problems in intelligent transportation systems, and we note that we are not the first to consider applying this to incident duration analysis. Speed data collected from roads is used in [51], where the authors include two features based on the speed series: if the difference between the 15th and 85th percentiles of the speed data is greater than 7 mph and if the 85th percentile for speed is less than 70 mph. Further, in [78, 82] the authors train two feed-forward neural networks, with input features that include the speed and flow for detectors near the incident. The first model provides a forecast just before the incident occurs, and the second has new data input whenever available, updating predictions as time progresses. In the first paper, the focus is on reducing the dimensionality of the problem through feature selection via a genetic algorithm. Our work in chapter 6 fundamentally differs from these for multiple reasons. With regard to [51], this was an analysis of which factors impact incident durations the most, and whilst we both utilise hazard based models, we use the sensor data to engineer dynamic features, either manually or through temporal convolutions. Additionally, our work differs from [82] and [78] through how we determine features and network structure, and through predicting an output distribution, not just a point estimate. This allows for some determination of risk and a survival function in-line with classic survival models, but retains the non-linear modelling power of a neural network. We additionally use the sensor information to determine when the end of the incident is in-terms of returning to 'normal operating conditions' which are defined entirely by the data, and as discussed is highlighted as a problem in this area to address.

With dynamic prediction being highlighted as an area to address in the literature, some recent papers have looked at this problem through different methods to those already discussed. One example of this is [83], where a topic model is used to interpret written report messages and predictions are made as new textual information arrives. Further, multiple regression models are built

in [84], and as different features become available, data-points are assigned clusters, and a prediction is generated using a regression model tailored to each cluster. The approach in [85] considers four types of incidents, defined by the ratio of measured speed and free-flow speed during the incident. As new speed measurements arrive, they are used to determine if and when the ratio of actual speed to free-flow falls below specified thresholds, and models are built incorporating this information. Lastly [86] consider a five stage approach, where a prediction is made at each stage and different features are available defining these stages. These include vehicles involved, agency response time and number of agencies responding. While this structured approach addresses some aspects of dynamic prediction, the purely data-driven approach which we present in chapter 6 provides much more flexibility.

2.4 Challenges in planning paths for autonomous vehicles

The previously discussed tasks consider network scale problems one might face and the existing work to address them. There are however a huge range of topics relating to intelligent mobility that consider problems on a more microscopic or single vehicle scale. A significant one from both an academic and industrial perspective is path planning for autonomous vehicles. An extensive overview of this area is provided in [87], highlighting methods typically used to generate paths. The four main methods reviewed are based on: determining paths through a graph, randomly sampling state-space, interpolating curves through way-points and numerical optimisation of functions subject to constraints. Many of these algorithms plan a route from some origin to a destination trying to meet some criteria, for example minimal path length, maximal comfort or lowest curvature. Fundamentally, these can incorporate static objects, but often the control scheme to follow this, and accounting for dynamic obstacles, is left as a later problem to solve. Indeed, accounting for dynamic environments is explicitly listed as a challenge in [87]. We now review each of these path generation methods, but note from the outset that the work in chapter 7 falls under the class of numerical optimisation approaches.

Graph based approaches are typically concerned with discretisation of a domain into a network representation, and then applying well know network theory algorithms such as Dijkstra's algorithm or A^* to generate a path through this. An example of such a work is [88] in the context of vehicle manoeuvres, where the network is taken to be a 'dynamically feasible lattice state space'. In short, this is a set of states, each of which consists of a position, heading and speed. Connections are made between states by considering an 'action

space', meaning pairs of acceleration and steering commands that allow one to transition between states. States in the lattice are connected if there is a valid action pair linking them, and then finding a route through the lattice ensures the resulting path is dynamically feasible. In [88], a variant of A^* is used to search through the graph, determining heuristics to decrease computational time. Further network based approaches are given in [89], where the authors again search the kinematic space of a vehicle using an algorithm based on A^*, and smooth the resulting path by optimising a cost function penalising being close to obstacles, excessive curvature and improving the smoothness of the path. A^* search is once again used in [90], specifically to route around a multi-level car-park and again the authors smoothed the result of the graph search to yield a final path. Whilst A^*, an approximation to Dijkstra's globally optimal search through a graph, improves the speed of searching, retaining the global optimality relies on choosing a heuristic that is 'admissible', meaning it never over-estimates the remaining cost of moving from some node to the target. Further, as pointed out in [88], the level of discretisation of the state and action spaces significantly impacts the final result, as well as the computational demand, so there are multiple trade-offs to consider when applying methods based on routing through a network.

Methods that are based on randomly sampling state space include rapidly-exploring random tree (RRT), discussed in [91]. Here, the initial vehicle state is specified as a single vertex in a tree, then a random point in state space is selected. The nearest point currently in the tree is chosen, then the control action that brings this point nearest to the random point in state space is applied, and the resulting state is included in the tree. This is repeated many times and rapidly the state space is explored. Applying the control action is done though the consideration of a system of differential equations representing the vehicle dynamics, and hence links between state space satisfy what is possible for a vehicle. When a path from the initial state to the goal is found, the algorithm terminates. Such a method allows state space to be explored rapidly, and when coupled with a collision detection system, can account for static obstacles. However, in its initial form, one takes the first feasible path as the result, which may not be optimal depending on what measure of optimality is used. One can continue the exploration until a path meeting some other criteria is found, however it is still a challenge to define exactly what these criteria should be if there is the potential for dynamic obstacles in a scene. Finally, the paths generated by this algorithm are known to be jerky (uncomfortable) as they are not continuous.

The third method mentioned, planning paths by considering interpolating curves, takes a set of way-points and then determines a smooth continuous path between them, discussed at length in [92]. This smooth continuous path

may be a polynomial, a spline interpolation or a clothoid to name just a few. Such a method is appealing in a static global sense, if one can predefine some intuitive way-points and generate a smooth path between them, the resulting experience in the vehicle will be comfortable and smooth for passengers. When dynamical obstacles are present, one has to re-compute paths in an on-line setting, and either determine a way to add additional way-points to manoeuvre around new obstacles, or wait until the obstacle has moved to proceed. How best to dynamically add new way-points is a complex problem, and if we were to consider this problem in the context of uncertain behaviour of other vehicles, we would perhaps search for some robust way-point generation based on a compromise between path length and expected interference from obstacles.

Numerical optimisation is used in path planning in numerous ways. The first is as a method to refine other planning results, as discussed in [89, 90] already. Such an application makes logical sense, as the initial path generated from say a graph based method should already be feasible and achieve the overall goal, and then making minor adjustments can further optimise the path. There are also a broad range of well established numerical optimisation routines that handle linear and non-linear constraints on states and actions that may be present in an autonomous driving context. Rather than just refinement however, other works consider constructing a trajectory directly through minimisation of a cost function. One such example of this is [93]. Here, a controller is developed to generate a path minimising a cost function which has components representing distance to walls, velocity difference from some desired value, and induced acceleration and jerk values. The idea is that the optimal path is a compromise between driving at a desired speed, comfort and avoiding walls. Whilst computationally expensive, designing a specific cost function and minimising it is appealing because constraints can be directly incorporated into the optimisation, and different weights can be applied on different components to represent the relative importance of them. In different scenarios, one might value different components of this cost more or less than others, so this can be adjusted as required to reflect this. Our work in chapter 7 is based on constructing a specific cost function and optimising the actions of a vehicle directly to minimise it. However, we consider a different concept of safety than this work, and focus on a conceptual, emergent behaviour model rather than trying to recreate a specific journey, as is done in [93]. It is important to note however than in the cited work, the final result was that a real vehicle drove 103 kilometres on roads and avoided any collisions, however the authors state it is tuned for a 'defensive style of driving' which might not always be preferable.

For numerical optimisation based methods, it is common to use a method known as model predictive control (MPC) [94] to overcome computation time

and account for modelling uncertainties. The idea is that, if we were to plan a 10 minute journey and had the option to apply a steering and acceleration action every 0.1 seconds, the size of the resulting optimisation problem would be large. In this case, it would require optimising $2 \times 60 \times 10 \times 10 = 12000$ variables. Whilst one could actually solve this, computation time would be restrictive, and there is some associated uncertainty with the results. The model used for the vehicle may present minor errors compared to an actual vehicles dynamics that accumulate over time, and the environment may change meaning the route has to be re-optimised. MPC tackles this by breaking up a large optimisation problem into many smaller ones. Given some look-ahead horizon τ^*, one solves an optimisation problem from the current time t up to $t + \tau^*$, then in a receding horizon fashion allows the system to evolve and re-solves an optimisation starting at the new point. This general methodology is a natural way to consider planning a path for an autonomous vehicle, as humans approach the problem in a very similar way: assessing the current situation, reasoning about what other actors might do in the near future, performing some actions, and constantly re-assessing and adjusting their actions based on how the situation develops. In the autonomous driving context, solving many smaller problems not only reduces computational effort, it also allows for continuously updating prediction uncertainty through time.

An example application of MPC in autonomous driving is given in [95], where a steering controller is developed, aiming to choose a set of actions that follow a pre-determined reference path with minimal error. The cost function utilised is based on the deviation from a reference path, and constraints represent physical and comfort limitations. Similarly, in [96], MPC is used to again follow a reference path, but also avoiding obstacles. Firstly, a global path is determined, then when static obstacles are detected, MPC is used to locally re-plan and avoid them. Another relevant work is [97], where MPC is used to both incorporate safety and comfort into path-planning. The work handles safety by introducing a sub-module that considers the risk of violating comfort constraints and the risk of violating a lateral tracking constraint. When this risk is sufficiently high, local re-planning is performed. However, there is no discussion on the concept of other drivers' behaviour or incorporating the variability and growing uncertainty in the future states of other vehicles.

A fundamentally important concept in autonomous driving is safety, and how to actually define this concept mathematically. All autonomous vehicles can be assumed to have some on-board 'watchdog' that will not allow input actions to roll a vehicle or present a physically impossible situation, however safety in terms of avoiding other vehicles and pedestrians in a scene is complex. There are a range of ways this is incorporated into path planning in the literature. In [93], when obstacles are detected, all corridors that the obstacle

can reach in a time horizon are determined, and then a trajectory for each corridor is predicted by assuming the vehicle follows its lane and maintains the same distance to the right bound. Polygons representing hard constraints on positions are created, varying through time along the predicted trajectories, specifying areas of the domain that cannot be occupied by the controlled vehicle. Other works generate a 'predictive occupancy map' of the domain, for example in [98], where feasible locations of other vehicles in a scene assuming fixed behaviour are determined. Twelve trajectories through space are then constructed, each with an associated risk considering the risk of changing lanes and of encountering other vehicles. Selection of the optimal trajectory is made by choosing the one from these twelve with minimal risk. In the context of ocean navigation, [99] utilise a probabilistic prediction of ship locations to consider what areas were most likely to be occupied by other ships. A cost function for planning a path is formulated by incorporating the risk of being in particular locations, hence resulting in a minimum expected risk planner. Our methodology in chapter 7 shares some ideas with these three works, as we will consider some hard constraint through time over the expected position of other vehicles as in [93], but we shall also utilise a predictive distribution of the states of other vehicles to augment the hard constraints. We will not generate a discrete set of trajectories and choose the lowest risk one as in [98], instead we will solve an online optimisation problem where safety is part of the cost function.

Other works assign a 'potential function' to obstacles that rapidly increases as distance to the obstacle decreases to ensure safety. Discussions of this in a dynamic setting are given in [100]. Here, the authors consider the relative velocity between the controlled vehicle and a moving obstacle, and choose some maximal allowed deceleration, which in turn defines a difference in distances one would travel over some time-period. A potential function is then described utilising concepts of a safe distance, the obstacle dimensions and at what ranges an obstacle influences decision making. Such a function guides a path planner to avoid being near obstacles whilst still pulling the trajectory towards some end goal. This is applied in an autonomous driving context in [101], where three types of potentials are considered. The first is for crossable obstacles (bumps in the road), the second is for non-crossable obstacles (other vehicles) and the third is for lane changing. Crossable objects have an exponential potential, non-crossable objects have a hyperbolic potential and lane boundaries have a quadratic potential. A cost function is then determined looking ahead some horizon and accounting for the potential functions, the distance from some target state and the cost of large tyre forces and steering angles. The ultimate goal is to reach some lane at some desired speed. A similar approach is taken in [102], where a 'potential crash severity index' is also incorporated,

accounting for velocity difference, mass difference and angle differences between the controlled vehicle and obstacle. This severity index is then incorporated into the cost function, along with the previously discussed components, and it is shown that doing so ensured if a crash between vehicles did occur, it would have minimal severity. We also consider looking some horizon ahead in our work throughout chapter 7, but rather than choosing a functional form for each obstacles potential, we instead allow this repulsion to be a consequence of the uncertainty of a predicted distribution of future obstacle states. We also do not include a concept of 'cut off' distance or similar, instead this is controlled by the tail behaviour of the predictive distribution. Finally we also include concepts of crash severity as in [102], ensuring the velocity difference between vehicles is accounted for in the concept of safety.

It should be noted that a huge number of papers address single complex manoeuvres. In [103] and [104], this takes the form of designing a methodology to perform an overtake. In-particular, [104] considers a three-phase overtake with only the controlled vehicle and a single other vehicle in the scene, and assumes the other vehicle travels with a constant velocity along a rectilinear route, and the overtake lane has enough space to complete the manoeuvre. Acceleration profiles are generated using sine curves in [105] and pre-specified acceleration and deceleration periods are used to design a system that smoothly overtakes obstacles. Such a construction makes fitting the parameters to data simple, but restricts the behaviour of the vehicle to follow a small set of all possible dynamics. Clearly, designing a large number of systems, each of which specialised to a single task, will make transitioning between simple and complex driving environments very difficult, as the number of individual tasks becomes very large as environmental complexity increases. In [106], a more general framework is considered, deciding when to follow and when to overtake vehicles on a highway. They do so first by computing a separate reference speed for the vehicle to follow in each lane. A rule based system then decides which lane to take, and an optimisation problem is solved to determine the sequence of actions to get there. Of particular relevance to our work in chapter 7 is their consideration of uncertainty, where elliptical boundaries are drawn over expected vehicle positions, with the ellipse shape representing different uncertainties in lateral and longitudinal positions of the vehicles. Assuming Gaussian tracking errors, they construct hard constraints in their optimisation specifying an elliptical region that the controlled vehicle cannot enter for each dynamic obstacle. The semi-major and semi-minor axis of these ellipses are determined by the lateral and longitudinal tracking uncertainty, multiplied by some constant which will in turn enclose some quantile of the associated distribution. Slack variables are then included to adjust the ellipse length under extreme conditions. In section 7.3.2 we discuss a similar idea, however

do not just consider the uncertainty as a hard constraint, rather we use it to determine the cost of occupying space and possessing various velocities throughout time. Our approach also does not generally assume Gaussian tracking errors, instead we take some output from a prediction model which can have any form. A further point to mention is that papers such as [106] have some fundamental rule based system attached, and an explicit understanding of following, overtaking, lane-changes and so forth.

Finally, of particular recent interest are imitation learning approaches to path planning. Examples of this include [107] for high-speed off-road driving and [108, 109] for urban driving. These methods aim to control the vehicle in such a way as to imitate an expert driver, defined by available training data. In some approaches, one assumes that the training data gives samples of a perfect 'policy', and the goal of training the method is to take an observation and determine the policy that maps this to the appropriate action that would have been taken by the expert driver. These policies are typically represented by neural networks, allowing for non-linear relationships between the scene input and driver behaviour. As they have little prior knowledge encoded into the modelling structure, they take large amounts of data to produce a model that can generalise across many scenarios, and there is a large cost associated with collecting this data.

2.5 Mathematical & statistical tools

We now detail some of the tools and methods that are central to work throughout this thesis.

2.5.1 Kernel density estimation

Kernel density estimation (KDE) is a statistical method that aims to take observed data-points and determine a smoothed representation of them. Given some samples from an unknown distribution, $p(\boldsymbol{x})$, with $\boldsymbol{x} \in \mathbb{R}^d$, KDE generates a non-parametric estimate of $p(\boldsymbol{x})$ by "smearing out" the samples using a predefined kernel. Given N samples, $\boldsymbol{X}_i \in \mathbb{R}^d$, $i \in \{1, 2, ..., N\}$, the kernel density estimate is

$$\hat{p}_{\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{1}{N} \sum_{i=1}^{N} k_{\boldsymbol{\Sigma}}(\boldsymbol{x} - \boldsymbol{X}_i) \tag{2.1}$$

where $\boldsymbol{\Sigma}$ is a $d \times d$ positive definite matrix called the bandwidth matrix and the kernel, $k_{\boldsymbol{\Sigma}}(\boldsymbol{x})$, is often taken to be the multivariate normal distribution

$$k_{\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}. \tag{2.2}$$

In denoting the estimated density by $\hat{p}_{\boldsymbol{\Sigma}}(\boldsymbol{x})$, we are suppressing explicit dependence on the samples, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ for the sake of notational compactness. The estimate $\hat{p}_{\boldsymbol{\Sigma}}(\boldsymbol{x})$ is strongly dependent on the choice of bandwidth matrix, $\boldsymbol{\Sigma}$, since it controls the amount of smoothing applied to the data. Choosing $\boldsymbol{\Sigma}$ involves a trade-off between under-smoothing and over-smoothing the data. See [110] for an extensive discussion. One principled way to select the optimal trade-off is to select the $\boldsymbol{\Sigma}$ that minimises the mean integrated square error (MISE), defined as

$$\mathrm{MISE}(\hat{p}_{\boldsymbol{\Sigma}}) = \mathbb{E}\left[\int_{\mathbb{R}^d} \left(\hat{p}_{\boldsymbol{\Sigma}}(\boldsymbol{x}) - p(\boldsymbol{x})\right)^2 d\boldsymbol{x}\right], \tag{2.3}$$

where the expectation value is with respect to the distribution of the samples, $\boldsymbol{X}_i$, $i \in \{1, 2, ..., N\}$. This optimisation cannot be done directly however since $p(\boldsymbol{x})$ is unknown. There are two approaches to address this problem. The first is to adopt a cross-validation approach whereby a subset of the data is used to estimate $\hat{p}_{\boldsymbol{\Sigma}}(\boldsymbol{x})$ and remaining data is then used to estimate the MISE in some way. The second is to exploit the fact that Eq. (2.3) simplifies significantly in the limit $N \to \infty$ when the number of samples become large. In this limit, an analytic formula for the optimal bandwidth can be found although this formula still depends on the (unknown) second derivative of $p(\boldsymbol{x})$. For example, in the univariate case where the bandwidth is a scalar, say ω, the asymptotic mean integrated square error (AMISE) is

$$AMISE = \frac{R(k)}{N\omega} + \frac{\omega^4}{4}\left(m_2(k)\right)^2 R(p'') \tag{2.4}$$

where $R(k) = \int k(x)^2\, dx$, $m_2(k) = \int x^2 k(x)\, dx$ and $R(p'') = \int p''(x)^2\, dx$ with $p''(x)$ denoting the second derivative of the unknown density, $p(x)$ and k denoting the kernel. AMISE clearly has a minimum as a function of ω. The location of this minimum is the optimal bandwidth and can be calculated analytically as

$$\omega_{\mathrm{optimal}} = \left(\frac{R(k)}{\left(m_2(k)\right)^2 R(p'')}\right)^{\frac{1}{5}} N^{-\frac{1}{5}}. \tag{2.5}$$

Details of these calculations can be found in [111, chapter 3]. Methods based on this second approach and extensions of it to the multivariate case are called 'plug-in' methods. Different strategies have been suggested for self-consistent estimation of the unknown second derivatives of $p(\boldsymbol{x})$ and most modern approaches to KDE are based on the 'plug-in' approach, see [112]. In our work in chapter 4, we use the method of Chacón and Duong described in [110, 113] to perform multivariate smoothing. In chapters 5 and 6, we utilise univariate smoothing. In each of these, KDE allows us to avoid specifying parametric forms of particular functions or distributions, and the specific

applications are discussed within each chapter. Further details can be found in appendix A.1.1.

2.5.2 Self-exciting spatial-temporal point-process models

From our review of the literature relating specifically to spatio-temporal incident modelling, we saw a number of spatial and temporal features were clearly present in incident data. Further, there was a specific mention of a cascading effect, meaning incidents occurring at one location and time influence the likelihood of others nearby in space and time. Finally, we saw point-process models had been applied, modelling traffic incidents on networks. A natural modelling extension based on these observations is to consider point-process models that capture this cascading effect through the use of a self-excitation component. Such a model is known as a Hawkes process, and these have been applied to a wide range of real-world problems, with a recent review given in [114]. One application discussed within is earthquake modelling, as in [115–117]. This is a natural application for such models, as there is strong evidence that initial large earthquakes lead to aftershocks, and hence there is a clearly interpretable self-excitation component to the application. Another application discussed is crime forecasting, as in [118, 119], where self-excitation can be seen in physical terms as retaliation crimes, among other things. Alternative applications discussed include epidemic forecasting in [120, 121], and modelling incidents on social-networks in [122]. Very recently, similar models have been applied to modelling the spread of COVID-19 [123], although this is still in its infancy.

In each of these works, there is some sense of a 'background' component that models the typical behaviour, and a 'triggering' component that allows for self-excitation. There is much discussion as to what functional form the components should take, in-particular for the triggering components. Typically, one supposes some reasonable parametric forms, then determines which is most appropriate through inspection of the log-likelihood value or information criteria. Standard choices include some form of exponential, Gaussian or power-law decay of triggering in time and space. However, recent work in [124] shows how one can determine both the background and triggering components in a non-parametric way, through kernel-smoothing of data. In-particular, the authors model crime data using a background comprised of periodic daily, periodic weekly, long-term trend and spatial components, as well as triggering in space and time. However, every component is determined without assuming any functional form, instead the authors show how when basing their methods on work in [125], one can determine which incidents in the data appear to be a result of the background and which appear to be triggered, and then smooth data based on this to reconstruct the desired components. It is on this we base

most of our work in chapter 5, and within the chapter we give an overview of the methodology and how we adapt it for our use.

Clearly, there is significant work on applying self-exciting point-process models to problems in crime, earthquakes and epidemics, but there is little on applying them to traffic incidents. One paper that does look at this is [126], where a self-exciting point-process model is proposed that could theoretically be fit to real data. However, only simulated data is used, generated from the proposed model, to then show how the fitting and evaluation would work. Additionally, a somewhat similar idea is considered in [127], however importantly here the authors consider traffic flow data to be 'incidents' and tried to use self-excitation to model the idea that often traffic flow occurs in clusters. They then apply the methods to model traffic flow in Sydney, however there is no clear conclusion as to if the model is statistically defensible and captured all features of the data, or if alternative traffic forecasting methods were preferable. There is still an enormous amount of work to be done applying this methodology to real traffic incident data, and understanding what components of it are important, and the amount of self-excitation present in traffic data, along with appropriate time and length scales it occurs on. Throughout our work in chapter 5, we are able to offer some discussions of these questions.

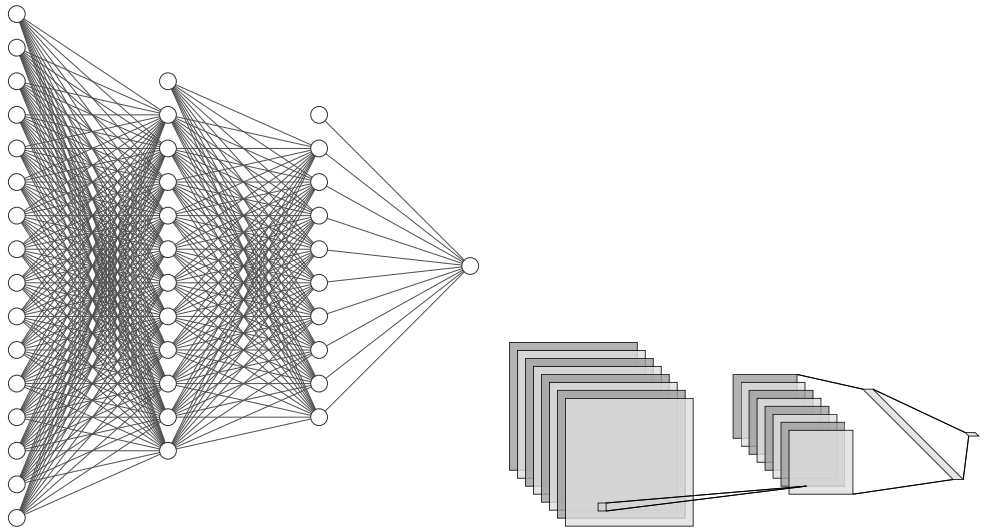
2.5.3 Deep learning

There has been enormous work in the areas of machine learning and deep learning in recent times, with tools from each branch of research being applied successfully to a wide range of applications. In the most general form, machine learning looks to approximate some function $f(\boldsymbol{x})$, which may be a non-linear function of the input variables $\boldsymbol{x}$. We use machine learning tools, in-particular neural networks, throughout chapter 6 and offer an overview of the relevant aspects here.

Neural network architectures

One choice to model the function $f(\boldsymbol{x})$ is through a neural network, the most basic form of which is referred to as a feed-forward network. A simple feed-forward network is pictured in Fig. 2.1a. In such an architecture, we have some number of layers ℓ that make up the network. We denote layer k to have d_k neurons, a weight matrix $\boldsymbol{W}^k$ that performs a mapping $\boldsymbol{W}^k : \mathbb{R}^{d_{k-1}} \to \mathbb{R}^{d_k}$ and some bias vector $\boldsymbol{b}^k$. The output for each layer can then recursively be defined as

$$\begin{aligned} F^1(\boldsymbol{x}) &= \sigma_{\mathrm{NN}}\left(\boldsymbol{W}^1\boldsymbol{x} + \boldsymbol{b}^1\right) \\ F^k(\boldsymbol{x}) &= \sigma_{\mathrm{NN}}\left(\boldsymbol{W}^k F^{k-1}(\boldsymbol{x}) + \boldsymbol{b}^k\right) \ \forall k \in 2, \ldots, \ell. \end{aligned} \tag{2.6}$$



(a) Diagram of a feed-forward neural network.

(b) Diagram of a convolutional neural network.

Figure 2.1: Examples of neural network structures. These are generated using the tool discussed in [128].

where σ_{NN} is some activation function. The ultimate aim is to determine the optimal weights and biases that, given a set of inputs, produces network outputs that match the observed outputs in the data as close as possible. By choosing σ_{NN} as some non-linear function, one can train a model to learn non-linear relationships between inputs and outputs, and there has been significant developments in different architectures to handle different datasets.

One such alternative architecture is known as a convolutional neural network (CNN), depicted in Fig. 2.1b. This figure illustrates the classic setup for such a model, when input data takes the form of an image being a $N \times M \times P$ tensor, with P being the number of channels, for example RGB information of the image. Convolutions between the input and a number of filters, that is small matrices of values to be tuned, are computed and the results are used as novel features of the input data. Tuning of these filters allows for the extraction of useful predictive features, which are often fed through feed-forward layers to determine a final network output. As we have discussed, [81] and other works have looked at applying these networks not to images, but to extract useful features from series data, and shown promising results. In such a case, the channels become different variables, for example speed and flow, and the filters are passed along some temporal window of the data.

Given some specified network architecture, one then trains the network by defining a loss function, and feeding training examples through the network, where both the input and output are observed. Often this is done by feeding a batch of training data at a time, computing the gradient of the loss with respect to each weight, and then updating the weights to improve the loss. Typically,

one separates their data into three sets, a training set with which the network weights are optimised, a hold-out set used to determine if the training should be stopped, and a separate set on which performance metrics are computed and reported. Doing so allows one to stop training the model before it becomes too specialised to the training data (overfit), retaining generalisation of the model. Further, a number of hyper-parameter combinations are considered and trained on, allowing one to determine the number of neurons, filters and so forth that result in an optimal model. We follow this work-flow throughout our work in chapter 6.

Model interpretability

Recently, there has been a significant effort to improve the interpretability of prediction models, both those involving neural networks and more general frameworks. One such example of this is in [129], where 'Local Interpretable Model-agnostic Explanations' (LIME) is proposed, a method to build a simpler 'explainer' model G that locally approximates some complex (in our context neural network) model F. The idea is to optimise the model G to match the outputs of F by generating a set of perturbed samples around a data-point of interest. One then minimises the squared loss between the explainer model and the complex model, and weights data-points in this loss by their distance to the instance we wish to explain. A penalty for model complexity is imposed on this optimisation, ensuring that the explainer model is truly simple enough for human interpretation. The optimisation function for LIME takes the explicit form

$$\sum_{\boldsymbol{z},\boldsymbol{z'}\in\mathcal{Z}} \pi_x(\boldsymbol{z}) \left(F(\boldsymbol{z}) - G(\boldsymbol{z'})\right)^2 + \Omega(G) \tag{2.7}$$

where $\mathcal{Z}$ is the set of samples, of which $\boldsymbol{z}$ are inputs to the full neural network model and $\boldsymbol{z'}$ are the explainer inputs, $\pi_x(\boldsymbol{z})$ is some weighting on the current sample and $\Omega(G)$ is a term that penalises complexity.

Alternative works consider neural network specific approaches to model interpretability, for example [130] in which 'DeepLIFT' is proposed. The idea here is to select some 'reference' or background point we want to compare to and a data instance we wish to explain the prediction of. If we denote the reference output of the network as $F(\boldsymbol{r})$ and the output for the data-point in question as $F(\boldsymbol{x})$, then DeepLIFT determines a contribution score $C_{\Delta x_i,\Delta r}$ for each feature i such that

$$\sum_{i=1}^{M} C_{\Delta x_i,\Delta r} = F(\boldsymbol{x}) - F(\boldsymbol{r}) \tag{2.8}$$

where M represents the number of features. Various forms of $C_{\Delta x_i,\Delta r}$ are

proposed in [130] dependent on network structure, however the key point of this method is that these contribution scores sum to the difference between the reference output and query output, and hence their values are informative of significant features. One computes these contribution scores by back-propagation of the final prediction through the network to the original feature values.

Later, in [131], it was shown that many existing model interpretability methods could be phrased in-terms of a concept from game-theory known as Shapley values. In short, they propose explainer models of the form

$$G(\boldsymbol{z'}) = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i \tag{2.9}$$

where z'_i is a binary value indicating the inclusion or exclusion of a particular feature and M is again the number of features. They then specify three properties that one might desire in a feature attribution method: local accuracy (predictions of the same input give the same output), missingness (no attributed impact of missing features) and consistency (for two models, if the output of one is more sensitive to a particular feature change than the other, then it achieves a higher attribution value). The authors show that under these properties and Eq. (2.9), the ϕ_i values actually coincide with Shapley values from game theory, and are computed as

$$\phi_i = \sum_{S \subseteq \mathcal{M} \setminus i} \frac{|S|!(M - |S| - 1)!}{M!} \left[F(S \cup \{i\}) - F(S)\right] \tag{2.10}$$

where $\mathcal{M}$ is the set of all features. In Eq. (2.10), we sum over all subsets of feature vectors that do not include feature i. For any one of these sets S, we compute the difference between the model output using the features in S and feature i, and the model output using only the features in S, shown by $F(S \cup \{i\}) - F(S)$. The remaining term $\frac{|S|!(M-|S|-1)!}{M!}$ accounts for all possible orderings of the feature vector. The ϕ_i values are then referred to as 'SHAP values' and the term ϕ_0 represents the 'null' model output, that is the output if all features were missing.

It is shown in [131] that various existing methods, such as LIME and DeepLIFT can be phrased to compute these SHAP values, unifying various feature importance measures. However, there is a clear problem in that neural networks cannot take arbitrary missing features as an input, so using Eq. (2.10) alone one cannot compute the SHAP values for such models. Instead, the idea of a background dataset is again used, and rather than set values to missing, they are replaced with values from this background. Typically, this background is not a single data instance, rather it is a large number of them, and the SHAP values are computed averaging across this background. In such a case, ϕ_0 is the

average model output in the background dataset. A further complication is the computational demand in evaluating Eq. (2.10) as it involves summing over all possible subsets of features, so a number of methods to improve computation speed have been proposed and are discussed in [132].

Once computed however, SHAP values are appealing as they are additive. The magnitude of the ϕ_i's reveals the overall importance of a feature, and the values themselves show if they increase or decrease the model output. We utilise SHAP values when considering the impact of specific features in our neural network models in chapter 6. Further discussion of SHAP values is given in appendix A.2.

2.5.4 Optimisation methods

Optimisation is fundamental to a number of problems and applications of methods, and we utilise it throughout our work. In-particular, an online optimisation problem underpins much of our work in chapter 7. We detail how the optimisation procedures actually minimise functions below.

Gradient descent is a simple and well established optimisation routine, with variants of it being used to train many modern neural networks. To minimise a differentiable function $f(\boldsymbol{x})$, the method iteratively steps towards a minima by choosing some starting point $\boldsymbol{x}^{(0)}$ and some step size η, then applying

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \eta \nabla f(\boldsymbol{x}^{(n)}). \tag{2.11}$$

The intuition is that we want to move against the gradient, so we are at some given point $\boldsymbol{x}^{(n)}$ and then choose at each iteration to step in a direction that the function appears to decrease in. When the gradient becomes 0, we have potentially arrived at the minimum of the function.

Newton's method is similar to gradient decent, however replaces Eq. (2.11) with

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \eta \left(\nabla^2 f(\boldsymbol{x}^{(n)}) \right)^{-1} \nabla f(\boldsymbol{x}^{(n)}), \tag{2.12}$$

which clearly adjusts the step taken by considering the second derivative of the function. The size of each step to take can be chosen using backtracking line search, discussed in appendix A.3.

Often, optimisation problems are accompanied by constraints representing practical considerations that limit the values solutions can take. One such set of constraints are equality constraints, which can be written in the form $h^c(\boldsymbol{x}) = 0$ for some constraint function $h^c(\boldsymbol{x})$. A second kind of constraint are inequality constraints, written as $h^c(\boldsymbol{x}) \leq \boldsymbol{0}$. One can extend Newton's method to incorporate equality constraints without significant difficulty. To incorporate inequality constraints, one can use the so called 'barrier' or 'interior

point' method. The intuition behind such methods is that one wants to solve an optimisation problem of the form

$$\min\left(f(\boldsymbol{x})\right) \text{ s.t. } h_i^c(\boldsymbol{x}) \leq 0 \tag{2.13}$$

$\forall\, i \in \{1, \dots M_c\}$, meaning we have a function to minimise and have M_c inequality constraints captured by the functions h_i^c. The barrier method incorporates these inequality constraints into the minimisation directly. If the so called 'log barrier function' is used, one instead solves

$$\min\left(f(\boldsymbol{x}) - \frac{1}{\kappa}\sum_{i=1}^{M_c}\log(-h_i^c(\boldsymbol{x}))\right) \tag{2.14}$$

for some parameter κ. The idea here is that as $h_i^c(x)$ approaches 0, an increasing penalty is applied to the quantity we wish to minimise, thereby still accounting for the constraints. One solves a sequence of problems, increasing the value of κ until some threshold is met, and in turn a solution to the original problem is produced that satisfies the inequality constraints. Each problem in this sequence is solved using Newton's method, initialised at the previous solutions end point. A review of these methods is found in [133], and further discussion of them and optimisation in general is given in appendix A.3.

2.6 Data & modelling uncertainty

Whilst we have reviewed the literature on four problems in intelligent transportation systems and intelligent mobility, and discussed a number of mathematical and statistical tools we shall use throughout this thesis, two further topics of interest remain that have relevance to our work. These relate to missing values in a dataset, and sources of uncertainty in modelling. These topics apply to a wide range of disciplines where real data is used and models of complex situations and processes are constructed. Since we do both of these throughout this thesis, we offer an overview of the two topics here, and discuss their relevance to specific aspects of our work in the conclusions to each chapter.

2.6.1 Missing data

When collecting datasets, there is often potential for certain values to be missing or erroneous. Practical examples of this include people refusing to respond to a particular survey question, faulty sensors and errors in data storage and transmission. Understanding the mechanisms that result in missing data, and the potential impact of this on modelling conclusions is a widely studied area of statistics, and an extensive overview is given in [134]. An important point

to note is that missing values often hide meaningful information. As examples in the context of transportation, these might take the form of missing speed values, flow values or incident flags that would provide a better understanding of the state of infrastructure if they were in the dataset. A result of this is that any modelling conclusions or recommendations to traffic operators we make are subject to this missing information being unknown.

It is important to highlight the difference between missing data patterns and missing data mechanisms. Missing data *patterns* describe which values are missing in a dataset. Examples of this applicable to our application include:

- Univariate or multivariate patterns, depending on how many variables report missing values
- Monotone patterns. If a sensor were to break at some time during the collection period and remain broken for the rest of the collection period, then we would describe the missing data pattern for that sensor as monotone.
- Non-monotone patterns. This describes the case where a sensor may break or temporarily malfunction, stop reporting data, but then be repaired and become functional again whilst we are still collecting data.

Understanding what patterns exist in the dataset is important when considering the potential influence of missing data.

Missing data *mechanisms* describe the relationship between missingness and the values of variables in the data. There are a number of different mechanisms that can lead to missing data, discussed in [134] and [135]. Following these, we consider a complete data matrix $\boldsymbol{Y}$ and denote a missingness indicator matrix $\boldsymbol{M}$, with an entry in $\boldsymbol{M}$ being 1 if the corresponding entry in $\boldsymbol{Y}$ shall be missing. A simplifying assumption is that the rows $\boldsymbol{y}_i$ and $\boldsymbol{m}_i$ are independent and identically distributed over i. We use the conditional distribution $f_{\boldsymbol{M}|\boldsymbol{Y}}\left(\boldsymbol{m}_i|\boldsymbol{y}_i, \boldsymbol{\phi}\right)$ to characterise missingness mechanisms, with $\boldsymbol{\phi}$ denoting unknown parameters. One mechanism is named missing completely at random (MCAR) and describes situations where missingness does not depend on values of the data (observed or missing). Formally, for two values $\boldsymbol{y}_i$ and $\boldsymbol{y}_i^*$ in the sample space of $\boldsymbol{Y}$, this is written as

$$f_{\boldsymbol{M}|\boldsymbol{Y}}\left(\boldsymbol{m}_i|\boldsymbol{y}_i, \boldsymbol{\phi}\right) = f_{\boldsymbol{M}|\boldsymbol{Y}}\left(\boldsymbol{m}_i|\boldsymbol{y}_i^*, \boldsymbol{\phi}\right). \tag{2.15}$$

Alternative mechanisms exist, one being missing at random (MAR). If we denote observed components $\boldsymbol{y}_{i,\text{obs}}$ and missing components $\boldsymbol{y}_{i,\text{miss}}$, then MAR describes situations where missingness only depends on $\boldsymbol{y}_{i,\text{obs}}$. This is written as

$$f_{\boldsymbol{M}|\boldsymbol{Y}}\left(\boldsymbol{m}_i|\boldsymbol{y}_{i,\text{obs}}, \boldsymbol{y}_{i,\text{miss}}, \boldsymbol{\phi}\right) = f_{\boldsymbol{M}|\boldsymbol{Y}}\left(\boldsymbol{m}_i|\boldsymbol{y}_{i,\text{obs}}, \boldsymbol{y}_{i,\text{miss}}^*, \boldsymbol{\phi}\right). \tag{2.16}$$

A further mechanism is missing not at random (MNAR), describing situations where the conditional distribution of $\boldsymbol{m}_i$ depends on the missing values in the data.

In many real-world datasets, values are often not MCAR, and missingness is in some sense informative. A practical example of this in our domain might be that, if a sensor breaks on an extremely busy section of road, it may not be justifiable to perform the necessary repairs as it would cause too significant of a delay to road users. Alternatively, the opposite could be true, that a sensor malfunctions but is on a section of road with so few problems historically that it is not considered a priority to address and remains inactive for a long time. We therefore take care to point out the potential impact of missing data on our results in the conclusions of each chapter. Further, whilst these two examples relate to missing sensor measurements, we may also have prior knowledge that informs us of missing data mechanisms for incident flags in traffic data. It is well known, and already discussed, that incident labels in data are often incomplete, and hence we expect some level of missing data with regards to these in our dataset. If these flags are manually input into the system by operators, as some are in our dataset with further discussion in chapter 3, then it is likely that periods of time where operators are particularly busy will contain more missing flags, and hence the values are not MCAR.

When missing values are encountered in a dataset, one can attempt to impute them. The method of imputation may vary by application, however in the context of our dataset, if particular sensor readings are missing, one might approach this in two ways. Firstly, if a small number of measurements are missing, but the surrounding values are present, one might use these surrounding values to construct an estimate of the missing values. We do this and discuss further in chapter 3, using linear interpolation. Alternative approaches include mean or median imputation using recent values, or constructing a distribution that we believe reflects the data values at a given point, and sampling values from it. However, if an entire day of sensor data is missing, and behaviour varies significantly over a single day, it is no longer reasonable to linearly interpolate such elongated periods of missing values. Instead, one might construct a seasonal estimate of the data values from the non-missing data and use this to impute values, however one must first be sure that there were no particular reasons or mechanisms for this data being missing that make a seasonal model unrepresentative of the true behaviour. Further, as we focus on incidents, that is highly atypical behaviour of traffic on a short time-scale (relative to the collection period) we are apprehensive to perform significant seasonal imputation, as the traffic state during incident periods will not be well modelled by the long-term seasonal behaviour a section of road experiences.

Imputation of missing incident flags is a more complex issue, as with sensor

data, we know some value must have existed at each time point, it may just have not been reported for various reasons. If we encounter a period in our data that appears extremely atypical but has no incident flag, we do not know what the operator was thinking at that time and if they were simply too busy to record the incident, purposefully choose not to raise a flag for some reason, or if despite the situation appearing unusual, there was no true cause of this behaviour they could identify. With the points discussed throughout this section in mind, we comment on missing data, its prevalence in our dataset and subsets of it in chapter 3 and appendix B.1.

2.6.2 Sources of uncertainty

We have discussed that missing data might make some useful information unavailable for analysis. However, even if all values are present in a dataset, there are often still a number of sources of uncertainty and imprecision when modelling complex problems. This is discussed at length in [136], and we outline implications for our work in this section. When attempting to perform analysis and develop methodologies to tack real-world problems, often three distinct steps are taken:

- Modelling: construct a mathematical or statistical model of a system or problem.
- Exploration: consider the behaviour of the model with various different inputs and in different scenarios.
- Interpretation: provide guidance or advise for an expert or decision maker based on exploration of the model.

An outline of the different uncertainties one faces in each of these steps is given in [136]. If we consider the modelling step, one might face uncertainties through:

- What might happen? For example, what situations might develop on a road network?
- Meaning & ambiguity. For example, if we want to improve traffic flow on a network, how do we measure the optimality of flow? It is simply the highest flow rate, or do we want to consider other metrics such as safety? If so, how do we define safety?
- Related decisions. For example, if a traffic operator temporary reduces the speed limit on a section of road, how many people will obey it? How long will it take for drivers to notice and act upon it?

If we then consider the exploration of models, relevant uncertainties include:

- Randomness & lack of knowledge. For example, how likely is it that a traffic incident disrupts flow for a given duration?

- Judgements. For example, what weight should be given to different features a model uses? What threshold do we set to distinguish between congested and uncongested traffic?

- Accuracy of calculations. For example, if we have to perform billions of computations to fit a model, how confident are we of the numerical accuracy during each of these calculations?

Finally, sources of uncertainty during the interpretation of models include:

- Appropriateness of model. For example, should a linear or non-linear model be used?

- Appropriateness of analysis. For example, should a Bayesian or frequentist approach be used be used to explore data?

- Depth of analysis. For example, how complex does a model need to be before we are happy it describes a situation sufficiently well?

Clearly, there are a range of different sources of uncertainty that are relevant to our work, and in addition we highlight potential uncertainty as a result of missing values in the data, or quantities that are not recorded in the NTIS system but likely influence the traffic state. Suppose we try to model the duration of traffic incidents, as we do in chapter 6. There may be factors that our dataset does not provide, but are highly informative of an incidents duration, for example detailed injury records. We therefore have an associated uncertainty as we have an aspect of randomness we cannot directly account for, and it is important to recognise when such problems arise throughout our work.

In [137], the authors discuss the importance of understanding and communicating sources of uncertainty to decision makers when models are built to aid humans in tasks. This is particularly relevant to the work in this thesis, as we design both tools that could be used for decision support in real-time, and perform retrospective analysis that might inform transportation planners of how to improve the network in the future. As a result, we comment on the sources of uncertainty for each aspect of our work in the conclusions for each chapter, and discuss their impact on the potential utilisation of our work in a transport management centre. There is no single optimal way to communicate uncertainties to decision makers. Instead, we consider the most appropriate form within each chapter. Example of this include using a continuous measure of severity of an incident in chapter 4, avoiding the use of an uncertain 'cut-off'

and instead allowing our system to act as a filter that an expert could use to prioritise attention. Further examples include utilising distributional outputs from models rather than point predictions in chapter 6, which inherently represents when a model is very unsure in a particular situation.

A further related aspect of note is that when complex models are proposed as a tool to support decisions by experts, one has to find useful and informative ways to communicate both the uncertainties these models are subject to, and ideally an explanation of why a model gives a particular result. Doing the later helps build trust between an expert and a new tool, and we have discussed one way we will do this in section 2.5.3. A further way to do this can be through the construction of the model directly, as we will see in chapter 4 where the model is built upon common engineering knowledge, and chapter 5 where the model explored can be decomposed into distinct, interpretable components.

Chapter 3

Data Pre-Processing

In this chapter, we outline the data collection, cleaning and pre-processing preformed during this project. A more technical outline is provided in appendix B.1, if readers wish to replicate the pipeline developed.

3.1 Data collection

NTIS provides both historic and real time data for various traffic measurements on roads throughout the SRN through the use of physical sensors (loops) positioned along the roads that detect passing vehicles. These loop sensors are part of a system called 'Motorway Incident Detection and Automatic Signalling' (MIDAS) and function based on electric induction, generating a current when the metal bodies of vehicles travel over them to record passing traffic. We choose to focus our analysis on one the countries busiest motorways, the M25 London Orbital, pictured in Fig. 3.1. Inside NTIS, roads are represented by a

Figure 3.1: Map view of the M25, marked in blue, which all data was taken from. The entire motorway circles London, with a small gap on the east side being the Dartford crossing where we do not collect data.

directed graph, with edges (links) being segments of road that have a constant

number of lanes and no slip roads joining or leaving. Each link in the network will have a number of loop sensors positioned along it We extract all links and loops that lie on the M25 in the clockwise direction, yielding a subset of 77 links and 447 loops to collect data for. We collect data at both the individual loop and aggregated link scales, describing when each source is needed and why when appropriate.

Starting in 2004, 'Smart Motorways' [138] have been introduced in some areas of the UK (one being the M25), hoping to provide traffic operators tools to better manage traffic on the network. In practice, operators use a series of cameras to investigate sections of the network, and can display temporary messages and speed limits to drivers using overhead signs. The hope is to reduce the likelihood of incidents and improve traffic flow by taking a proactive approach to managing situations. The state of this smart-motorway infrastructure is also provided in NTIS, allowing us to collect data describing the state of all message and speed limit signs on our selected links, as well as the opening and closing of lanes. Further details on the control aspects of smart motorways can be found in [139].

3.2 Data selection

Given a set of links and loops, we collect data for all of these sites between April 7th 2017 and November 1st 2018. Extracting this much data was a computational challenge because of the amount of memory, storage and computation time required. After doing so however, we attain three sets of data detailing the network during the collection period: time series at the link and loop-level and incident information and message sign information, both at the link-level.

3.2.1 Time series processing

The following loop level time series are collected from NTIS:

- Flow: The number of vehicles per hour that pass a given loop.
- Occupancy: The percentage of time a loop is occupied in the time period between updates.
- Speed: The average speed in kilometres per hour of the vehicles passing a given loop. This is computed using the average of measured speeds when vehicles are going over the loop.
- Headway: The average time in seconds between vehicles, measured from the front of one vehicle to the front of the following vehicle.

Each of these series are also aggregated internally in NTIS to the link level and are collected for analysis. Two further series are provided at the link level: travel times and profile times. The travel time is the average time in seconds to travel along a selected link, measured through a combination of sensors and tracked vehicles. The profile time is a prediction of the travel time for each link at each minute of the day, using Highways England's prediction algorithm. Details of this algorithm are not publicly known.

Specifically for the loop-level data, we note that these time series are originally provided separately for each lane. They are further accompanied by potential error flags, specifying when the system considers a measurement to be the result of 'suspect equipment'. We treat any data-points with this error flag as missing values. We do not believe there is any reason to suspect missing values will be more common in particular lanes. We also combine each lane measurement into a single, aggregate measure for each loop site for any given minute in the collection window. To do so, the reported occupancy, speed and headway is averaged over all lane measurements. If this involves averaging a missing value, the result is set to missing. Flow is reported per lane but also segmented into four categories dependent on the length of vehicles. We are agnostic to vehicle length in our work, so first sum the results for each length category, and then further sum the flow over all lanes. Missing values in this sum again give an overall missing value.

After extraction, we ensure all time series have an entry at each minute of the collection period. Missing values are linearly interpolated if at-least 3 values from a surrounding 5 minute window are not missing, otherwise they are marked as missing. The time-frames and link or loop sites used for each piece of work are detailed within each chapter.

We accept that one could perform some form of imputation to fill in missing values that do not meet the criteria set for linear interpolation. One candidate methodology to do so would be to construct a seasonal profile of the variables and replace elongated stretches of missing values with this profile. However we choose not to do this for a number of reasons. Firstly, we have a significant amount of non-missing data available to use that allowed us to construct and verify methodologies without any imputation other than the simple linear interpolation for intermittent missing values. Secondly, we explore models and methodologies relating to significant events on the network, that perturb the traffic from its typical state throughout chapters 4, 5 and 6. As a result, we are interested in the short-lived but extremely atypical behaviour these rare but significant incidents result in, and we feel these periods would not be well modelled by any form of profile we could use to impute the missing values. To take our methods into practical implementation, one could consider the most appropriate form of imputation to use in these extreme periods, and how this

impacts the models and methodologies we develop in this thesis.

We finally estimate a density at the loop and link level by dividing the flow series by the speed, yielding a quantity dimensionally equivalent to density (vehicles per kilometre). We note from the outset that the variables recorded by NTIS are temporally averaged, which raises a point of particular note in regards to speed. The average speed value computed by taking the arithmetic mean of individual speed recordings in some given time-frame produces a result known as 'time-mean speed' (TMS). The alternative measurement one could take is known as 'space-mean speed' (SMS), which under stationary conditions is the harmonic mean of the individual measurements [140]. Whilst TMS describes the average speed in some time-window at a given point in space, SMS describes the average speed over some space at a given instant in time [141]. These quantities are further discussed in [142] and [143]. TMS is biased above SMS, and hence density estimates are lower than expected when using TMS in-place of SMS for macroscopic models. The difference between these two quantities is of practical relevance when considering the fundamental diagrams of traffic flow [144], and macroscopic models based on fluid dynamic principles.

Denoting SMS $= \bar{v}_{\mathrm{SM}}$ and TMS $= \bar{v}_{\mathrm{TM}}$, it was shown by Wardrop in [142] that, assuming constant road topology, one can write

$$\bar{v}_{\mathrm{TM}} = \bar{v}_{\mathrm{SM}} + \frac{\sigma^2_{\mathrm{SM}}}{\bar{v}_{\mathrm{SM}}} \tag{3.1}$$

where σ_{SM} is the standard deviation of the instantaneous vehicle speeds. However, this quantity is not known for many datasets. As we do not collect measurements for individual vehicle speeds, we cannot compute σ_{SM}. It is therefore non-trivial to convert between SMS and TMS with our dataset, however work in [141] considers a heuristic way to convert between the two. Specifically, the authors take NTIS data and collect accompanying individual vehicle data, and derive an approximation for σ_{SM} based on this. We take the heuristic approximation used in this work and also generate series of SMS and density using these values. Further details of the approximation in [141] are given in appendix B.2. Note that throughout this work, if we discuss speed or density, we are using the temporally averaged values provided by NTIS. We will specifically point out in any section when space-mean values are utilised.

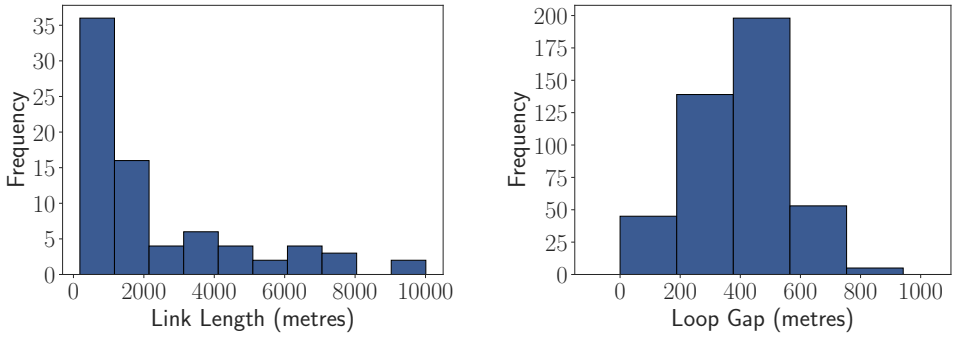
3.2.2 Incident flag processing

We annotate the time series data with incident flags provided by NTIS. These flags are manually entered into the system by traffic operators, specifying any one of the following incident types:

- Accidents - Collisions between vehicles or the surrounding environment.
- Vehicle & General Obstructions - Vehicles, debris or spillages blocking parts of a link.
- Poor Environmental Conditions - Weather related incidents on links.
- Road Management - Operators opening or closing lanes on a link.
- Maintenance - Work carried out to maintain the road or equipment.
- Abnormal Traffic or 'Deviation from profile' - Periods when the travel time on a link is higher than expected. We are told that 'all but very short links' have such an incident declared when the measured travel time exceeds the profile time by at-least 120 seconds, for 5 consecutive minutes.

In addition to type, incidents have an accompanying start and end time, and are specified to have occurred on a particular link in the network. Further incident features are discussed in chapter 6 when they are utilised for modelling. Poor environmental conditions are so common that they are considered redundant for analysis.

In chapter 5, we require more fine-grained locations for incidents than just the link it occurred on, so we perform further localisation. The issue with link-level localisation data for incidents is that NTIS links vary vastly in size, with the distribution of link lengths given in Fig. 3.2a, along with the distribution of gaps between successive loop sensors across the entire network in Fig. 3.2b. The wide range of link lengths shown in Fig. 3.2a clearly indicate



(a) Distribution of link lengths through the M25.

(b) Distribution of gaps between loop sensors on the M25.

Figure 3.2: Comparison of link lengths and distance between successive loop sensors. We see that there is wide range of link lengths, with a small number being close to 10 kilometres in length. On the other hand, distances between successive loop sensors are highly concentrated between 0 and 800 metres.

that further localisation needs to be performed to narrow down a specific point

on the link before one can apply point-process methodologies considered in chapter 5. However, there is hope that the loop sensors are fine-grained enough in space to provide an estimate of this, with Fig. 3.2b showing that any point in the network should have a loop sensor less than around 400 metres away. This is a far higher level of detail than initially provided by NTIS incidents, so we decide to refine our data further by using the time series provided by the loop sensors. This is not a trivial task however as we do not have labelled data on which we can train a model, and instead must define sensible criteria that indicate the location of an incident.

3.2.3 Incident localisation methodology

We note from the outset that detection of traffic incidents in both space and time is an active research field, and various approaches are being considered using an increasing variety of data sources. We are in an atypical situation in that we know a temporal and spatial window in which an incident occurred, and are only searching for a more fine-grained spatial location within this window. As such, we do not aim to completely solve the incident detection problem using inductive loop data, rather we try and develop an effective methodology to take a given window with a known incident in and argue what pair of sensors the incident may lie between. From informal discussions with industry experts, we consider the following properties to be clear signatures of a significant traffic incident:

a) upstream of an incident location, speed will be decreased and occupancy will be increased compared to seasonal values

b) downstream of an incident location, speed may still be decreased, but less so than upstream of the accident

c) downstream of an incident location, occupancy will be decreased compared to seasonal values

Given the industry expert criteria, we first develop simple seasonal models for each loop sensor in our network. There is clear seasonality on the weekly scale in traffic data, with commuting days in the UK being Monday to Friday, and Saturday and Sunday having fewer vehicles on the road. Taking this as the leading seasonal component, we construct a simple seasonal model by taking all data on a given weekday at a given time of day, and then computing the median value, using this as a reasonable seasonal estimate. Doing so, we have a model for each weekday, at the minute level, fit to each loop site separately. We produce one such seasonal model for each variable the loops record.

We then consider what is reasonable to develop with our available data. Ideally, we would design and validate some localisation methodology incorpor-

ating spatial-temporal information from the loop data, inferring a location from the behaviour of all loops. However, we have no data with the ground truth locations, so we cannot reasonably develop such a model. Instead, we can use a simple 'rule of thumb' approach based on existing methodology. Numerous historic methodologies in traffic theory [13, 16] compare adjacent sensors to determine two points, one where the data appears to show an incident, and another where the data does not. We can do the same, and given we know there must be an incident in the given window, we can simply ask at what point do we see the largest discrepancy for two adjacent sensors. First, we consider any two sensors in our network, $i-1$ and i $\forall\, i \in \{2, \cdots, N\}$. We construct residual series for speed and occupancy by subtracting the seasonal profiles attained from the historical median models, and denote the residual series for speed and occupancy on sensor i as RS_i and RO_i respectively. We then compute the spatially differenced values as

$$\Delta RS_i = RS_i - RS_{i-1}, \quad \Delta RO_i = RO_i - RO_{i-1}. \tag{3.2}$$

We then define an 'incident impact score' between the two loops as

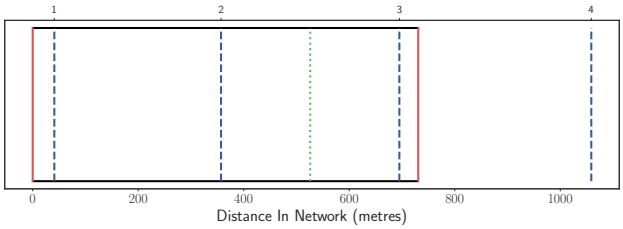
$$IIS_i = \Delta RS_i - \Delta RO_i. \tag{3.3}$$

Since ΔRS_i will likely be positive and ΔRO_i will likely be negative if an incident occurs between $i-1$ and i, then we look for the pair of loops at which the value of IIS_i is largest considering an average over the incident duration, and place our localised incident halfway between these two loops. There is of course huge scope to improve upon such a model, but without data to validate more advanced approaches it is sufficient to provide a method that agrees with existing literature and common sense rules.

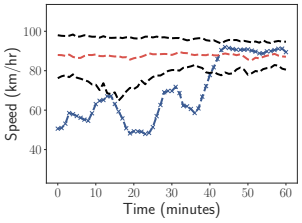
In Fig. 3.3 we plot an example of localising an incident based on our simple methodology applied to loop sensor data.

3.2.4 Message & speed limit sign processing

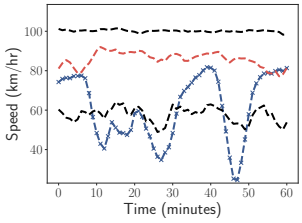
The final components of the data we attain from NTIS are the text written on variable message signs and the speed limits displayed overhead whenever they are changed during the collection period. Speed limits can be set to any of 40, 50, 60 or 70 miles per hour (64.37, 80.47, 96.56, 112.65 kilometres per hour), and currently if a speed limit is active at a particular location, all lanes at that location will have the same speed limit enforced. Variable messages on the other hand vary vastly, from the lowest importance message given as 'do not drive when tired' to the most important 'danger oncoming traffic'. In total, we record the state of 997 signs displayed to drivers.



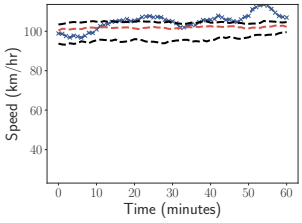
(a) Simple Diagram of the link with sensors marked (- - -), the localised incident location (····), and the start and end of the link (——).



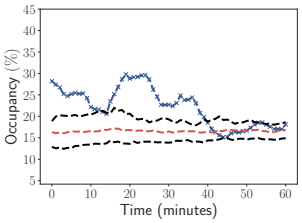
(b) Loop 1 - speed.



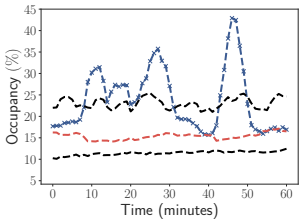
(c) Loop 2 - speed.



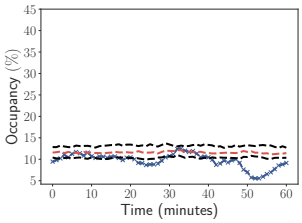
(d) Loop 3 - speed.



(e) Loop 1 - occupancy.



(f) Loop 2 - occupancy.



(g) Loop 3 - occupancy.

Figure 3.3: An example result of our localisation procedure. We plot the data (- - -), seasonal median (- - -) and the 20th and 80th percentiles for the particular traffic variable (- - -). These give a sense of how much variation there is in the data. This is the first link in our network, so we show the next loop sensor along for a sense of scale. We see that sensors 1 and 2 have large drops in speed and increases in occupancy, however sensor 3 appears reasonably seasonal. Our methodology has then placed the incident between sensors 2 and 3.

3.3 Data summary & further details

Having preformed the extraction and processing described throughout this chapter, we attain a dataset that will be used extensively for exploratory analysis, calibration and validation of models throughout chapters 4, 5 and 6. There are a number of additional aspects we discuss here that offer a more detailed description of the data and choices made when collecting it.

3.3.1 Why the M25?

In principle, the data extraction and processing described throughout this chapter could be performed on any part or parts of the SRN covered by NTIS. There are a number of reasons we choose to focus on the M25. The first is that it is a significant part of the UK's infrastructure, both in terms of the number

of vehicles that use it daily and its location, circling England's capital and carrying significant amounts of both freight and commuter traffic. Secondly, during the period in which we collect and process data, there are no elongated periods of time in which special events should make the observed data atypical. As an example, had we collected data from the same location but during the summer of 2012, the presence of the Olympic games in London would likely make this subset unrepresentative of the long-term behaviour, and hence conclusions made on such subsets would be specific to these conditions. No such events are present during our collection period. However, specific days in the data on which rare but recurrent events take place may still look atypical relative to the wider dataset. Practical examples of this include major sporting events held at Wembley stadium, which typically constitute the final games of football tournaments. On these days, traffic volumes on the M25 near these locations of interest may seem higher than on a typical week, however such events occur every year, just infrequently. More recent considerations relate to the highly atypical situation caused by national travel restrictions in England due to the COVID-19 pandemic. During these elongated periods where the majority of the population did not commute to their offices, the traffic volumes are likely to be far lower than in the preceding years. Finally, the M25 offers a large spatial domain to study and attain understanding about the spatial heterogeneity in traffic data, which is discussed at length in chapter 5.

As it is used by both commuters and freight, traffic data from the M25 typically exhibits high volumes of traffic during two 'rush periods', one relating to commuters travelling from their homes to work and the other relating to them doing the opposite. Late in the evening, one might expect the average speed on the M25 to hits its peak, however this is not necessarily true. The high prevalence of heavy goods vehicles during this time, which are generally limited in speed to around 60 mph (96.56 km/hr) can actually result in a slightly lower average speed that one would intuitively expect. Another practical factor to consider is that during the weekends, the lack of commuter traffic often results in a reduction in the number of vehicles on the road. Finally, there are a number of national holiday periods (bank holidays) in the UK, where again the lack of commuter traffic will influence the data recorded on the roads.

As a result of the discussed factors, improvements that are applicable to the M25 have the potential to benefit both a large number of businesses and individual commuters. We also note however that the methodologies and ideas we apply throughout chapters 4, 5 and 6 can be applied to other sections of the SRN in the future to gain similar insights and tools that can aid management across the entire road network. The data collection period, both in space and time, offers a dataset that is representative of the expected behaviour across the M25 during typical times of operation, and offers the opportunity to study

temporal variation over different time-scales, and spatial variation over a major component of the UK transportation infrastructure. In terms of future work, one can further analyse the data we have to collected if consideration of smart motorway features is desired, as the M25 has such features on some sections.

3.3.2 The data generating process

Whilst loop level data is recorded by physical sensors placed along the road, the link level data is generated through a data fusion process before being input into NTIS. This process constitutes generating a link-level measurement using both readings from the loop sensors on the link and floating vehicle data. This floating vehicle data is purchased from a third party, INRIX, and as a result no details are given on the penetration levels of vehicles being tracked. Further, the exact methodology for fusing it with the loop sensor readings is not detailed publicly. It is therefore difficult to make comments on the influence this floating vehicle data will have on the final link-level values. We are however informed that the tracked vehicles constitute a fleet of vans used by various companies that operate on a national scale throughout the UK. One would expect that, if no tracked vehicles are present on a stretch of road during a particular time-period, the loop data alone would be enough to generate a sensible link-level estimate. It may be that very late in the night, this is exactly how the data is generated, and during the day when more vehicles used the roads, the tracked vehicles have a greater influence on measurements. Investigation of the optimal methodology to fuse heterogeneous data sources, and the range of potential data sources one could use is an interesting avenue for future research, however in this context would require a significant amount of cooperation from multiple industrial partners.

Further, there are specific aspects relating to the generation of incident labels in the system that are relevant to the work throughout this thesis. Recall that incident types in NTIS include physical instances such as accidents and obstructions, as-well as abnormal traffic incidents where the measured travel times rise above some link-specific threshold for a given duration. Human operators are able to log information about incidents into the system, attained through the use of cameras covering the network and phone calls from local authorities. In the control room, an operator sees two things displayed at any given time. The first is a schematic of the network, which can be manipulated to view specific locations and the current traffic state at those locations. The second is a list of tasks to resolve. Abnormal traffic incidents are automatically recognised by NTIS but not automatically logged into the system. Instead, when the conditions to raise a abnormal traffic incident are met, the system adds an entry to the operators list of tasks, detailing where the incident is

taking place. The incident is then only recorded in the NTIS feed if the operator manually selects it from their list and accepts the incident. If the travel times on a link return to a level such that an abnormal traffic incident should no longer be active, the task is removed from the operators list. As a result, there may be a number of periods in the data where the criteria to declare an abnormal traffic incident are met, but no record of this is given in NTIS. The number of periods in which this happens depends heavily on how busy an operator is, and how common the conditions resulting in abnormal traffic incidents are, which itself depends on the variability in the data and the thresholds used to define the incidents. This discussion clearly indicates there are fundamental reasons in the data generation process that will lead to potential missing event flags in our data. It is important to consider this when evaluating our models performance, and we discuss it further when reflecting on the work in chapter.

3.3.3 Summary statistics & exploratory analysis of the data

Within chapters 4, 5 and 6, we will use varying subsets of data either for illustrative purposes, or to separate training and testing of models. Here we detail some finer points and summary statistics of the overall dataset. The first point of note is that although we collect data between April 7th 2017 and November 1st 2018, we do not have data-records for every single day in this window. Rather, data collection was first performed between April 7th 2017 and June 20th 2017. After this, more processing code was written to record loop-level data, and once complete, more data was attained for analysis. Loop data is therefore accessible to us between September 1st 2017 and November 1st 2018. We could not attain the raw data files before this date to further extract the loop data, as due to their size Highways England removes them from storage after a set time. Highways England were contacted and asked if there was any potential to get this data but revealed that it no longer existed in their system to provide.

Summary statistics

The summary statistics, computed for each link using all available data, are given in table 3.1. It is important to also note that 'Num Events' in table 3.1 includes all accident, obstruction and abnormal traffic events. A large number of abnormal traffic events exist in NTIS, and industry experts tell us these are the most common form of incident, however they vary enormously in severity and duration. Whilst table 3.1 contains a significant amount of information due to the studied domain consisting of a large number of links, we highlight some aspects here. The first is that generally, longer links have more loop sensors on

than shorter links. This makes intuitive sense and shows loop sensors provide a higher level of detail spatially than link averaged values. We do have a single link, numbered 17 in table 3.1, that is 0.5 kilometres in length, and has no loop sensors on it but still reports data at the link-level. Data values from this link will be a result of the fused data provided by the NTIS system.

Inspecting speed values in table 3.1, we see that median speeds for links vary between 88 and 110 kilometres per hour, showing some spatial variation. Many links report a minimum speed value of 0 kilometres per hour, however note the values displayed in table 3.1 are rounded to the nearest integer, so in practice this shows many links experience a minimum speed less than 0.5 kilometres per hour. From this, we see most links experience nearly stationary traffic at some point during the collection window. Maximum speeds by link vary between 116 and 130 kilometres per hour. These are larger than the maximum allowed speed of 112.654 kilometres per hour (70 miles per hour) on UK motorways, however during very low density periods, we could have individual drivers violate the speed limit and hence give rare large readings. Finally, the amount of variation observed in speeds on each link is shown with the inter-quartile range. We see links with the lowest observed variation have an inter-quartile range of 6 kilometres per hour, where as the highest observed reaches 51 kilometres per hour. This is a significant difference and highlights that different spatial locations can experience quite significantly different speed behaviour.

Performing the same analysis for flow, we see that median values across all links vary between 1680 and 5460 vehicles per hour. This discrepancy indicates certain spatial locations routinely carry significantly more traffic than others, which is expected as some links will be adjacent to particular locations that many people will need to access. Minimum flow values of 0 vehicles per hour are often observed. Maximum values on the other hand reach between 4298 and 10200 vehicles per hour, again showing significant variation of traffic metrics across the M25. The inter-quartile range for flow varies between 1635 and 4980 vehicles per hour across links.

Comparing travel times across links is less informative as we have vastly different link lengths, and hence expect significantly different travel times. One can however see that the maximum travel times are between 6 and 34 times higher than median values observed on links, which poses significant problems for individual vehicles on the links and disrupts both commuters and freight transport. In exceptionally rare cases where measure speed is exactly 0 kilometres per hour, one might expect infinite travel times to be reported, however in practice these values are always finite. This is likely a consequence of the reporting architecture and data fusion previously discussed.

Moving to inspect density in table 3.1, we observe median values between

20 and 64 vehicles per kilometre, further highlighting the spatial variation in observed traffic metrics on the M25. Minimum values of 0 vehicles per kilometre are often observed however again note that these are rounded, indicating many links yield speed and flow values such that when flow is divided by speed, we see less than 0.5 vehicles per kilometre on many links. Maximum density values vary significantly across links, attaining values between 117 vehicles per kilometre and 390 vehicles per kilometre. The inter-quartile ranges varying between 18 vehicles per kilometre and 89 vehicles per kilometre show, as with all other metrics, that our dataset possesses significant variation in traffic metrics across the studied domain. Similar summary statistics of different subsets of data used throughout this thesis are given in appendix B.4.

Missing data

Inspecting table 3.1, we see that our dataset does contain missing data, and therefore comment on the patterns observed and potential mechanisms. From table 3.1, we see that only a single link, numbered 65, reports missing data for the entire collection period across all variables. Is is most likely that this is a result of broken equipment that has not been repaired in the collection window. There is a clear pattern when inspecting the prevalence of missing values for different variables, with flow being missing more often than speed and travel time values. Specifically, links 60, 66, 68 and 69 report only missing flow values but have a significant number of non-missing speed values. It is unclear why flow would be more likely to be missing than other variables, however this consistent pattern observed in our dataset may point to a specific aspect of the sensors that allow them to report one variable but not another, or an aspect of the data-fusion that is able to determine speed values but not flow values. The fact that these links with entirely missing flow values are nearby in space suggests that this data is not MCAR, and instead it may be that addressing the problem, if it does involve changes to the physical sensors at the locations, would be too disruptive to traffic to justify.

A further pattern of missingness emerges when we inspect links 12 and 13 in table 3.1. These report significant amounts of missing speed, flow and travel time values, and are adjacent in space to each other, again suggesting an aspect of informative missingness. They do however report speed values for over half of the collection period. An obvious question may be, did these missing values occur continuously over some interval, for example a broken sensor that was fixed at some time, or did they occur randomly through the collection window? In appendix B.4, we provide analysis and summary statistics for subsets of data used throughout the thesis, and our conclusions there suggest that almost all of these missing values were reported in 2017. Clearly, this missingness is

therefore dependent on the collection time and may indicate a malfunctioning set of sensors that were fixed or an aspect of the data fusion that was altered in early 2018.

Taking a less granular view, we see the majority of links have less than 0.5% of speed values missing. Three links have over 50% of their speed values missing, the first of which is within the first half of the domain (link 31 in table 3.1) and the remaining two are towards the end of the M25 (links 63 and 64 in table 3.1). It is interesting to see that the links surrounding link 31 do not report any significant amount of missing speed values, suggesting the mechanism responsible for these missing values was solely localised to the individual link. Our analysis of subsets of data in appendix B.4 suggests that a significant amount of non-missing values on links 63 and 64 were reported from June 2018 onward, again suggesting structure in the missingness and revealing information about the underlying generation process.

As discussed, flow generally has more missing values than speed for each link, with 5 links reporting only missing flow values for the entire collection period. In addition to these, 5 other links have more than 50% missing data reported, however the majority of links report 1% or fewer missing values. We most commonly observe that 7%-8% of travel time values are missing across links, which suggests a systematic issue with NTIS. Only one link reports missing travel time values for the entire collection period, and a further three report above 50% of their values as missing. Finally, since density is constructed from speed and flow, missing values in either of these quantities at a given time lead to missing density values at that time.

Recall that it is very difficult to reason about how many missing incident labels the links have, so uncertainties in these are present in all of our projects, and indeed most datasets in the literature. Given these observations, we consider two aspects that might impact our work, what informative missingness is present and what bias might our conclusions be subject to as a result of it. Elongated periods of missingness are clearly informative of a systematic problem in the collection or fusion of the data. Further, the patterns observed comparing the prevalence of missing flow values compared to speed values indicate that missingness patterns are dependent on the variable of interest.

Considering potential bias missing data may introduce in our work is best done when reflecting on the content of each chapter, and as a result we discuss it within the conclusions of each chapter. As a general overview, the most obvious potential bias we might be exposed to is the inability to train and validate models that require flow data on links 60, 65, 66, 68 and 69. Since these are nearby in space, and are clearly subject to missingness that is not MCAR, they may possess particular properties or traffic volumes that we cannot expose our modelling approaches incorporating flow data to. Further, missing speed

values for elongated periods may result in a bias where we neglect the reason for specific periods being missing, for example a change in traffic behaviour that occurs before the links start to report data. Finally, potentially missing incident flags may mean that models to detect incidents (considered in chapter 4) have higher reported false alarm rates than they would when evaluated on perfect data. On the other hand, methods trained to predict the duration of incidents (considered in chapter 6) may not be exposed to short-lived or very minor incidents that occur on the network but are often omitted from the data.

Seasonality

Since seasonality is a clear feature of traffic data, an exploratory analysis of our dataset would not be complete without a discussion of it. We provide plots of seasonal speed, flow and density for all of our links in appendix B.3 and note the main points here. We extract the seasonality by segmenting the data by time of week and taking the median value for each variable. Considering flow (Fig. B.2), the morning and evening commuter rush periods are quite evident across most links, however they are more prominent on some links compared to others. We also observe some links that experience considerably less flow than the links adjacent to them, which from a practical perspective will be a result of them being after a major exit that many vehicles often use. There is generally a noticeable, but sometimes small, decrease in flow on weekends compared to weekdays during these rush periods.

Considering speed (Fig. B.3), we generally see significant reductions on weekdays during rush periods, but significantly less so on weekends. There is also a group of links, starting at number 43 and ending at number 54 that do not see a significant speed drop as a seasonal feature, even during weekdays, suggesting these are not commonly used by commuters. Further, link 8 consistently has a lower seasonal speed than its surrounding links. The most extreme reductions in speed, both in terms of magnitude and duration are observed between links 25 and 42, with links 25 to 37 exhibiting clear speed drops even during weekends. Finally, inspecting seasonal density (Fig. B.4) highlights that links 7 to 13 appear to be significantly impacted by the morning rush period, more so than the evening one. We also see links 33 to 36 attain the highest seasonal density values, and even show significant increases in density on weekends. The lowest seasonal densities are observed near the very start and end of the M25.

Incident variation in space

It is also useful to visualise how the intensity of physical incidents, that is accidents and obstructions, varies spatially across the network. To do so, we

look forward to the methods in chapter 5, where we formulate a model that incorporates the spatial background rate of events along the M25. Using the output of this model, we generate a visualisation of the incident rate on the M25, shown in Fig. 3.4. Inspecting Fig. 3.4, we see two peaks in incident

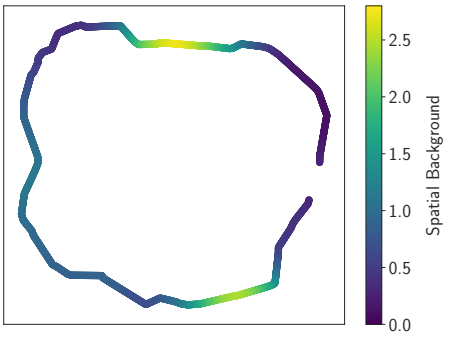


Figure 3.4: Background intensity of incidents on the M25, constructed using data between September 1st 2017 and September 30th 2018 using the methods detailed in chapter 5.

intensity along the M25. The largest peak, around 140 kilometres along the motorway, is located near the 'Potters Bar' junction. The second largest, around 25 kilometres along the motorway, is located between where the M25 meets the M26, and where the M25 meets the M23. Readers should refer to chapter 5 for a detail description of what this intensity means and how it is constructed.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	94	100	6	0	127	3374	0	1980	1940	0	4298	67251	9	26	1	22	438	56347	8	20	20	0	140	67825	9
1	3.6	13	133	102	7	0	117	3190	0	3420	3360	0	7131	3296	0	129	9	112	935	9025	1	34	35	0	176	3989	1
2	1.0	3	82	106	6	11	126	3179	0	2716	2685	0	5943	3344	0	34	2	29	333	8932	1	26	27	0	130	3426	0
3	0.5	2	56	105	8	7	129	3607	1	2280	2273	0	5213	3580	0	18	2	14	269	16344	2	22	22	0	142	3906	1
4	0.4	1	51	105	8	7	129	3513	0	2280	2278	0	5213	16758	2	14	1	11	210	9118	1	22	22	0	158	17053	2
5	4.5	2	68	108	8	0	129	3083	0	2908	2940	0	6540	3417	0	151	12	126	1800	50075	7	27	29	0	165	3458	0
6	0.9	2	33	109	8	5	130	3760	1	2289	2276	0	5160	3487	0	29	2	24	624	50557	7	21	22	0	171	3897	1
7	5.1	2	104	110	8	0	128	2880	0	2609	2600	0	5760	3514	0	167	12	145	2301	10969	2	24	24	0	244	3918	1
8	1.0	1	159	90	5	0	130	34198	5	1680	1635	0	4358	3319	0	42	2	29	753	41706	6	20	18	0	381	36618	5
9	0.7	1	121	100	6	0	128	3301	0	3109	2874	60	7098	291644	41	23	1	18	469	9774	1	31	30	1	340	291644	41
10	0.9	2	171	103	7	4	130	70916	10	3372	3273	120	8160	61803	9	31	2	25	798	79265	11	35	31	2	390	112911	16
11	5.9	13	547	103	10	3	130	101721	14	3300	3180	60	7800	300189	42	206	20	163	7081	110082	15	32	32	1	238	300189	42
12	1.0	2	200	108	8	14	126	327598	45	3092	2985	0	7680	327700	46	34	3	29	263	335895	47	29	30	0	168	327700	46
13	6.8	17	494	105	9	8	122	300162	42	3303	3191	84	6900	327582	45	234	23	199	3041	337536	47	32	33	1	309	327582	45
14	1.0	2	73	105	10	0	122	26300	4	3060	2760	0	6120	402026	56	34	4	29	448	73196	10	29	29	0	243	402129	56
15	1.2	4	72	106	9	0	122	3147	0	3645	3648	0	7516	3463	0	40	4	35	377	16116	2	34	37	0	162	3625	1
16	1.5	6	67	106	8	0	122	3029	0	3589	3611	0	7380	3414	0	52	4	45	486	15998	2	33	36	0	184	3576	0
17	0.5	0	53	108	8	0	127	3074	0	3600	3600	0	7317	4064	1	16	1	13	242	8239	1	33	34	0	357	4245	1
18	0.2	1	48	109	8	6	130	3212	0	2454	2579	0	5820	3442	0	6	0	5	110	56283	8	22	24	0	280	3548	0
19	1.2	4	78	108	7	6	127	3252	0	2292	2449	0	5580	5308	1	41	3	35	739	56332	8	21	24	0	207	5410	1
20	2.0	4	67	103	10	8	120	3328	0	4020	3841	0	8251	3330	0	71	7	60	636	52772	7	39	39	0	243	3461	0
21	3.0	6	89	103	9	6	125	3327	0	3391	3240	0	7143	3905	1	104	10	86	1756	52771	7	32	33	0	243	3907	1
22	10.0	22	187	110	8	0	126	3196	0	3900	3720	0	8335	3301	0	328	24	286	2762	52642	7	35	36	1	292	4112	1
23	0.5	1	114	110	9	8	130	3244	0	3195	3073	0	7382	14117	2	17	2	14	272	52690	7	29	31	0	176	14117	2
24	6.4	13	463	109	11	0	125	3199	0	3780	3641	0	7854	3314	0	212	22	184	3291	56255	8	35	38	1	200	5045	1
25	0.8	2	321	108	13	0	127	3256	0	3456	3379	0	7715	95444	13	26	4	22	517	56312	8	33	39	0	206	96510	13
26	2.6	9	294	106	15	0	126	3218	0	3827	3720	0	7980	3321	0	87	14	74	2166	56325	8	36	43	0	198	4401	1
27	0.9	2	311	106	16	0	127	3639	1	2852	2776	0	6262	3836	1	29	4	25	723	6582	1	28	33	0	190	6995	1
28	6.9	13	471	104	31	5	124	4187	1	4411	4283	0	8247	3739	1	242	102	201	4929	51179	7	43	63	0	207	4291	1
29	0.9	3	105	101	25	6	130	4050	1	3849	3752	0	7260	3947	1	33	10	26	559	50840	7	39	52	0	152	4294	1
30	2.5	4	71	100	17	0	122	3144	0	4834	4680	0	8335	3399	0	90	16	74	564	8154	1	48	57	0	232	5526	1
31	1.2	1	37	102	12	16	120	529542	74	3180	3104	0	6240	527684	73	42	6	36	270	532579	74	34	34	0	117	529618	74
32	3.9	6	358	105	10	0	125	3279	0	4624	4548	0	9600	3309	0	133	13	112	1468	55158	8	47	50	0	221	9480	1
33	1.0	3	731	106	13	0	123	3234	0	3714	3980	0	8700	3332	0	34	4	28	683	51878	7	42	49	0	334	5079	1
34	1.9	3	660	103	20	0	123	3233	0	4743	4691	0	10200	3367	0	68	21	57	858	55202	8	53	68	1	309	3389	0
35	1.7	5	882	101	37	0	124	3118	0	3480	3867	0	8138	3139	0	61	44	49	1022	55096	8	47	68	0	275	3196	0
36	1.6	3	304	96	39	18	116	3189	0	5460	4755	0	9926	3596	0	60	38	49	346	55158	8	64	89	0	360	3596	0
37	1.2	3	436	98	51	6	127	3163	0	3167	3096	0	5860	3286	0	46	42	35	739	36084	5	36	67	0	197	3286	0
38	7.1	17	416	97	26	6	120	3152	0	5280	4980	0	7977	3247	0	267	80	214	4288	15464	2	58	68	0	226	3247	0
39	0.7	2	164	98	14	0	130	3544	0	3865	3642	0	6808	3312	0	26	3	20	516	15856	2	41	42	0	155	3956	1
40	0.6	2	224	100	14	0	130	6178	1	3000	3044	0	5100	42115	6	23	3	18	467	18228	3	31	34	0	194	45842	6
41	1.9	5	199	98	15	0	120	3894	1	4374	4409	0	7560	3664	1	69	10	56	1355	25150	3	47	51	0	216	11590	2

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	130	88	19	0	126	4208	1	4298	4374	0	7560	3659	1	55	11	38	954	25248	4	50	57	0	382	11231	2
43	4.6	11	160	97	12	0	128	3849	1	4286	4354	0	7560	3563	0	170	20	129	2656	37168	5	46	49	0	229	11196	2
44	0.8	1	75	101	9	0	121	3331	0	4345	4339	0	7380	53417	7	28	3	24	573	36533	5	44	46	0	344	54401	8
45	1.6	5	75	101	9	0	121	2996	0	4471	4568	0	8008	3321	0	58	5	48	1171	36154	5	45	49	0	194	3631	1
46	0.2	1	70	101	9	0	130	3170	0	4270	4380	0	7819	7808	1	9	1	7	176	36091	5	43	47	0	202	7867	1
47	0.9	3	55	100	11	0	130	3161	0	4747	4906	0	8820	3329	0	33	3	25	502	16204	2	48	52	0	354	3415	0
48	1.4	4	63	101	10	0	126	3098	0	4738	4884	0	8640	3338	0	49	4	39	704	16141	2	47	52	0	245	3401	0
49	1.8	4	54	101	8	0	129	3098	0	4740	4877	0	8798	3572	0	66	5	54	1107	36019	5	47	50	0	213	3635	1
50	2.3	7	51	103	10	0	129	3130	0	4075	4213	0	8620	3281	0	80	7	64	1200	31224	4	40	42	0	237	6373	1
51	0.9	3	35	101	9	0	126	3454	0	3572	3683	0	7260	16020	2	33	3	26	661	30722	4	36	39	0	174	25861	4
52	3.8	12	72	104	8	0	124	3154	0	4227	4382	0	8256	3265	0	132	10	111	1667	13977	2	41	44	0	306	6059	1
53	1.0	3	56	108	7	0	130	3775	1	2729	3079	0	6105	3317	0	33	2	27	588	16004	2	25	30	0	158	5612	1
54	0.4	2	70	105	7	0	130	3267	0	2048	2412	0	5520	4385	1	14	1	11	478	14512	2	20	24	0	161	5572	1
55	0.5	1	82	103	7	0	124	3874	1	3300	3414	0	7560	4170	1	18	1	15	379	16769	2	32	35	0	236	6180	1
56	6.1	16	153	105	8	0	122	3152	0	3601	3766	0	7433	3253	0	209	15	180	3128	35176	5	35	38	0	289	8033	1
57	0.8	3	258	103	7	0	129	3156	0	3195	3276	0	6713	3322	0	26	1	21	542	31491	4	31	34	0	172	3586	0
58	4.0	12	453	102	9	0	121	3153	0	3557	3637	0	7054	3267	0	142	12	120	2072	36074	5	36	39	0	179	5532	1
59	0.9	3	273	103	10	0	130	10301	1	2858	2856	0	5880	3463	0	33	3	26	682	41552	6	29	30	0	209	11355	2
60	3.4	11	373	100	13	0	130	4007	1	-	-	-	-	720001	100	122	16	95	2436	50213	7	-	-	-	-	720001	100
61	0.8	2	273	106	10	0	130	218976	30	3540	3271	0	6600	476120	66	29	3	23	606	259178	36	35	31	0	225	483616	67
62	7.8	20	692	101	12	0	129	321955	45	3610	3535	0	7440	219735	31	275	30	218	5611	347728	48	42	28	0	397	337586	47
63	1.0	3	128	92	10	0	130	547741	76	3120	2895	0	5820	476453	66	40	4	28	732	551449	77	40	18	0	340	549607	76
64	4.7	11	335	97	12	0	130	482525	67	3584	3325	0	6720	605400	84	176	20	131	2843	500709	70	37	35	0	259	606150	84
65	0.6	2	126	-	-	-	-	717854	100	-	-	-	-	720001	100	-	-	-	-	720001	100	-	-	-	-	720001	100
66	1.8	4	184	101	9	0	126	3451	0	-	-	-	-	720001	100	64	5	51	1289	18641	3	-	-	-	-	720001	100
67	4.5	11	289	101	7	0	123	3174	0	3355	3283	0	6968	3304	0	159	11	136	1789	47533	7	33	33	0	168	5330	1
68	0.9	2	142	105	9	0	130	206202	29	-	-	-	-	720001	100	31	2	25	649	213080	30	-	-	-	-	720001	100
69	0.5	1	103	101	7	0	122	3198	0	-	-	-	-	720001	100	16	1	13	329	47780	7	-	-	-	-	720001	100
70	1.9	6	102	104	8	0	122	3161	0	3501	3612	0	7782	3324	0	66	5	57	1152	16388	2	34	37	0	300	3553	0
71	9.5	27	211	105	8	0	122	3168	0	3491	3614	0	7655	3266	0	330	25	281	4289	45838	6	34	37	0	154	3500	0
72	0.9	3	90	101	8	0	120	3840	1	2350	2448	0	5340	4785	1	31	3	26	494	49945	7	23	26	0	147	5821	1
73	3.4	10	83	104	7	0	122	5905	1	3254	3363	30	6840	4209	1	120	8	102	1210	48570	7	32	34	1	158	6213	1
74	1.1	4	74	106	8	0	130	7115	1	2466	2535	0	5460	7193	1	36	3	30	772	49780	7	24	25	0	114	9069	1
75	7.8	24	255	104	8	0	127	7276	1	3099	3090	0	6291	4600	1	269	19	220	3993	53073	7	31	32	0	196	7845	1
76	1.2	2	406	100	9	0	126	6604	1	2040	2048	0	4646	6664	1	42	4	33	1376	51001	7	21	22	0	234	9130	1

Table 3.1: Summary statistics of data, computed using all available data. For readability, the table has been split across two pages, and headers are included on both pages. Num: number. IQR: inter-quartile range. Link 0 is the link just below the Dartford crossing, and links are numbered increasingly as one moves clockwise along the M25 from this point. Link length is given in kilometres, speed in kilometres per hour, flow in vehicles per hour, travel time in seconds and density in vehicles per kilometre.

Chapter 4

Anomaly detection from fluctuations in the flow-density relationship

In this chapter, we describe and validate a novel data-driven approach to the real time detection and classification of traffic anomalies based on the identification of atypical fluctuations in the relationship between density and flow. For aggregated data under stationary conditions, flow and density are related by the fundamental diagram. However, high resolution data obtained from modern sensor networks is generally non-stationary and disaggregated. Such data consequently show significant statistical fluctuations. These fluctuations are best described using a bivariate probability distribution in the flow-density plane. By applying kernel density estimation to high-resolution data from NTIS, we empirically construct these distributions for London's M25 motorway. Curves in the flow-density plane are then constructed, analogous to quantiles of univariate distributions. These curves quantitatively separate atypical fluctuations from typical traffic states.

Although the algorithm identifies anomalies in general rather than specific incidents, we find that fluctuations outside the 95% probability curve correlate strongly with the spikes in travel time associated with significant congestion incidents. Moreover, the size of an excursion from the typical region provides a simple, real-time measure of the severity of detected anomalies. We validate the algorithm by benchmarking its ability to identify labelled incidents in historical NTIS data against some commonly used methods from the literature. Detection rate, time-to-detect and false alarm rate are used as metrics and found to be generally comparable except in situations when the speed distribution is bimodal. In such situations, the new algorithm achieves a much lower false alarm rate without suffering significant degradation on the other metrics. This method has the additional advantages of being self-calibrating and adaptive:

the curve marking atypical behaviour is different for each section of road and can evolve in time as the data changes, for example, due to long-term roadworks.

4.1 Introduction & problem relevance

As discussed, an important problem with significant existing research in highway traffic management is incident detection: monitoring the data streams from embedded sensors that report the traffic state on infrastructure in real-time and raising flags to alert operators to potential problems as they happen. A high level understanding of the typical behaviour of traffic on a section of road is provided by the fundamental diagram which assumes a functional relationship between vehicle density and vehicle flow. See [144] for a review. This relationship is valid for 'near steady-state' conditions [145], where data has been aggregated on an appropriate time-scale and therefore describes the average behaviour of steady-state traffic. In this chapter, we explore how analysis of the fluctuations in the high resolution, non-aggregated and non-steady-state flow-density data can be used to perform real time anomaly detection, by identifying periods of time when the flow pattern on a stretch of road is atypical. Our objective is to provide a systematic, data driven definition of 'atypical' behaviour and to suggest a principled way of tracking the severity of incidents as they occur in real time.

We consider anomalies to correspond to cases when the traffic state is atypical in a statistical sense. We emphasise from the outset that anomaly detection and incident detection are not exactly the same thing. Network operators, road users and other stakeholders are generally interested in *incidents* such as collisions, obstructions or lane closures. Such incidents may cause disruption resulting in an atypical flow pattern that can be picked up by an anomaly detection system. However, this is not necessarily the case a priori. For example, a broken down vehicle during a period of low demand may not impact upon the traffic flow at all. This incident would therefore not result in any anomaly. Conversely, anomalies like so-called 'phantom traffic jams' are sometimes observed that are not associated with any underlying incident or cause. In this chapter, we investigate the correspondence between incidents and anomalies in detail using labelled incident data to establish that the overlap between the two is sufficient to be of use in practice.

The methodology presented throughout this chapter is a purely data-driven method for identifying significant deviations from the typical behaviour of the traffic on a road section using time series data from the M25. The key idea is exploit the large volume of data to arrive at a robust understanding of the range of typical fluctuations about the flow-density relationship that is tailored

to each road section and to interpret big excursions from this range as proxies for significant incidents. We adopt a 'macroscopic' perspective meaning that the objective is to detect when the collective behaviour of the traffic on a road section is unusual in some sense. This macroscopic perspective means that we do not directly detect individual incidents like collisions, stationary vehicles or lane closures but rather changes in flow patterns that can be a consequence of such incidents. To validate our approach, we measure how periods identified as atypical relate to existing labels of incidents on roads provided by NTIS, and benchmark this comparison against some other available methods from the literature. Although we are initially agnostic about *why* a particular configuration is atypical we nevertheless find our criteria typically detect the most extreme congestion incidents. Our entirely unsupervised, data-driven methodology offers comparable or better performance than some existing models based on learning patterns from the data.

The main advantages of our approach, compared to specifically trained incident detection methods are as follows. Firstly, calibration is straightforward, requiring no labelled data to determine typical behaviour, instead only a representative sample of data taken from a stretch of road is required. Since road incidents are rare in absolute terms, this means that collecting a dataset to calibrate the method on is far easier than collecting a representative set of incidents on a specific section of road. We show in section 4.3.3 that 3 weeks of data is sufficient to identify stable periods of typical behaviour, whereas it is highly unlikely that the same time-period would provide a sufficient number of incidents on a single section of road to adequately train a deep learning model for example. Secondly, it is self-adaptive. One can account for long-term changes in the behaviour on certain sites by simply reapplying the methodology on a new representative sample. Thirdly, it is fully interpretable, with a clear reason as to why a data-point is marked atypical, and allows for direct comparison between different locations around a network with reference to the typical behaviour at each site. Finally, it is shown to also capture periods of labelled incident data, showing that physical incidents on the network such as accidents and breakdowns correspond to a subset of the atypical periods a section of road experiences. In terms of practical application, this analysis should be thought of as a filter that could be help human operators of the smart motorways infrastructure to prioritise their attention. Regardless of cause, it may be useful to inform operators which parts of their infrastructure are experiencing the most atypical situations, and hence where intervention might be able to help the system to return to normal.

In chapter 2, we reviewed existing work in automatic incident detection. We note that the approach proposed in this chapter takes some ideas from this broad range of literature, but has distinct advantages and applicability. Like the

McMaster algorithm, we segment the flow-density diagram into distinct regions that separate the typical from the atypical. Unlike the McMaster algorithm, we do not pre-specify how the segmentation of the diagram should be done. Instead, the segmentation is defined by the data itself in a parameter-free way, similar to the philosophy underpinning the SND algorithm. As is the case for SND, using data to determine its own bounds requires that statistically robust approaches are used. As such, our method should be seen as combining the idea of segmenting the flow-density diagram with the robust approaches in the literature, whilst side-stepping the two difficult problems of obtaining labelled data and calibrating an incident detection system with many parameters.

The remainder of this chapter is structured as follows. First we detail what parts of the data are used and how we define typical behaviour on a link. We then consider time-scales of atypical incidents in our data, and compare this to existing incident flags. After, we discuss when to raise atypical incident flags, methodologies to determine the severity of incidents, and consider how our incidents correlate with travel time spikes. Finally, we consider how our method performs if one were to use it only to detect labelled incidents in the dataset, comparing it to existing incident detection methodologies from the literature. We then summarise our findings and results.

4.2 Data details

For the work in this chapter, we first develop the methodology and illustrate its usage on a subset of our data, the first 10 weeks collected. This data is recorded between April 7th 2017 and June 16th 2017, and summary statistics of this subset are given in appendix B.4.1. As this data was taken purely to determine if our methodology was viable and to perform exploratory analysis, we arbitrarily took a window starting at the earliest collection point we had.

After verifying our methodology is viable, we utilise a much larger subset of the dataset when considering model validation in section 4.6 to show how the proposed approach compares to existing methods over an extensive range of observed incidents. This subset is taken between September 1st 2017 and November 1st 2018. Such a long window ensures we have a sufficient amount of data to apply our model on a large representative sample of real scenarios that it would encounter if taken forward for industrial use. This subset, and the separation into training and testing is further discussed in section 4.6.

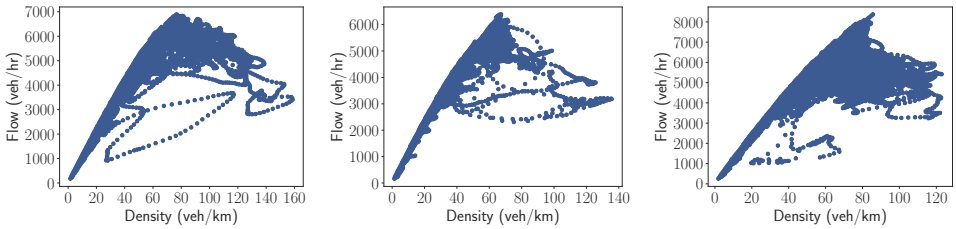
4.3 Data-driven characterisation of the flow-density relationship and identification of atypical configurations

We recall from chapter 3 that NTIS does not provide direct measurements of density. Throughout this chapter, when we refer to 'density' we are really referring to a proxy measure obtained by dividing the average flow by the average speed. It has been noted that this proxy quantity for density can be biased [146]. To obtain quantitative estimates of parameters like critical density, for example, one must be careful to account for this bias. We do not need to measure such quantities however since our objective is to distinguish between typical and atypical fluctuations in the traffic state. Direct use of a proxy measure for density is therefore appropriate provided this does not introduce significant artificial scatter into the data. To test if this is the case, we consider the empirical relation derived in [141], where NTIS data is taken along with microscopic measurements and a formula to convert from time-mean quantities to space-mean quantities is given. Applying this to our dataset, we see qualitatively similar properties hold, there is still significant variation, and our methods give generally the same conclusions at each step. Based on these calculations, we use the raw data provided by NTIS from here onwards. One could formally test for introducing scatter by binning the data based on the two definitions of density and applying a number of paired statistical tests comparing the distributions of flow in each bin, however this is not fundamentally important to our methodology. Rather, the important aspects will be that all anomalies are defined relative to some determined baseline, this baseline can be constructed using existing NTIS data, and we are consistent with our use of variables when using the methodology. Example flow-density series using the raw NTIS data are given in Fig. 4.1a, and the transformed data using the methods in [141] are given in Fig. 4.1b.

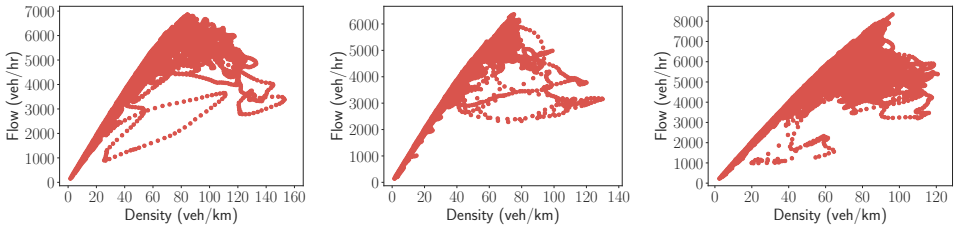
4.3.1 Using kernel density estimation to model the flow-density relationship

Whilst a classical way to consider the properties of a link are to inspect the corresponding fundamental diagram, the raw disaggregated flow-density series shows statistically significant variability. The reasons for this variability are discussed at length in [145]. Our methods rely on utilising this variability in the data, retaining valuable information in the joint distribution of flow and density that is typically lost when this is replaced by a single curve representing the average relationship.

By modelling the flow-density relationship as a probability distribution,



(a) Example flow-density series using time mean speed data and a density inferred from this.



(b) Example flow-density series using approximated space mean speed data and a density inferred from this.

Figure 4.1: Example flow-density series using 10 weeks of M25 data, taken from 3 distinct links. Left column: link between junctions 2 and 3 (3636 metres long). Middle column: link between junctions 3 and 4 (4525 metres long). Right column: link between junctions 8 and 9 (10,009 metres long).

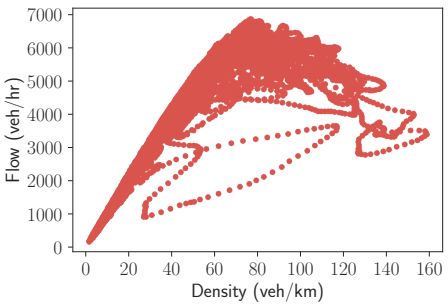
we can incorporate this variability which would typically be removed when aggregating and filtering parts of the raw data. The price to be paid is additional analytical complexity: reasonable functional forms are required for the joint distribution of flow and density. We suggest sidestepping this question by using the data to construct the required distribution directly.

A simple and computationally efficient way to do this is using kernel density estimation (recall KDE from section 2.5.1). As discussed, we aim to approximate an unknown distribution $p(\boldsymbol{x})$, with $\boldsymbol{x} \in \mathbb{R}^d$, using samples $\boldsymbol{X}_i$, $i \in \{1, 2, ..., N\}$ with an estimate of the form

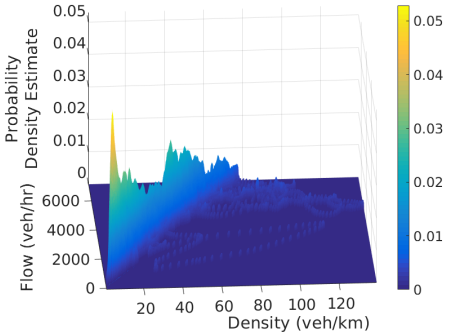
$$\hat{p}_{\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{1}{N} \sum_{i=1}^{N} k_{\boldsymbol{\Sigma}}(\boldsymbol{x} - \boldsymbol{X}_i). \tag{4.1}$$

As before, $\boldsymbol{\Sigma}$ is the bandwidth matrix controlling the amount of smoothing we apply to the data. We determine such an approximation for the flow-density data, using the method of Chacón and Duong, detailed for the multivariate case in [110]. We specifically use the R implementation described in [147]. Doing so yields a two-dimensional distributional interpretation of the flow-density data for any given link.

Fig. 4.2 shows an illustrative example of this process. Fig. 4.2a shows a scatterplot of the data from a representative M25 link between junctions 2 and 3. Fig. 4.2b shows a 3D rendering of the corresponding joint distribution



(a) Scatterplot generated from 10 weeks of data from a representative M25 link between junctions 2 and 3.



(b) Kernel density estimate of the joint distribution of flow and density generated from this data.

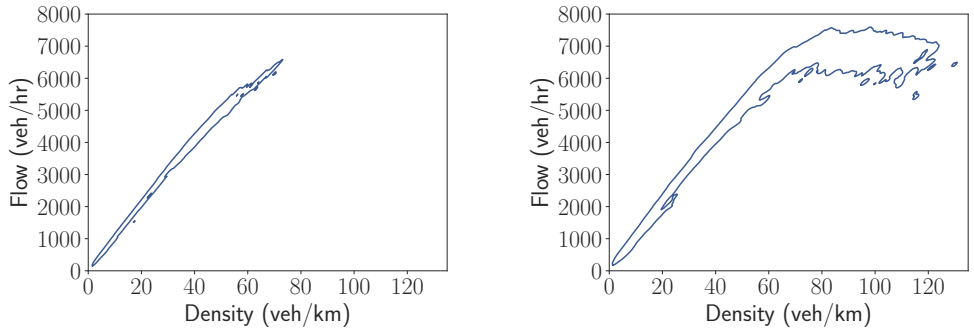
Figure 4.2: Representative example of the use of KDE to obtain a representation of the joint distribution of flow and density for a single link of the M25 using NTIS data.

obtained using KDE. The latter makes clear that most of the mass is centred around two modes, a feature that is much less evident from the scatterplot. We refer to these as the low density and high density modes. Note that this multimodal structure is not a sampling bias since the data is uniformly sampled at 1 minute intervals across the shown 10 week data collection window. Rather it reflects the fact that this link spends a lot more time in the low density regime compared to the high density regime.

4.3.2 Separating typical and atypical configurations

We now turn to the central point of the chapter: using KDE to systematically draw a distinction between typical and atypical behaviour of the traffic on a link. The idea is to calculate a level curve of the KDE representation, $\hat{p}_{\boldsymbol{\Sigma}}(\boldsymbol{x})$, of the joint distribution of flow and density that encloses a predefined portion, $1-\alpha$, of the total probability. Level curve here means a curve satisfying $\hat{p}_{\boldsymbol{\Sigma}}(\boldsymbol{x}) = \text{constant}$. Points inside this level curve are considered typical whereas points outside are considered atypical. By adjusting the parameter α we can adjust the relative frequency between typical and atypical. This is essentially a two-dimensional bivariate analogue of a one-dimensional confidence interval for a univariate distribution. This construction is a topic of active research in the statistics literature because inference of level curves is analytically and computationally highly non-trivial. See [148] and the references therein. In our application however, the large volume of data available from NTIS means practically useful level curves can be obtained without addressing the much harder problem of quantifying the uncertainty in these curves.

Once we have obtained the KDE representation, $\hat{p}_{\boldsymbol{\Sigma}}(\rho, f)$, of the joint



(a) Data taken between junctions 2 and 3 on the M25.

(b) Data taken between junctions 8 and 9 on the M25.

Figure 4.3: Level curves corresponding to a threshold $\alpha = 0.05$ created using data from two different links. Each contains 95% of the probability mass of the respective KDE representations.

distribution of density, ρ, and flow, f, the computational task is to find z^* such that the integral of all mass above this height equals some threshold value $1 - \alpha$. We determine z^* by finding the (unique) root of the function

$$H(z) = 1 - \alpha - \int_{\rho_{\min}}^{\rho_{\max}} \int_{f_{\min}}^{f_{\max}} \hat{p}_{\mathbf{\Sigma}}(\rho, f)\, \Theta\left[\hat{p}_{\mathbf{\Sigma}}(\rho, f) \ - z\right] df\, d\rho. \tag{4.2}$$

Here $\Theta(x)$ is the Heaviside function,

$$\Theta(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases},$$

which serves to threshold the integrand by setting it to zero when $\hat{p}_{\mathbf{\Sigma}}(\rho, f)$ is lower than z. In Eq. (4.2), the root of $H(z)$ will simply be the value at which the integral of our thresholded distribution is equal to the desired threshold. Note that as our data has bounds on density and flow, we only need to integrate between these.

Having found the desired height z^*, we slice the surface $\hat{p}_{\mathbf{\Sigma}}(\rho, f)$ at this value, thereby defining a level curve $c(\rho, f)$ that contains the desired amount of mass, with points lying outside the curve being outliers at the α threshold. For example, if we choose $\alpha = 0.05$, we would have 95% of the data inside $c(\rho, f)$, with the remaining 5% considered atypically large deviations from the standard behaviour. The level curves for taking $\alpha = 0.05$ are shown in Fig. 4.3 for some representative links. These two examples are illustrative of how we capture known features of traffic flow, in this case recurrent bottlenecks. Flow breakdown is generally absent on the link shown in Fig. 4.3a, however the link shown in Fig. 4.3b clearly sustains a high flow as we move from densities around 60 veh/km to 120 veh/km, suggesting it is impacted by a recurrent

bottleneck. As the traffic state evolves in time on any given link, it traces out a trajectory in the $\rho - f$ plane. NTIS can track this trajectory almost in real time. When the trajectory makes an excursion outside of the curve delimiting the region of typical behaviour, we call it an 'atypical traffic incident'. One can numerically check if a flow-density data-point is outside of this typical region by solving the 'point in polygon' problem. Further discussion of this is given in appendix C.2.

4.3.3 Stability of the flow-density relationship over time

In order for the concept of an atypical traffic incident to be useful, the level curves delimiting the region of typical behaviour should be stable in time. Our analysis showed this to be the case. Another way to phrase this is that, if we are to use this contour approach for any real-time applications, we should first be confident that the system we are observing is stationary over relevant timescales. For our purposes, stationary means that when we look at subsets of our data, the flow-density relationship and the corresponding contour do not significantly change. Whilst we may get different incidents and severity of incidents in different periods of time, we would hope that the typical behaviour of the flow-density series remains reasonably constant. To test this, we construct the contours of typical behaviour using 3 disjoint 3-week subsets of data. These subsets begin on April 7th 2017, April 28th 2017 and May 19th 2017, and were chosen to investigate the stability at sequential time-periods in the data. Summary statistics for these subsets are given in appendix B.4.2, B.4.3 and B.4.4. For each of the 3 week periods, we perform kernel density estimation and determine the 95% contour. We plot each for two different links, showcasing how different types of contour, one with a significant amount of data present at high densities and one without, can be stationary in Fig. 4.4.

Considering Fig. 4.4, we see generally similar results across each 3 week-period, both for the contours with only low density (0-70 veh/km) data, and for the one with high-density data (up to 120 veh/km) present. This showcases that despite the clearly different dynamics a link may experience, we can still find a stationary distribution of the flow-density series due to the recurrent nature of traffic. All of the windows shown have incidents in them, however the typical behaviour remains constant on each link. After verifying stationarity, we consider only contours fit to 3 weeks of data throughout the remainder of this work.

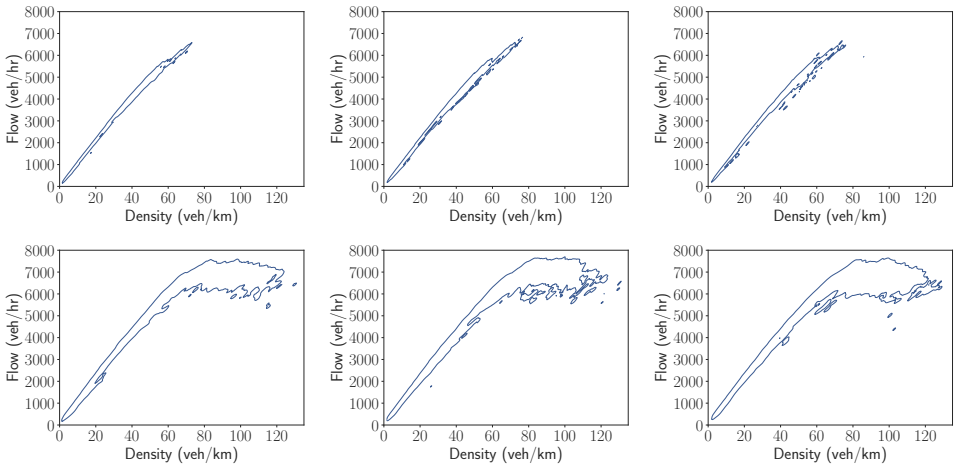
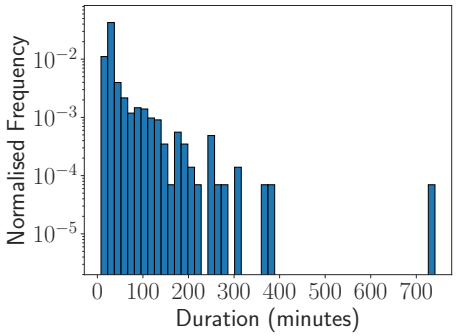


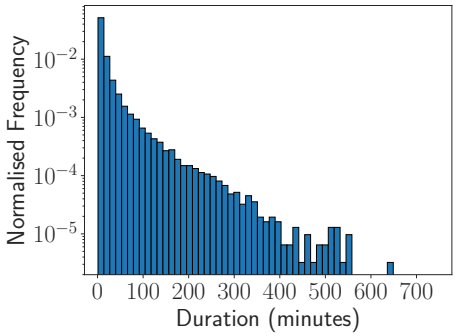
Figure 4.4: Contour plots of data segmented into 3-week periods. Top row: data taken from a link between junctions 2 and 3, 3635 kilometres in length. Bottom row: data taken from a link between junctions 15 and 16, 7146 kilometres in length. We see generally stable contours using 3-week segmentation of the diagrams, shown by the similar shape of each region of typical behaviour. Clearly, the diagrams in the top-row typically experience significantly lower density values than those in the bottom-row, however all 6 diagrams shown have accidents, obstructions and abnormal traffic incidents in, so this behaviour appears to be a recurrent feature rather than caused by occasional incidents.

4.3.4 Timescales of atypical traffic incidents and comparison to NTIS incidents

We might expect that many atypical traffic incidents correspond to small excursions that rapidly return to the typical region. Very brief excursions are unlikely to be of much practical interest. It is therefore useful to compare the distribution of durations of atypical traffic incidents generated by the procedure described above with the distribution of durations of incidents flagged by NTIS. We scan all the links in our data and determine the start and end points of each NTIS accident and obstruction incident and each atypical traffic incident. This gives us a set of durations for each category. The resulting histograms of durations are shown in Fig. 4.5. From Fig. 4.5a, we see the most common duration of an NTIS accident or obstruction incident is between 25 to 40 minutes. In contrast, abnormal traffic incidents, shown in Fig. 4.5b, are most commonly found to be less than 10 minutes in duration and the distribution simply decays as duration increases. This clearly highlights two things. The first is that a large number of the abnormal traffic incidents may be so short lived that the actual impact on traffic on a link is minor. An infrastructure operator, for example, might not be interested in seeing these minor deviations. It can therefore make sense to apply secondary thresholding to single out the most significant ones. This is done in section 4.4. Secondly, when an incident



(a) Accident and obstruction incidents flagged by NTIS.



(b) Atypical traffic incidents generated by our algorithm.

Figure 4.5: Comparison of the empirical distributions of durations of NTIS accidents and obstructions incidents and durations of atypical traffic incidents defined through our methodology. The two distributions are somewhat distinct, with atypical traffic incident durations most commonly being between 1 and 10 minutes, and mainly concentrated at small values less than 100 minutes. Accidents and obstructions on the other hand are most commonly between 25 and 40 minutes, and the histogram is also reasonably constant between durations of 60 and 140 minutes. To ensure a representative set of NTIS incident flags, we have used data between April 7th 2017 and June 20th 2017 in both of these plots. Summary statistics of this dataset are given in appendix B.4.5. This gives us a continuous period of operation to consider data from, and is reflective of the wider period we have been using for exploratory analysis of our methodology.

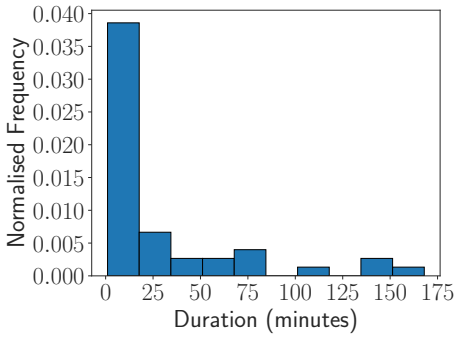
occurs, it may not perturb the traffic state significantly for a number of minutes, causing only minor obstructions to begin with. However these can grow in time so that one eventually sees strongly atypical flow-density behaviour. There is therefore a natural notion of the severity of an incident which can evolve in time. An approach for quantifying severity is discussed in section 4.5.

Finally, one should consider that the differences observed between Fig. 4.5a and Fig. 4.5b are likely also a consequence of the definitions of accidents and obstructions compared to the definition of atypical traffic incidents. Accidents and obstructions require either a collision between a vehicle and one or more vehicles or the environment, or some form of physical obstruction to be present on a link. As a result, they are unlikely to be resolved on very short time-scales, for example 5 minutes or so. Conversely, atypical traffic incidents do not necessarily have to correspond to a physical problem, and can just be the result of short-lived drops in speed or flow that occur for a variety of reasons. As a result, we would intuitively expect that some of these atypical incidents may be very short lived, and others that do overlap with accident and obstruction flags in the data last for a comparable time-scale.

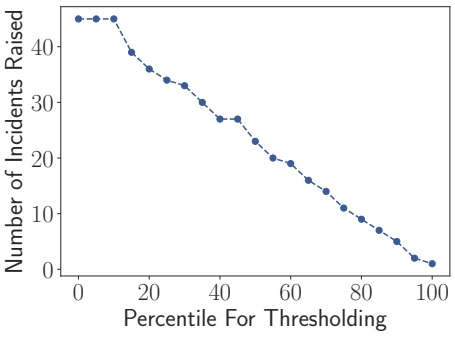
4.4 Identification of significant 'Deviation from Typical Behaviour' (DFTB) incidents

In this section we describe a protocol for identification of significant abnormal traffic incidents by filtering out those of short duration. We call these 'Deviation from Typical Behaviour' (DFTB) incidents to distinguish them from the incident categories NTIS already records, discussed in section 3.2.2. We expect significant commonality between the two types of incident and some comparisons are made in the remainder of this chapter.

4.4.1 Determining when to raise flags



(a) The duration of atypical traffic incidents observed in our 3-week training window. The majority are short term as shown by the initial large peak.



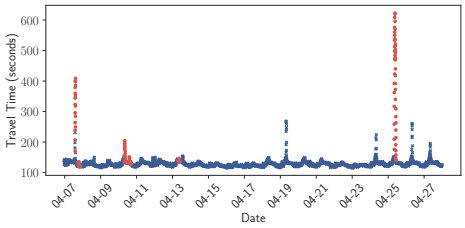
(b) The number of DFTB incidents raised in the training window as a function of the percentile chosen as the threshold for duration. At the extremes, we have every single deviation registering an incident (threshold = 0th percentile) or no incidents raised (threshold = 100th percentile).

Figure 4.6: Analysis of the duration of atypical traffic incidents and the number of incidents raised. Plots are shown for a representative link, between junctions 2 and 3 on the M25.

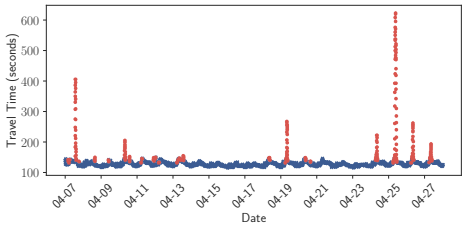
We propose to make decisions on raising DFTB incident flags by defining a time threshold for which trajectories must be outside of the typical region in order be considered significant. This filters out the aforementioned minor fluctuations. Throughout this section, we use 3-weeks of data to perform the KDE and to identify the curve enclosing the typical region. Two further 3-week subsets of data are then used as a test set within which we seek to identify incidents so that decisions are based solely on past data. As before, the 3-week subsets used begin on April 7th 2017, April 28th 2017 and May 19th 2017, with the first being the training window. Summary statistics for these subsets are given in appendix B.4.2, B.4.3 and B.4.4. To begin, we plot the histogram of durations of atypical traffic incidents observed on a representative link during the training window. A example is shown in Fig. 4.6a. We see a large number

of short-duration excursions from the typical region, and far fewer large ones. These extreme deviations represent the most severe incidents on the link, with the less extreme ones being located in the short-term deviations. By setting the threshold at a given percentile of the duration distribution we obtain a systematic and purely data driven notion of significance that does not rely on any details of individual links. Obviously selecting a higher threshold percentile gives fewer incident flags. The relationship between the threshold and the number of incidents is shown in in Fig. 4.6b.

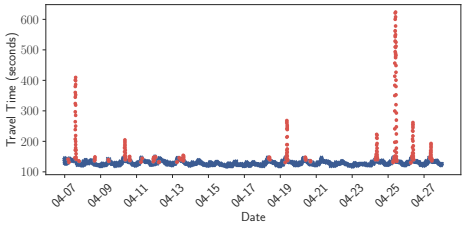
It is natural to look where DFTB incidents appear in travel time series and to compare them to incidents flagged by NTIS. In Fig. 4.7, we plot the travel times for the link in question, separated by colour and symbol into cases with and without DFTB flags raised. For reference, we also show the same plot using NTIS flags, using deviation from profile, accident and obstruction flags, and finally show how our flags change for a range of different thresholds.



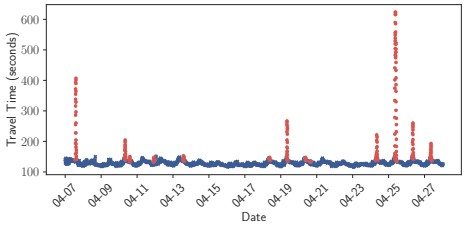
(a) NTIS incident flags in the training data.



(b) DFTB flags in the training data without thresholding.



(c) DFTB flags raised in training data with threshold at 40th percentile.



(d) DFTB flags raised in training data with threshold at 80th percentile.

Figure 4.7: Travel times for a single link, separated into cases with and without incident flags. On each plot, blue (×) represents a data-point with no incident flag, whereas red (•) represents a raised flag. Plot (a) is of travel times coloured and symboled to represent NTIS incident flags, whereas the rest are done to represent the presence of a DFTB flag at various thresholds. For reference, the threshold values for the 40th and 80th percentiles are 4 and 23 minutes respectively.

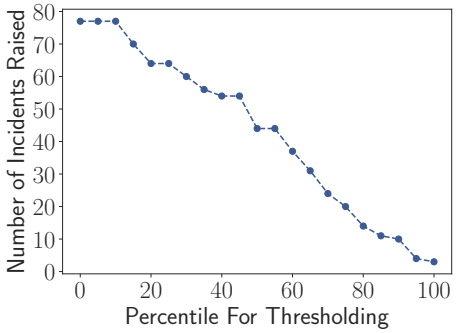
From Fig. 4.7, it is clear that atypical fluctuations in the flow-density relationship identify spikes in the travel times on a link. This identification is indirect since we never modelled travel times. For low thresholds, we see many DFTB flags are raised during low-travel times. However as we increase the threshold we identify only the large spikes in the series. Fig. 4.7a also shows

that some DFTB incidents are associated with significant spikes in travel times that are not associated with any NTIS incident. It would be interesting to investigate these further.

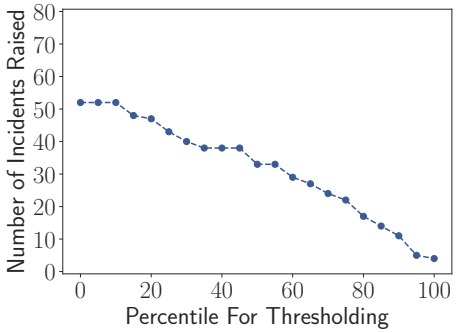
From an operational perspective, recall from chapter 3 that some flags are automatically entered into the system, where as others have to be accepted by a human operator. As a result, some incidents may not have been entered into the system simply when an operator was particularly busy, or had their attention elsewhere. Hence, whilst NTIS provides useful data to compare our methodology to, we have to acknowledge that the process by which the data is generated is subject to these practical considerations. In such cases, our algorithm may be raising flags at periods where the data should have a flag, but a flag was not accepted by the operator. Doing so acknowledges that we do not have perfect 'ground truth' information.

4.4.2 Results for DFTB flags in the test data

The plots in section 4.4.1 are for the training data that was used to determine the contour defining the typical region. It is therefore not surprising that the DFTB incidents selected are in close correspondence with atypical incidents observed in the travel time series. We now check that the process generalises to unseen data, using two subsequent 3 week periods of test data that has not been used to construct the curve enclosing the typical region. Recall that the first 3 week test period starts April 28th 2017 (test set 1) and the second starts May 19th 2017 (test set 2). In Fig. 4.8, we show the number of incidents raised as a function of threshold again, but for each of the two test periods using unseen data.

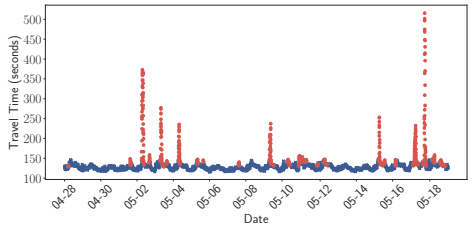


(a) Test set 1: 3 weeks of data starting April 28th 2017.

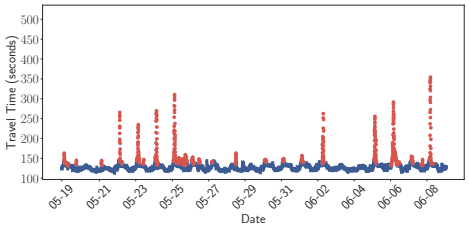


(b) Test set 2: 3 weeks of data starting May 19th 2017.

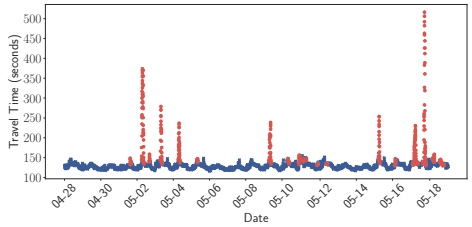
Figure 4.8: Plots of the number of DFTB incidents raised as a function of threshold percentile for each of our test datasets. We see similar relationships to those observed in Fig. 4.6b, with slightly more incidents raised in these windows at low thresholds.



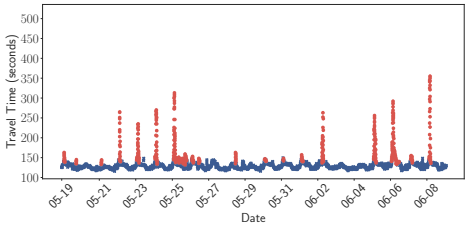
(a) DFTB flags in test set 1 with threshold at the 40th percentile.



(b) DFTB flags in test set 2 with threshold at the 40th percentile.



(c) DFTB flags in test set 1 with threshold at the 80th percentile.



(d) DFTB flags in test set 2 with threshold at the 80th percentile.

Figure 4.9: Travel time series for test set 1 (left column) and test set 2 (right column). On the top row, we show results when thresholding at the 40th percentile, and on the bottom row results for thresholding at the 80th percentile. As in the training case, we see the significant spikes in the series are captured by our methods, however thresholding at the 40th percentile also identifies many seemingly small perturbations in the travel times. As we move to the thresholding at the 80th percentile, we see our DFTB flags match only the significant spikes and a small number of low travel time periods. For reference, the threshold values for the 40th and 80th percentiles are 4 and 23 minutes respectively. As before, blue ($\times$) represents a data-point with no incident flag, whereas red ($\bullet$) represents a raised flag.

The series of travel times with colours and symbols used to represent the presence of incident flags in are shown in Fig. 4.9. Inspecting our results in Figs. 4.8 and 4.9, we see that our results are both quantitatively and qualitatively similar in the training and testing scenarios considered.

4.5 Severity ranking of DFTB incidents in real time

A clear disadvantage of duration thresholding for raising DFTB incident flags is that the process is retrospective: the duration of an incident is only known once traffic conditions return to normal. Since NTIS updates every minute, this problem can be avoided by assigning a dynamical severity level to incidents based on the evolution of the trajectory in the density flow plane. Minor fluctuations can then be removed almost in real time by filtering on this severity level.

4.5.1 Dynamical severity measures for DFTB incidents

There are many ways one could ascribe a severity level to an excursion from the typical region. Relevant features include how far the trajectory has deviated from the typical region and how long the trajectory has spent outside the typical region. Since these could be combined in different ways, perhaps with other features, there is no single 'correct' quantitative measure of severity. We argue that any proposed measure should satisfy two practical requirements:

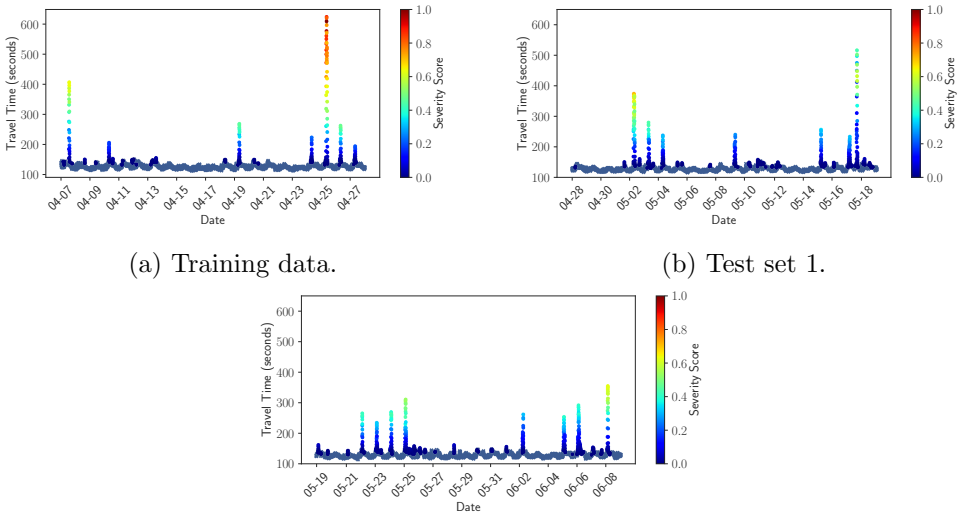
1) severity should be 0 when a trajectory is inside the typical region,

2) the numerical value of severity should be quickly and intuitively understandable.

The first criterion simply reflects that we are only interested in ranking abnormal traffic incidents. The second is a question of normalisation. We choose to normalise our measure so that a severity score of unity implies an incident as extreme as the worst observed in the training data. Doing so permits an intuitive understanding: a score between 0 and 1 suggests traffic is in a state no more severe than seen in the training window, whereas a score above 1 means the situation has developed more so than previously seen. We tried multiple different definitions of severity. We ultimately decided that calculating the current distance of a trajectory to the nearest point on the boundary of the typical region alone provides a useful and computationally feasible severity measure. We did this by approximating the boundary of the typical region as a polygon and calculating the distance to the nearest point - see appendix C.1. Excursions from the typical region almost always start small, reach some maximal distance and then return to the typical region after some time. A severity score based on distance to the typical region means that severity evolves in such a way that maximal severity is reached when the largest perturbation from the contour is observed, and we slowly increase to and decrease from this value in time. The duration aspect of severity is captured indirectly since excursions that reach a larger distance from the typical region are generally also of longer duration.

4.5.2 Results for severity ranking of DFTB incidents

Fig. 4.10 shows a representative example of how the dynamical severity score of incidents raised using the distance based severity method relate to travel times. Considering Fig. 4.10, we clearly see the most severe scores are attained with the largest peaks in the travel time series. Importantly, in Fig. 4.10, DFTB flags are raised, and severity computed, whenever a data-point leaves the typical region, eliminating the need to wait to decide whether or not to raise a flag, but allowing instead thresholding of incidents on the severity value.



(a) Training data.

(b) Test set 1.

(c) Test set 2.

Figure 4.10: Comparison of severity scores in relation to travel times using the training and test data. We colour the points by their severity score, normalised by the maximum value in the training window. In all cases, we see low severity scores when travel times are only marginally above the typical values, and higher scores when large deviations occur.

We note that there is a single spike (around 05-18 in Test Set 1) that has the second largest observed travel time, but does not have the second highest severity. This can occur because we do not model travel times directly in our methodology. This spike simply has a less extreme deviation in-terms of distance from the typical contour than other cases, which may have smaller travel time spikes but larger excursions from the typical region. Nevertheless, the overall conclusion is that the severity of deviations from the typical region strongly correlate to travel time spikes.

4.6 Validation & comparison to existing NTIS incidents

It is clear from section 4.5 that our methodology identifies travel time spikes though atypical fluctuations with varying severities. This is one form of validation showing that these atypical situations we would display to operators correspond to practical situations where road users are experiencing delays. However, another form of validation is to question how many atypical fluctuations in the flow-density relationship correspond to actual incident flags recorded in the data. As discussed in section 3.2.2, NTIS has multiple incident categories, each of which are practically important to traffic operators. Here, we compare how alerts one might raise at different severity thresholds using our method compare to accident, obstruction and abnormal traffic incident

flags in the data, and how alternative methodologies perform at the same task.

4.6.1 Comparison models

A number of methods for incident detection have been investigated throughout the literature, and were discussed in chapter 2. We also recall that NTIS only provides incident locations at the link-level, and hence we never know the ground truth of which two loop sensors an incident actually occurred between. As such, the classic California algorithms are not appropriate for comparison as they rely on comparing adjacent sensors and flagging when discrepancies emerge between them. However, in practice our method segments the flow-density diagram like the McMaster method, and defines a robust threshold for normality like the robust variants of the SND method. We therefore choose to compare to these two methods as closely as possible using the available link level data.

To implement the SND methodology, one isolates non-overlapping windows of speed time series, and then computes summary statistics measuring the average and spread of the data. Initially these were the mean and variance, however more robust statistics, being the median, inter-quartile range (IQR) and median absolute deviation (MAD) have since been applied. For a given window w, if we denote the mean and median speeds as μ^w_{mean} and μ^w_{median} respectively, and the standard deviation, IQR and MAD as ζ_{sd}, ζ_{IQR} and ζ_{MAD} respectively, we then compute threshold speeds for this window as

$$\begin{aligned} \tilde{s}_1 &= \min\left(s_1^{\max}, \mu^w_{\text{mean}} - c_1\zeta_{\text{sd}}\right), \\ \tilde{s}_2 &= \min\left(s_2^{\max}, \mu^w_{\text{median}} - c_2\zeta_{\text{IQR}}\right), \\ \tilde{s}_3 &= \min\left(s_3^{\max}, \mu^w_{\text{median}} - c_3\zeta_{\text{MAD}}\right). \end{aligned} \tag{4.3}$$

This approach is taken from [24], where one avoids 'swamping' of the robust statistics. As in the cited work, we set the threshold values $s_1^{\max}, s_2^{\max}$ and $s_3^{\max}$ to be 45 mph, informed by the work in [149]. To apply this methodology, one then needs to determine optimal values for either c_1, c_2 or c_3 depending on methodology choice. We choose to use windows of length 15 minutes, computing a different set of statistics for windows across a week as in [24]. A window length of 15 minutes is a reasonable and standard choice consistent with existing literature, for example [21] and [24], offering enough data in each bin to measure variation without being too fine to introduce large uncertainty in these statistics. We see consistently better performance on our dataset using robust methods, and generally a slight performance improvement when using the median and IQR so only consider the threshold defined by $\min\left(45, \mu^w_{\text{median}} - c_2\zeta_{\text{IQR}}\right)$, which we refer to as SND (Robust) from now on. After fitting, we raise alarms when the measured speed on a link falls below this threshold for at-least 3 consecutive

minutes. Note that this 45 mph value is actually converted to km/hr to align with the units of our data in implementation.

Another insightful method for comparison is a variant of the McMaster algorithm, in the sense that one takes the flow-occupancy diagram at the link level and then segments it into distinct regions. Occupancy is used here to exactly follow the work in [16]. This has more parameters than the SND methodology, as one has to specify a 'lower bound of uncongested data' (LUD), a critical occupancy and a critical flow. We consider a quadratic form for LUD as in [18], giving us 5 total parameters to fit for the model. It should be noted that the full McMaster logic, as detailed in [16], contains an initial step of determining a congested, bottle-neck or uncongested state, and then has a second step where one distinguishes between recurrent and incident congestion. As we are dealing with link-level data, it is not sensible to conduct this secondary check, however we can still compare to a method that in spirit segments the flow-occupancy diagram and determines when the link state is outside reasonable bounds. If we consider only links not subject to recurrent congestion, it is sensible to make comparisons between this model and our own. As such, one should not entirely consider this the McMaster algorithm, however it is also an adept comparison as the proposed methodology similarly attempts to segment the flow-density diagram, however by discovering the bounds on typical behaviour in an entirely data-driven way.

4.6.2 Measuring performance

To tune the comparison models and measure performance, we first isolate all aforementioned incidents on a link. We then split the data into a training set, which we calibrate our models with, and a test set which we measure performance on. By calibration, we specifically mean three things. For the DFTB model, we are aiming to choose the optimal severity threshold that will distinguish between fluctuations representing labelled incidents and those not. For the SND model, we compute the median and IQR of the data, and choose the optimal c_2 value that distinguishes between typical speed variation and decreases relating to incident flags. Finally, for the McMaster algorithm, we fit all parameters to segment the diagram in such a way that when flags are raised by the method, they correspond optimally to real incident flags. As such it is clear that for the DFTB and SND methods, we are fitting a single parameter, and for the McMaster we are fitting 5.

Since the amount of data used for training could influence model performance, for a fair comparison we used the same 3-week subset of data (April 7th 2017 to April 27th 2017) used to fit the typical behaviour contour, summarised in appendix B.4.2. To verify that training window length does not strongly

influence the results, we also fit the models using 12 weeks of data and found no significant changes. This 12 weeks of data started on September 1st 2017 and constituted the 12 week period following this. We selected this to ensure we were training on a continuous window of data, as-well as testing if providing a longer window of training data altered model performance. This subset is summarised and discussed in appendix B.4.6. To ensure that we evaluate our models on a truly representative set of data, we take our test period to be 45 weeks, taken between December 21st 2017 and ending on November 1st 2018. This provided a continuous window of unseen data that spanned a representative set of weather conditions, traffic volumes, national holidays, and typical commuting days and weekends. Further details on this subset are given in appendix B.4.7.

Given this, we then compute the standard performance metrics used throughout incident detection work. The first of these is detection rate (DR), calculated as the number of labelled incidents detected divided by the total number of labelled incidents. Another is false alarm rate (FAR), calculated as the number of times an incident flag was raised when there was no flag in the data, divided by the number of applications of the algorithm. The final one is mean time to detect (MTTD), being the average time between the start of an incident in the data and the identification of the incident by the method. We express DR and FAR as percentages from now on. As with all multi-objective optimisation problems, there is no immediately clear mathematical answer to the question of how to trade these performance measures off against each other. This depends on the application. We choose to combine DR, MTTD and FAR into the single optimisation criterion, denoted performance index (PI), that was used in [24], [150] and [151] and is written as

$$\mathrm{PI} = \left(\epsilon_{\mathrm{DR}} - \frac{\mathrm{DR}}{100}\right) \cdot \left(\frac{\mathrm{FAR}}{100} + \epsilon_{\mathrm{FAR}}\right) \cdot (\mathrm{MTTD}) \tag{4.4}$$

where ϵ_{DR} and ϵ_{FAR} are constants that avoid trivial minima of the criteria. While we recognise that there is arbitrariness in this choice, it does facilitate easy comparison to existing literature. We set $\epsilon_{\mathrm{DR}} = 1.01$ and $\epsilon_{\mathrm{FAR}} = 0.001$ as in [24]. We choose the parameters of each model that minimise Eq. (4.4), and then compare their performance to our anomaly detection methodology. In general, calibration of automated incident detection methods is considered to be a difficult task [152], yet to fit a set of models over many links, one must comprise and use some reasonable criteria as an optimisation objective.

It is important to note that with the discussed methods, one can easily envision how an operator could interpret and change the single parameters of the DFTB and SND methods. Since both have essentially one tuning parameter, the severity threshold in the DFTB case and the c parameter in the SND case,

they are entirely interpretable to an operator experienced in traffic management but not in the technical aspects of the model. An increase in either of these thresholds will allow more variation before flagging, but it is reasonably clear how each of the DR, FAR and MTTD would naturally increase or decrease when doing this. This is not true however for the McMaster algorithm, as altering a 5 parameter model and understanding how your changes will impact the performance criteria is not as simple.

4.6.3 Validation results

After choosing the parameter set for each model that has optimal PI in the training period, we apply the tuned models to the unseen test data. Results for this comparing the DFTB performance in DR, FAR and MTTD to the robust SND model are give in table 4.1, considering a variety of link lengths and locations along the M25. We have selected this subset of links as those that appear to have minimal data quality problems in terms of missing data. After presenting aggregated statistics in the final rows of table 4.1, we then consider the typical differences between the models in table 4.2, where we also test if there is statistically significant evidence that distinguishes between the two models for each performance metric. Inspecting table 4.1, we see that the mean and median DR for the DFTB method is lower than that of the robust SND methodology across the considered links. The mean and median FAR across the links is lower for the DFTB compared to the robust SND method, and we see that the DFTB results in a marginally lower mean value of MTTD, but a marginally higher median value of the same metric.

Before any further discussion of these performance scores, let us first determine which of the observed differences are statistically significant. This is to avoid basing conclusions on small differences that could have arisen by chance. Statistical significance of differences in model performance is a topic that has been discussed extensively in the machine learning literature, with [153] providing an overview of the hypothesis tests that are most appropriate for comparing performance across multiple datasets. This is further discussed in [154]. They conclude that two statistical tests are generally appropriate for model comparison: the Wilcoxon signed-rank test and the sign test. We therefore apply both of these, and show results in table 4.2. To perform a Wilcoxon signed-rank test, we first compute the paired differences in the data, and record the sign of this (+1 or -1). We then compute the absolute values of each paired difference, and rank them from smallest to largest. Our test statistic is then the sum of these ranks, multiplied by +1 or -1 depending on the sign of the paired difference. This statistic is then compared to known distributions under the null hypothesis: assuming the difference in medians

Link		DR (%)		FAR (%)		MTTD (minutes)	
Location	Length (km)	SND (Robust)	DFTB	SND (Robust)	DFTB	SND (Robust)	DFTB
East	0.7	80.39	86.28	0.94	0.97	5.71	4.66
East	1	89.13	84.78	1.99	2.25	4.34	2.59
South East	0.5	82.35	79.41	2.14	1.78	10.14	12.22
South East	0.4	83.87	80.65	2.15	1.78	7.50	9.32
South East	5.1	81.97	78.69	0.74	0.55	14.26	14.77
South	0.7	96.55	96.55	2.66	0.71	4.21	4.25
South	1.2	69.09	49.09	0.71	0.04	10.79	14.74
South West	0.9	94.27	81.25	10.06	1.35	3.54	7.03
South West	6.9	44.11	46.39	2.40	0.44	14.96	12.88
West	2.5	64.29	78.57	0.49	0.46	9.59	9.76
West	3.9	35.06	89.08	1.58	2.57	10.67	7.17
West	1.2	91.16	81.86	9.40	0.65	12.10	15.34
North West	1.9	98.13	90.65	5.32	1.59	7.41	7.10
North West	1.3	47.95	42.47	1.49	0.33	17.37	13.90
North West	0.9	54.29	48.57	0.52	0.41	17.84	13.41
North	4.5	69.14	61.71	1.03	0.74	15.29	14.12
North East	1.9	89.09	87.27	1.03	0.85	11.25	11.60
Mean		74.76	74.31	2.63	1.03	10.41	10.29
Median		81.97	80.65	1.58	0.74	10.67	11.60
Std Dev		19.58	17.43	2.91	0.73	4.54	4.14
IQR		24.84	24.56	1.47	1.14	6.85	6.80

Table 4.1: Comparison of model performance across a set of representative links of varying lengths and locations around the M25. All models are shown 3 weeks of training data. SND (Robust) is the standard normal deviate methodology using the median and inter-quartile range (IQR), as defined in Eq. (4.3). DFTB is the proposed deviation from typical behaviour methodology. The metrics compared are: DR (detection rate, given as a percentage of all labelled incidents in the data), FAR (false alarm rate, given as a percentage of the total applications of the algorithm) and MTTD (mean time to detect, given in minutes). The length of each link is given in kilometres. For reference, the links used here are numbered 0, 2, 3, 4, 7, 9, 19, 27, 28, 30, 32, 37, 41, 42, 51, 67 and 70 in table 3.1.

	DFTB DR - SND (Robust) DR	DFTB FAR - SND (Robust) FAR	DFTB MTTD - SND (Robust) MTTD
Mean	-0.45	-1.60	-0.12
Median	-3.28	-0.36	0.04
Wilcoxon signed-rank test p-value	0.170	**0.004**	0.890
Sign test p-value	0.077	**0.0127**	> 0.999

Table 4.2: Aggregated model performance summary. We consider the mean and median differences between the 3 evaluation metrics, and question if there is statistically significant evidence to suggest the two models have differing performance in each metric. This is done through both a non-parametric Wilcoxon signed-rank test and a sign test, both considering the measurements to be paired. Statistically significant values at the 5% significance level are shown in bold. Abbreviations and units are as defined in table 4.1.

between the paired samples is 0. Hence, in our context, rejecting the null hypothesis under the Wilcoxon signed rank test suggests there is statistically significant evidence that the median difference between the two models is not 0.

A sign test on the other hand does not account for ranks, and acts as more of a 'tournament' between two models. A 'match' in this tournament represents training and then testing each model on a single link. In each match the two models get the same training data and calibrate to this, then have to predict the same test data. We then only record the 'winner' of each match, asking which model had superior DR, FAR, or MTTD. If one model consistently wins when we compare the results for each match, then we suspect that the methods are not equally skilled at the particular task. Formally, the null hypothesis in such a test is that given a pair of measurements, it is equally likely that either the first or second is larger or smaller than the other. This suggests that rejecting the null hypothesis in our context means one model is consistently producing a larger DR, or lower FAR and MTTD. Further discussion of both of these tests is given in appendix C.3 for those interested, but one can also consult [153] or papers that have since used the discussed methods, such as [155] and [156] for other examples of modelling works using this approach.

The results from these tests in table 4.2 suggest there is sufficient evidence to reject the null hypothesis of the median difference between the DFTB and robust SND FAR being 0 (Wilcoxon signed-rank test). Further, there is sufficient evidence to reject the null hypothesis that the probability the FAR for the DFTB method is larger than that of the robust SND method equals 0.5 (sign test). None of the other tests in table 4.2 suggest any statistically significant results. This analysis is based on a 0.05 significance level, with a p-value falling below 0.05 indicating statistically significant results, providing evidence to reject the null hypothesis.

If we compare to the McMaster algorithm with the discussed link-level limitation, we see that it consistently has a higher MTTD compared to the DFTB model on our dataset, taking a mean of 4.5 and median of 4 minutes longer to detect incidents than the DFTB does. It also has a lower DR, 6.7% lower in mean and 2.2% lower in median. Finally, it achieves a better false alarm rate, with a mean difference of 0.19% and a median difference of 0.28%, but this is at the expense of the other two metrics, and being far harder to fit.

We also note that the FAR reported above might seem higher than in the literature. We are aware some incidents can be missed in the system, an indeed informed of this by system experts. Additionally, we have not pre-filtered the data to identify only incidents that blocked lanes as in [24] and other works. We find FAR to be consistent across models showing this is a systematic feature of the data and not the models. Our typical MTTD values are slightly larger

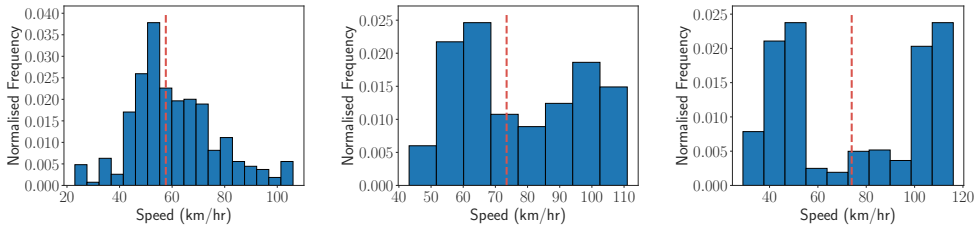
than those discussed in [24], however again this may relate to the data quality problems, and could be further impacted by the definition of deviation from profile incidents that could be minor for some time then develop into significant incidents on the link.

Finally, one has to also consider from a practical perspective, what might be important to a traffic operator who would be in an environment using any of these models? A survey of such operators was conducted and analysed in [152], and one of the principle reasons for limited integration and operational use of automated incident detection systems was difficulty of algorithm calibration. Difficulty of calibration can be thought of in two ways. The first is deciding on an exact optimisation criteria that captures what peak performance means exactly to an operator. Whilst we have used PI for this, it is an open problem to determine what might be more appropriate. However, our model, as with the SND methodology, is fully interpretable and since both have a single parameter, calibration is much more reasonable than say, a deep learning model that might have billions of parameters that an operator could not feasibly adjust without expert knowledge.

4.6.4 Analysis in bimodal speed cases

The above performance tests indicate that our algorithm is generally comparable to the well calibrated robust SND methodology in terms of performance on NTIS incidents data although there are a few instances where there are very large differences, particularly in the FAR. We now investigate why this is so. In the process, we identify situations in which our bivariate approach provides some fundamental advantages over SND. Again referring to table 4.1, we see that robust SND method attains a 10.058% FAR on the first link in the south-west investigated, which is enormous even considering the data-quality problems we might have. First, one might say this is just a 'poor' training window choice, not representative of the long-term link behaviour, however this does not appear to be the case. If we extend the training period, even providing 12 and 24 weeks of data, we still see similar false alarm rates from the SND method. Additionally, the DFTB achieves a FAR of 1.3%, at the expense of a higher MTTD and lower DR. This suggests that, on this link, some property of the speed leads to it not being well modelled by a single location and scale parameter pair. We investigate this by taking all false alarms raised by the robust SND method, and considering what time-bin each lies in through the week. We then consider the distribution of the speeds in each bin. If the SND assumptions were valid, the speeds in each bin would be unimodal, and a sensible measure of outliers is $\mu - c\sigma$. This bin could still contain extreme values or a heavy tail, which are accounted for through the use of robust

statistics. However, if the speed distribution in a bin is bimodal, then it is not appropriate to judge outliers using a single location and scale parameter. To illustrate this we refer to Fig. 4.11, where we plot the distribution of speeds in 3 time-windows. In Fig. 4.11a, we show an example where the use of the SND threshold is reasonable, whereas in Figs. 4.11b and 4.11c we show cases where there appears to be bimodality in our data.



(a) An example unimodal speed distribution.

(b) An example bimodal speed distribution (1).

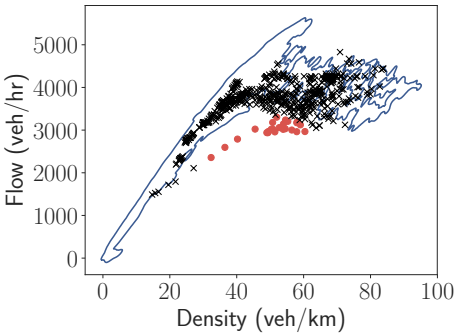
(c) An example bimodal speed distribution (2).

Figure 4.11: Distribution of speed in 3 example time-windows, each of length 15 minutes. We take the same 45 weeks of data as discussed in section 4.6.2 and plot the resulting histogram for each bin, marking the median of the data with a dashed red line.

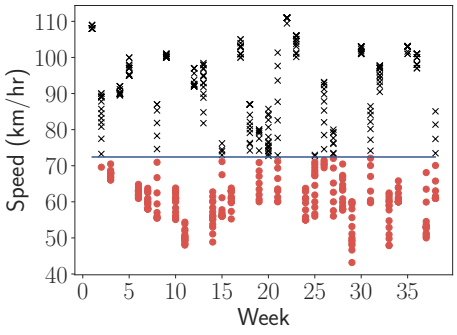
Inspecting Fig. 4.11a, we see an example time-window on the link where it is completely reasonable to judge outliers as values being some threshold away from the median of the distribution. However, in Fig. 4.11b, we see a bimodal distribution, one mode around 100 km/hr and another at 60 km/hr. Similarly, in Fig. 4.11c, we see one mode at 110 km/hr and another at 50 km/hr. This behaviour does not appear to be the result of roadworks, as we have removed points marked with these flags before plotting the distributions. With this bimodal behaviour, optimising the SND parameters essentially faces a dilemma. Since the inter-quartile range, median absolute deviation and standard deviation will all be very large, the method can choose a very small c value to still flag in these bimodal cases, but then be more prone to false alarms, or choose a larger value and almost completely ignore flagging in these bins.

We can further show this point by plotting the actual data inside each bin, along with the differing thresholds for them. To do this, we take bins (1) and (2) as shown in Figs. 4.11b and 4.11c, and collect all the reported data during them for the specific link. We then plot this data, along with the thresholds from the DFTB and robust SND methodology, with results shown in Fig. 4.12.

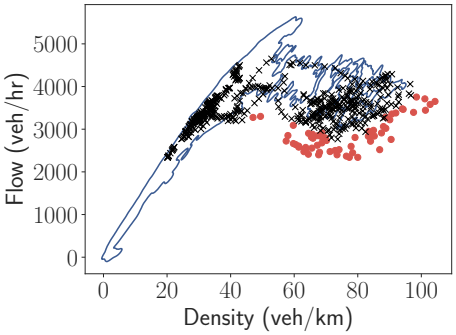
The main interpretation one should make from Fig. 4.12 is that data-points in the shown time-windows clearly fall in two regimes of the flow-density diagram. Some have densities around 20-40 veh/km and flows between 2000 and 4500 veh/hr. Additionally, there is a further cluster of data-points at high



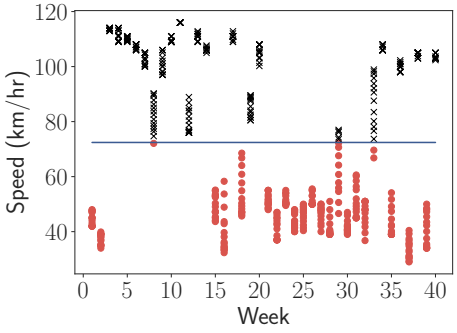
(a) Example flags for the window (1) shown in Fig. 4.11b, plotted on-top of the contour of typical behaviour used in the DFTB method.



(b) Example flags for the window (1) shown in Fig. 4.11b, plotted with the speed threshold used in the robust SND method.



(c) Example flags for the window (2) shown in Fig. 4.11c, plotted on-top of the contour of typical behaviour used in the DFTB method.



(d) Example flags for the window (2) shown in Fig. 4.11c, plotted with the speed threshold used in the robust SND method.

Figure 4.12: Example flags in a specific time-window for each method. We show the threshold for each method with a solid blue line, points flagged as 'typical' with black (×), and points flagged as atypical in red (•). Notice that points can be outside of the contour of typical behaviour, but still be typical as they are not far enough away to have a severity past some threshold. Note that these SND speed thresholds have been set based on optimal performance in the training window, and the very large spread of data in some bins appears to have set the c parameter to a very small value, which in turn has lead to quite high thresholds for these two bins, but over the entire dataset gave better performance.

density high flow points, around densities of 60-80 veh/km and flows of around 3000-4000 veh/hr. A weakness of the SND method is highlighted as it cannot capture the two regime component in the bin as the DFTB method does.

Exactly what causes this bimodality is unclear without a significant amount more data on the physical properties of the road and environment throughout the year. One could see if there is some sort of seasonal bottleneck impacting the link, for example once every 2 weeks or month where the flow becomes restricted, speed drops but not due to an incident. An example of such a

situation could be sporting or cultural events nearby that influence the traffic. If this were true, one could adapt the SND methodology to use a different number of bins and so forth, but we see that our model does not need this adaptation. By using a different definition of a typical threshold, we are able to distinguish better in these particular bimodal states than a simple univariate approach as with the SND algorithm, however all of the complexity in defining our threshold is automated, with the kernel smoothing, bandwidth selection and so forth, so the end user still receives a single parameter model that is fully interpretable.

4.6.5 Performance on high density links

In its raw form, the proposed methodology does not apply any temporal segmentation of the data. This can cause problems on links that recurrently reach very high density states and experience flow breakdown. For such links, flow-density trajectories that are highly unusual during periods of low demand may enter regions in the flow-density plane that are typical during periods of high demand. Such trajectories therefore might not get flagged as atypical despite being very unusual given the time that they occur. To illustrate this scenario, we look for the link that periodically reaches the highest densities in our data and exhibits flow breakdown as a recurrent feature, not just due to incidents. The contour of typical behaviour for this link is shown in Fig. 4.13. The point is illustrated by examining an incident trajectory in flow-density space that occurs during a typically low-density period. This is done in Fig. 4.14a, where we plot the same contour as in Fig. 4.13, but overlay a trajectory that corresponds to an incident that occurred at 12:14pm. We see that although the incident starts and ends in a low density region of typical behaviour, the traffic state whilst the incident is progressing mostly stays within a typical region at high densities. Our method in its raw form would not raise an alarm in this case until very late into the incident.

There is however a natural way to account for this by providing some level of temporal segmentation before fitting any contours of typical behaviour. A natural temporal segmentation is into 'rush periods' and 'non-rush periods' representing commuting patterns that strongly influence traffic on motorways. We show in the remainder of this section that doing so solves the identified drawback for links that regularly experience flow breakdown and high densities. We first take the same 3 weeks of data that the contour in Fig. 4.13 is fit to (period beginning April 7th, summarised in appendix B.4.2), and split the data into points that are taken from the rush period and points that are not. We then determine contours of typical behaviour for each of these datasets separately, with the non-rush period typical region shown in Fig. 4.14b and the

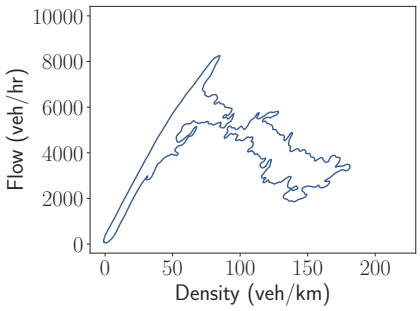
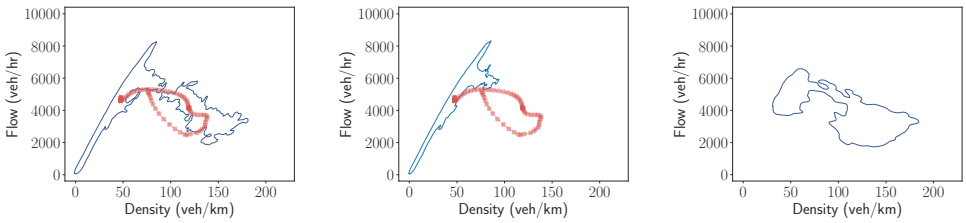


Figure 4.13: Example of the contour of typical behaviour for a link with significant flow breakdown as a recurrent feature. We see a clear aspect of the 'typical' behaviour includes a breakdown of flow as density increases, around 100 veh/km.

rush period typical region shown in Fig. 4.14c. We see from Fig. 4.14b that the



(a) Contour fit to full dataset, with an incident trajectory imposed in red.

(b) Contour fit to only data outside of the 'rush period' of operation in the UK. The same trajectory as in (a) is overlaid.

(c) Contour fit to only data during the 'rush period' of operation in the UK.

Figure 4.14: An incident trajectory overlaid on contours constructed from different subsets of data. The particular incident occurs at 12:14pm, and the flow-density data during it is plotted in Figs. 4.14a and 4.14b. The contour in Fig. 4.14c represents only the typical region during rush periods, and hence the trajectory we observe in this example is abnormal relative to what we expect the flow-density data to produce during non-rush periods.

incident trajectory deviates from the typical region during non-rush periods, and would correspond to raising a flag at this time. Clearly, an incident flag would be raised much sooner when we consider a temporal segmentation of the data rather than the full dataset as in Fig. 4.14a. This segmentation is simple and intuitive and clearly demonstrates that for links that regularly experience a large range of density values and flow breakdown, partitioning the data before contour construction is important for identifying atypical cases. Doing so, one would ideally choose two severity thresholds, as we see potential for far more significant deviations from the contour of typical behaviour outside of the rush period compared to inside it. We see examples of trajectories corresponding to incidents deviating from the typical region during rush periods in Fig. 4.15.

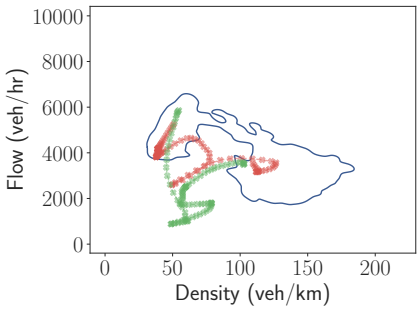


Figure 4.15: Two example trajectories corresponding to incidents that occurred during rush period, overlaid on the contour of typical behaviour fit to only data during the rush period.

When we compare the performance metrics for this link, we see that applying a simple temporal segmentation of the data into rush and non-rush periods significantly improves the balance between DR, FAR and MTTD that the method achieves, with results given in table 4.3. From table 4.3, we see

Method	DR	FAR	MTTD
SND (Robust)	72.868	6.189	8.191
DFTB	36.951	0.898	39.182
DFTB (With Temporal Segmentation)	93.023	3.300	5.847

Table 4.3: Comparison of methods on a link with a high density range. We see that the DFTB method in its raw form performs poorly on such a link, however if we segment the data into rush and non-rush periods before fitting the contour, we observe a large increase in performance and achieve better results than the SND (Robust) method across all 3 metrics.

that the robust SND method clearly outperforms the DFTB in its initial implementation. It is clear why this poor performance is observed, thinking back to the example in Fig. 4.14a. A trajectory can deviate from one region in the flow-density space that is typical, however move into another region that is still typical, however not typical at the time the incident occurs. The raw DFTB methodology does not perform any temporal segmentation and hence for extremely high density links, this can mean that incidents during non-rush periods occupy the same flow-density space as typical behaviour during rush periods. Hence, we have to either wait a very long time for the trajectory to move out of all regions of typical behaviour, or flag at very low severities, which would cause a large number of false alarms. As a result, we see that the raw DFTB method has detected very few incidents, and taken a very long time to do so, but has a very low FAR. When we instead provide an intuitive temporal segmentation before fitting the contour, we see an enormous improvement in balance of performance metrics for this link. This extra layer of information we provide the model achieves a higher DR and lower MTTD

than both the original DFTB method and the robust SND method. It also achieves a lower FAR than the robust SND method, however higher than the raw DFTB method. This has to be considered with a balance of the other variables however, as the FAR is so low in the raw DFTB case as a result of the MTTD being very high and DR being very low. Overall, it is clear that the temporal segmentation improves the overall performance of the DFTB method on links that recurrently reach very high densities. If we show the robust SND method more data, the detection rate improves to match that of the temporally segmented DFTB model, but the FAR and MTTD still remain inferior.

4.7 Summary & conclusions

In this chapter, we have explored how empirical analysis of fluctuations in the flow-density relationship can be combined with streaming NTIS data to identify significant deviations from typical behaviour of traffic flow on a road section almost in real time. Starting by considering a kernel density estimate of the raw flow-density data, we showed how to define a contour of typical behaviour in which standard traffic behaviour lies, and investigated the deviations from this contour. We showed how the time-scales of these deviations can be used to infer the presence of significantly atypical behaviour on a link, and hence identify problematic traffic states for operators to address. To avoid creating a delay in raising incidents, we looked at methods to automatically and in real time assess the severity of a deviation, finding the distance of a flow-density data-point from the typical behaviour contour offered a severity measure that directly correlated with the spikes in travel time series. Our methods indirectly identify spikes in the travel time series and incident flags raised by the NTIS system.

This approach may present some practical advantages over existing methods of incident detection. One advantage is that our approach is entirely model free in the sense that the boundary of the region of typical behaviour for each link is autonomously learned entirely from data. There is no need for complicated time series analysis of travel times or calibration of travel time profiles at the level of individual links in order to identify incidents. Furthermore, by relearning the boundary of the typical region at periodic intervals, the system could be made self-adaptive to automatically capture long term changes in the behaviour of traffic on a link. A second advantage is that identification of incidents and assignment of a severity score can be done almost in real time since the process operates directly on the speed and flow data. From a practical perspective, we receive data from the NTIS system on a minute-scale, and the computation of a severity takes significantly less than this.

We envision that our methodology for detecting anomalies on the network and raising flags when they are sufficiently severe could provide the basis for a practically useful decision support system that can help operators better manage and monitor the road network. Before discussing this further, it is useful to recall from chapter 3 that in the existing operational infrastructure, human operators view a display summarising the current state of the network and can filter this to different locations. Along with this display, current incidents appear as a list on the screen. Locations where the criteria for a deviation from profile event are met enter this list, but have to be manually accepted by operators before being logged in NTIS. One could extend this existing infrastructure to incorporate our DFTB flags in a reasonably simple way. We have seen two important aspects relating to this throughout our methodology. Firstly, we have observed throughout this chapter that DFTB flags have a significant amount of commonality with non-recurrent congestion flags in the NTIS data, hence capturing many observed incidents already logged in the data. Secondly, we are able to rank anomalies by their severity, measured relative to the typical variation observed on each individual link. As a result, we imagine integrating our DFTB flags into the existing operator display, raising flags when a particular severity threshold is reached. A key point in integration of our methodology would be to use the ranks to determine what order to display these flags to operators, so the operator sees the most severe situations first, and less severe anomalies can be addressed later if time allows. Computationally, our methodology does not require any specialist hardware to be used when run in real-time, further highlighting potential for integration.

The data available to us has required some compromises during validation that one would ideally investigate further in the future. The first of these is a somewhat uncertain 'ground truth', meaning the incident labels NTIS provides that we have measured our DR, FAR and MTTD on in section 4.6. We are aware that both our dataset and data from the wider literature does not provide perfect incident labels, as start times may be delayed when entering incidents into the system and incidents may not be recorded at all if operators are particularly busy or their attention is required elsewhere. Collecting a dataset that does not suffer from these data quality issues would be an expensive and time consuming task, however this would be required to truly measure the effectiveness of any incident detection methodology. We therefore believe that such a dataset should be attained, potentially on a smaller time and space domain if required for practically, and the validation completed here repeated on this dataset before our method could be implemented into the NTIS system and be used by operators. Doing so would remove a significant element of uncertainty in the modelling and conclusions we have drawn, further justifying its use in a practical setting and building trust with operators.

It should also be noted that when validating our methodology in section 4.6, we have selected a range of links of different lengths, from different locations around the M25 and that experience different traffic volumes. However, to avoid excessive inspection of the data, we have used a subset of links that have minimal data quality issues, regarding missing sensor readings and clearly missing event labels. Whilst we are confident we have tested our behaviour on a wide range of links and long window of data, far more than is typical in the literature, we also recognise that in selecting this subset of links, we have preformed a process similar to complete case analysis. If all missing values are missing completely at random (recall MCAR), this would not cause significant bias, however as we have discussed, there may be a number of reasons our data possesses missing incident labels and sensor data. Examples of these include malfunctioning sensors on sections of road that are recurrently extremely busy, where maintenance on these sensors cannot be justified due to the distribution it would cause. Further, operators may themselves be more likely to focus their attention on a subset of links they know are structurally important to the transport network, and miss more event flags on other links. For both of these reasons, the validation we have performed may be biased, and in the future significant efforts should be made to verify if the conclusions made in our work remain the same when the methods are applied on datasets without these issues. As discussed, generating such datasets is costly and may prove difficult, particularly for the durations we have considered. However, subject to the constraints our data places on our work, we believe our analysis has covered a range of different links and locations, on time-scales that allow one to observe a significant amount of heterogeneity in behaviour that might occur across the M25.

If one were to use our system to augment the information already displayed to operators as discussed, this would introduce a new incident category in NTIS (DFTB), however we believe the visualisations given throughout this chapter clearly define why such an event would be raised, so this would not be a major hurdle for operators. Another potential application would be to use our methodology as an automated filter that determines inputs to a more complex and intensive simulation model. Given an incident occurs on the network, it may be of interest to model directly what the impact on the surrounding infrastructure will be, potentially at the microscopic scale. One could build a model for this purpose at the level of individual vehicles, and calibrate it to NTIS data. However, such a model is likely to be very computationally intensive, and not realistic to run for every incident that occurs on the network. Instead, only incidents that reach a particular severity could be input to such a model, which might provide useful predictions for the short term future of the network in the cases where the output would be most helpful for operators.

The results from this expensive simulation model could then be presented to operators to identify the optimal control actions to take in managing an incident.

We have already discussed that the uncertainty in the incident labels lead to uncertainty in the DR, FAR and MTTD for both our proposed methodology and the comparison models, and how we would ideally want to overcome this. A further level of uncertainty is introduced through the missing flow values on links 60, 65, 66, 68 and 69 (recall table 3.1). As we cannot evaluate our methodology without speed and flow data, we are unable to perform our analysis, fitting or validation on these links on any subset of our data. We are therefore unable to say with absolute certainty that these links will not exhibit some unseen density-flow behaviour that degrades our models performance. However, the same is true for any comparison methods that use both variables, and we have considered a range of different links when testing and validating our methodology. As a result, we concede that our conclusions are subject to some bias and would ideally like to repeat our analysis on these links when data is available and before our methodology is used in a real-world scenario.

A final point is that if our method was used in the control room, but a particular set of sensors did not report flow data, we would be unable to raise DFTB incidents on these links. This could lead to situations where control room operators miss significantly atypical situations on particular stretches of road, and they cause increased levels of congestion as they are not dealt with for long periods. We therefore believe that even after all of the other uncertainties highlighted have been overcome, operators should still be given a list of links that do not report data, and they can then check these manually either through camera feeds or other means. The operators face a similar problem if links report missing travel times in the current implementation of NTIS, so they will already have some experience in working with missing data on sections of the network.

Chapter 5

A non-parametric Hawkes process model of primary and secondary incidents

In this chapter, a self-exciting spatio-temporal point process is fitted to NTIS incident data to model the rates of primary and secondary incidents observed in the data. This process uses a background component to represent primary incidents, and a self-exciting component to represent secondary incidents. The background consists of periodic daily and weekly components, a spatial component and a long-term trend. The self-exciting components are decaying, unidirectional functions of space and time. These components are determined via kernel smoothing and likelihood estimation. Temporally, the background is stable across seasons with a daily double peak structure reflecting commuting patterns. Spatially, there are two peaks in intensity, one of which becomes more pronounced during the study period. Self-excitation accounts for 6-7% of the data with associated time and length scales around 100 minutes and 1 kilometre respectively. In-sample and out-of-sample validation are performed to assess the model fit. When we restrict the data to incidents that resulted in large speed drops on the network, the results remain coherent.

5.1 Introduction & problem relevance

The United Kingdom has one of the lowest per-capita death rates from traffic accidents in the world, estimated by [157] at 3.1 per 100,000 of population in 2016. Nevertheless 1782 deaths and 25,484 serious injuries resulted from accidents on UK roads in 2018 [158]. Aside from the direct human cost of serious accidents, indirect economic costs result even from relatively minor incidents. This is because crashes, collisions and breakdowns can cause severe congestion leading to significant drops in the efficiency of the road transport

network. For these reasons, there is an imperative to further reduce the incident rate on UK roads. However traffic incidents are rare in absolute terms and are not distributed uniformly across the network. Further rate reductions are therefore likely to require targeted interventions. Targeted interventions might try to to improve safety at specific locations where the congestion risk is known to be high compared to the baseline or might try to mitigate against particular mechanisms that are known to account for a significant proportion of incidents. Infrastructure modifications to improve the safety of accident-prone junctions is an example of the first type of intervention. The deployment of MIDAS to reduce the number of secondary incidents on motorways is an example of the second. Secondary incidents occur when when a driver fails to react appropriately to the disruption caused by an existing incident leading to a subsequent incident upstream of the first one.

Throughout this chapter, we use data from NTIS to model the distribution of motorway incidents as a spatio-temporal Hawkes process. The objectives of the study are two-fold. The first is to quantify how incident risk on the M25 varies in space and time relative to the baseline. The second is to use the self-excitation component of a Hawkes process to quantify the likely contribution of secondary incidents to the observed totals. We would like this model to be helpful in addressing the question of how best to target interventions when the baseline incident rate is low in absolute terms. We therefore perform extensive in-sample and out-of-sample validation to verify the models performance on seen and unseen data.

5.2 Data details

Throughout this chapter, we consider all accident and obstruction incidents in the data between September 1st 2017 and September 30th 2018, defining incident locations through the localisation methodology discussed in section 3.2.3 and times as given by NTIS. This subset of data was selected as we obtained both the link and loop level data during this period, allowing us to perform the localisation of NTIS incidents. We did not extract the loop data before this time-period, and hence could not perform the same localisation. This subset is still sufficiently large to capture long-term behaviour across the M25, and as we will see throughout this chapter, identify natural features of traffic behaviour as-well as novel insights into incidents on the studied domain. Summary statistics for this subset of data are given in appendix B.4.8.

5.3 Methodology

5.3.1 Model formulation

Our objective is to model the number of incidents observed between any two positions x_1 and x_2 in any time interval between t_1 and t_2. We denote this quantity by $N_{[x_1,x_2),[t_1,t_2)}$. Since we know that incidents cluster in both space and time, the simplest model is a non-homogeneous Poisson point process. This model is specified by an underlying intensity function, $\lambda(x,t)$, which is the local incident probability per unit length per unit time. It is then assumed that for any intervals, $[x_1,x_2)$ and $[t_1,t_2)$, $N_{[x_1,x_2),[t_1,t_2)}$ has a Poisson distribution, written as

$$\mathbb{P}(N_{[x_1,x_2),[t_1,t_2)} = n) = \frac{\Lambda^n_{[x_1,x_2),[t_1,t_2)}}{n!} e^{-\Lambda_{[x_1,x_2),[t_1,t_2)}}, \tag{5.1}$$

where

$$\Lambda_{[x_1,x_2),[t_1,t_2)} = \int_{t_1}^{t_2}\int_{x_1}^{x_2} \lambda(x,t)dx\,dt. \tag{5.2}$$

It is natural to assume that the intensity is multiplicatively decomposable, giving

$$\lambda(t,x) = \mu_0\,\mu_d(t)\,\mu_w(t)\,\mu_t(t)\,\mu_s(x). \tag{5.3}$$

Here μ_0 is a uniform background intensity which has units of incidents per unit time per unit length. This uniform background is then modulated by the functions $\mu_d(t)$, $\mu_w(t)$, $\mu_t(t)$ and $\mu_s(x)$ to capture spatio-temporal variation. The spatial modulation, $\mu_s(x)$, accounts for the fact that different locations will naturally have differing rates of incidents, for example junctions with low visibility having higher rates than straight, simple sections of road. The temporal modulation, $\mu_d(t)\,\mu_w(t)\,\mu_t(t)$ consists of three components: $\mu_d(t)$ represents daily variation, $\mu_w(t)$ represents weekly variation and $\mu_t(t)$ represents any long term trend that may be present. Daily and weekly seasonality is a ubiquitous feature of traffic data reflecting daily 'rush-period' commuting patterns, and weekly differences between the 5-day working week to the 2-day weekend. Annual seasonality may also be present but since our data spans 13 months, any such variation is captured by the trend.

Inhomogeneous point processes describe clustering solely in terms of variations in the intensity function. A more sophisticated model is a self-exciting point process, known as a Hawkes process, to capture the distinction between primary and secondary incidents. Self-excitation means that when an incident occurs, the probability of observing a subsequent incident nearby increases. A second term depending on the previous incidents is added to Eq. (5.3) to give

what is called the *conditional* intensity function, taking the form

$$\lambda(t,x) = \mu_0\,\mu_d(t)\,\mu_w(t)\,\mu_t(t)\,\mu_s(x) + A\sum_{\substack{t_i<t\\ x_i>x}} g(t-t_i)h(x-x_i), \tag{5.4}$$

where A is the triggering rate, g and h are triggering functions that describe how the triggering mechanism decays in time and space respectively and (t_i, x_i) are the times and locations of the observed incidents. The word 'conditional' here reflects the dependence of the intensity on the realisation of the process.

Our methodology is based on the work of [124], where it is shown how to construct the conditional intensity of a Hawkes process in a non-parametric way by applying kernel smoothing to the observed data on crime. We modify this model to make it applicable to traffic data. We begin in the following sections by detailing the estimators of model components and details of the fitting algorithm, following closely [124]. We then discuss the three main changes made to the original work. Firstly, the spatial triggering mechanism is one-dimensional and unidirectional: secondary incidents cannot occur downstream of the primary incident. Secondly we enforce monotonicity of the triggering functions, g and h, in Eq. (5.4) as a constraint to help with identifiability. Thirdly, we apply boundary correction to the kernel density estimates to reduce bias.

Derivation setup

Throughout the following sections, we will denote a kernel function with bandwidth ω as $k_\omega(x)$, and let m_d, m_w be the number of time-points (in our case minutes) in a day and week respectively. Then, as discussed in [124] and [125], for a spatio-temporal point process N_p with conditional intensity function $\lambda(t,x)$ and a predictable process $f(t,x) \geq 0$, over a time domain $[T_1, T_2]$ and space domain S, we can write

$$\mathbb{E}\left[\int_{[T_1,T_2]\times S} f(t,x)dN_p(dt\times dx)\right] = \mathbb{E}\left[\int_{T_1}^{T_2}\int_S f(t,x)\lambda(t,x)dxdt\right]. \tag{5.5}$$

Using this, we then outline the methodology provided in [124] to construct first the background components of our desired process then the triggering components. To construct each component of our model, we have a dataset of incidents, each with a time of occurrence t_i and location x_i, with $i \in \{1, 2, ..., N\}$, and our data collected through a time-space range $[0, T] \times [0, X]$. Note also that throughout, $\mu_d(t)$, $\mu_w(t)$, $\mu_t(t)$ and $\mu_s(x)$ have average value 1, allowing μ_0 to represent the background rate and each of these functions to modulate it, and for ease of comparison across components. The triggering functions $g(t)$ and $h(x)$ are probability density functions, as in [124].

Periodic temporal background components

Our model contains two periodic background components, capturing daily and weekly variation in the background. These are constructed very similarly, so we only detail the daily component here. To begin, we first denote

$$w^{(d)}(t,x) = \frac{\mu_d(t)\mu_s(x)}{\lambda(t,x)}, \tag{5.6}$$

and substitute this into Eq. (5.5), giving

$$\begin{aligned}
&\sum_{i=1}^{N} w^{(d)}(t_i,x_i)\mathbb{1}_{t_i\in\cup_{k\in\mathbb{Z}}[t+km_d-\Delta t,t+km_d+\Delta t]} \\
&\approx \int_0^T\int_0^X w^{(d)}(\tau,\chi)\lambda(\tau,\chi)\mathbb{1}_{\tau\in\cup_{k\in\mathbb{Z}}[t+km_d-\Delta t,t+km_d+\Delta t]}d\chi d\tau \\
&= \int_0^T\int_0^X \frac{\mu_d(\tau)\mu_s(\chi)}{\lambda(\tau,\chi)}\lambda(\tau,\chi)\mathbb{1}_{\tau\in\cup_{k\in\mathbb{Z}}[t+km_d-\Delta t,t+km_d+\Delta t]}d\chi d\tau \\
&= \left[\int_0^T \mu_d(\tau)\mathbb{1}_{\tau\in\cup_{k\in\mathbb{Z}}[t+km_d-\Delta t,t+km_d+\Delta t]}d\tau\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&= \left[\sum_{\gamma=0}^{\lfloor\frac{T}{m_d}\rfloor}\int_0^{m_d}\mu_d(\tau)\mathbb{1}_{\tau\in[t-\Delta t,t+\Delta t]}d\tau\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&= \left[\sum_{\gamma=0}^{\lfloor\frac{T}{m_d}\rfloor}\int_{t-\Delta t}^{t+\Delta t}\mu_d(\tau)d\tau\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&\approx \left[\sum_{\gamma=0}^{\lfloor\frac{T}{m_d}\rfloor} 2\Delta t\mu_d(t)\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&\propto \mu_d(t).
\end{aligned} \tag{5.7}$$

Here, we have the indicator function $\mathbb{1}_{t_i\in\cup_{k\in\mathbb{Z}}[t+km_d-\Delta t,t+km_d+\Delta t]}$ incorporating a data-point into the estimate if it lies in some interval $[t-\Delta t, t+\Delta t]$, on any day in the dataset, where $t\in[0,m_d)$ and Δt is a small positive value. From Eq. (5.7), it follows that

$$\hat{\mu}_d(t) \propto \sum_{i=1}^{N} w_i^{(d)}\mathbb{1}_{t_i\in\cup_{k\in\mathbb{Z}}[t+km_d-\Delta t,t+km_d+\Delta t]}, \tag{5.8}$$

$$w_i^{(d)} = \frac{\mu_d(t_i)\mu_s(x_i)}{\lambda(t_i,x_i)}. \tag{5.9}$$

We can simplify this expression by introducing $\tilde{t}_i = t_i - m_d\left\lfloor\frac{t_i}{m_d}\right\rfloor$, $\forall i \in \{1,2,\ldots,N\}$, mapping the raw incident times from the domain $[0,T]$ onto the

domain $[0, m_d)$. Eq. (5.8) then becomes

$$\hat{\mu}_d(t) \propto \sum_{i=1}^{N} w_i^{(d)} \mathbb{1}_{\tilde{t}_i \in [t-\Delta t, t+\Delta t]}. \tag{5.10}$$

In Eq. (5.10), we essentially have a histogram estimate of $\mu_d(t)$, and it is natural to considering smoothing such an estimate as in [124]. We can again turn to kernel functions to do this, using a bandwidth ω and Gaussian kernel of the form

$$k_\omega(x) = \frac{1}{\sqrt{2\pi}\omega} e^{-\frac{x^2}{2\omega^2}}. \tag{5.11}$$

Applying this smoothing, we then attain

$$\hat{\mu}_d(t) \propto \sum_{i=1}^{N} w_i^{(d)} \frac{k_{\omega_d}(t - \tilde{t}_i)}{\int_0^{m_d} k_{\omega_d}(\tau - \tilde{t}_i) d\tau} \tag{5.12}$$

where ω_d is the smoothing bandwidth specific to the daily background component, and we have prevented the 'leaking of mass' problem by ensuring that each kernel function is normalised such that its integral is 1 over the domain in question. We have constructed an estimate for the daily periodic background component by smoothing the observed incident times mapped onto the domain $[0, m_d)$, however one subtly remains.

The smoothed contribution of the data-point t_i to the function $\mu_d(t)$ is $k_{\omega_d}\left(t - \left[t_i - m_d \left\lfloor \frac{t_i}{m_d} \right\rfloor\right]\right)$. However, since the domain is periodic, we account for the fact that some incident t_i may also have non-zero contribution from its location on the previous and following day. In the context of a periodic temporal function, we would consider an incident before our domain as $t_i - m_d \left\lfloor \frac{t_i}{m_d} \right\rfloor - m_d$ and after our domain as $t_i - m_d \left\lfloor \frac{t_i}{m_d} \right\rfloor + m_d$. We denote the contributions of these components as $k_{\text{daily}}^{\text{lower}} = k_{\omega_d}\left(t - \left[t_i - m_d \left\lfloor \frac{t_i}{m_d} \right\rfloor - m_d\right]\right)$ and $k_{\text{daily}}^{\text{upper}} = k_{\omega_d}\left(t - \left[t_i - m_d \left\lfloor \frac{t_i}{m_d} \right\rfloor + m_d\right]\right)$. We can then incorporate these into the estimate as

$$\hat{\mu}_d(t) \propto \sum_{i=1}^{N} w_i^{(d)} \frac{k_{\text{daily}}^{\text{lower}} + k_{\omega_d}\left(t - \left[t_i - m_d \left\lfloor \frac{t_i}{m_d} \right\rfloor\right]\right) + k_{\text{daily}}^{\text{upper}}}{\int_{-m_d}^{2m_d} k_{\omega_d}(\tau - \tilde{t}_i) d\tau}. \tag{5.13}$$

In Eq. (5.13) we have normalised the kernel over the entire domain, including the extended points. Notice here that we have approximated the function over the domain $[-m_d, 2m_d]$ and then we retain only the chunk in between 0 and m_d for $\mu_d(t)$, meaning we do not need to provide boundary correction as there is no boundary. One could actually incorporate an infinite sum over all past and future periods, however in practice the decay of a Gaussian kernel leads to it being sensible to truncate this sum to just these terms. An explicit example

comparing estimates with and without this correction incorporated can be found in appendix D.3.

Similarly, one can write the weekly periodic component as

$$\hat{\mu}_w(t) \propto \sum_{i=1}^{N} w_i^{(w)} \frac{k_{\text{weekly}}^{\text{lower}} + k_{\omega_w}\left(t - \left[t_i - m_w \left\lfloor \frac{t_i}{m_w} \right\rfloor\right]\right) + k_{\text{weekly}}^{\text{upper}}}{\int_{-m_w}^{2m_w} k_{\omega_w}(\tau - \tilde{t}_i)d\tau} \tag{5.14}$$

with

$$w_i^{(w)} = \frac{\mu_w(t_i)\mu_s(x_i)}{\lambda(t_i, x_i)} \tag{5.15}$$

and $k_{\text{weekly}}^{\text{lower}}, k_{\text{weekly}}^{\text{upper}}$ are defined as in the daily case but considering the number of points in a week and weekly bandwidth choice. As before, ω_w is the smoothing bandwidth specific to the weekly background component, and in this case $\tilde{t}_i$ maps to a one week domain.

Temporal trend background component

Definition of the temporal trend component follows similarly to the previously discussed components. One begins by specifying

$$w^{(t)}(t, x) = \frac{\mu_t(t)\mu_s(x)}{\lambda(t, x)}. \tag{5.16}$$

Substituting this into Eq. (5.5) yields

$$\begin{aligned}
&\sum_{i=1}^{N} w^{(t)}(t_i, x_i)\mathbb{1}_{t_i \in [t-\Delta t, t+\Delta t]} \\
&\approx \int_0^T \int_0^X w^{(t)}(\tau, \chi)\lambda(\tau, \chi)\mathbb{1}_{\tau \in [t-\Delta t, t+\Delta t]} d\chi d\tau \\
&= \left[\int_0^T \mu_t(\tau)\mathbb{1}_{\tau \in [t-\Delta t, t+\Delta t]} d\tau\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&= \left[\int_{t-\Delta t}^{t+\Delta t} \mu_t(\tau)d\tau\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&\approx [2\Delta t \mu_t(t)]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&\propto \mu_t(t).
\end{aligned} \tag{5.17}$$

Again, one can then write a smoothed estimate of the trend component as

$$\hat{\mu}_t(t) \propto \sum_{i=1}^{N} w_i^{(t)} \frac{k_{\omega_t}(t - t_i)}{\int_0^T k_{\omega_t}(\tau - t_i)d\tau} \tag{5.18}$$

$$w_i^{(t)} = \frac{\mu_t(t_i)\mu_s(x_i)}{\lambda(t_i, x_i)}, \tag{5.19}$$

where ω_t is the smoothing bandwidth for the trend component.

Unlike the periodic components, the trend component is restricted onto the domain $[0, T]$. As such, the smoothed estimate is truncated both at 0 and T. It is well known that, on a truncated domain, kernel estimates are biased near the boundary, with discussion of this in [159] and references within. There are many ways to account for this throughout the literature, with a simple method being to 'mirror' the data, supposing we have extra data-points. If we have our 'real' data-points $X_1, X_2, \dots X_N$ and wish to truncate our domain at 0, we introduce extra points $-X_1, -X_2, \cdots - X_N$. The smoothed estimate $\hat{\nu}(x)$, using a bandwidth ω, is then

$$\hat{\nu}(x) = \frac{1}{N} \sum_{i=1}^{N} \left[k_\omega \left(x - X_i \right) + k_\omega \left(x + X_i \right) \right]. \tag{5.20}$$

We apply such a correction, both for the truncation at 0 and T for the trend component.

Spatial background component

The spatial background component can be constructed much like the temporal background components. First, we write

$$\psi(t, x) = \frac{\mu_0 \mu_d(t) \mu_w(t) \mu_t(t) \mu_s(x)}{\lambda(t, x)}, \tag{5.21}$$

and then substitute this into Eq. (5.5), giving

$$\begin{aligned}
&\sum_{i=1}^{N} \psi(t_i, x_i) \mathbb{1}_{x_i \in [x - \Delta x, x + \Delta x]} \\
&\approx \int_0^T \int_0^X \psi(\tau, \chi) \lambda(\tau, \chi) \mathbb{1}_{\chi \in [x - \Delta x, x + \Delta x]} d\chi d\tau \\
&= \int_0^T \int_0^X \frac{\mu_0 \mu_d(\tau) \mu_w(\tau) \mu_t(\tau) \mu_s(\chi)}{\lambda(\tau, \chi)} \lambda(\tau, \chi) \mathbb{1}_{\chi \in [x - \Delta x, x + \Delta x]} d\chi d\tau \\
&= \mu_0 \left[\int_0^T \mu_d(\tau) \mu_w(\tau) \mu_t(\tau) dt \right] \left[\int_{x - \Delta x}^{x + \Delta x} \mu_s(\chi) d\chi \right] \\
&\approx \mu_0 \left[\int_0^T \mu_d(\tau) \mu_w(\tau) \mu_t(\tau) d\tau \right] 2 \mu_s(x) \Delta x \\
&\propto \mu_s(x)
\end{aligned} \tag{5.22}$$

with Δx being a small positive value. This suggests that we can write the smoothed estimate of the spatial background component as

$$\hat{\mu}_s(x) \propto \sum_{i=1}^{N} \psi_i \frac{k_{\omega_s}(x - x_i)}{\int_0^X k_{\omega_s}(\chi - x_i) d\chi}, \tag{5.23}$$

$$\psi_i = \frac{\mu_0 \mu_d(t_i)\mu_w(t_i)\mu_t(t_i)\mu_s(x_i)}{\lambda(t_i, x_i)}, \tag{5.24}$$

where ω_s is the smoothing bandwidth specific to the spatial background component.

We make a further modelling assumption that the spatial background is periodic, as the M25 is an almost continuous ring around London. Recall from Fig. 3.1 that a small section to the east of the M25 is not a motorway and reports no data, but it is negligible in comparison to the wider motorway, so assuming a spatial background on a ring is reasonable. We therefore incorporate the same correction to this estimate as we have with the periodic temporal components, including potential non-negligible contributions from smoothed data-points one period before and after the domain in question.

Triggering components

To determine our triggering functions in an entirely data-driven way, we first consider two data points $\left(\tau^{(1)}, \chi^{(1)}\right)$ and $\left(\tau^{(2)}, \chi^{(2)}\right)$. We then define $\rho(\tau^{(1)}, \chi^{(1)}, \tau^{(2)}, \chi^{(2)})$ as

$$\rho\left(\tau^{(1)}, \chi^{(1)}, \tau^{(2)}, \chi^{(2)}\right) = \begin{cases} \frac{Ag(\tau^{(2)}-\tau^{(1)})h(\chi^{(2)}-\chi^{(1)})}{\lambda(\tau^{(2)},\chi^{(2)})}, & \text{if } \tau^{(1)} < \tau^{(2)} \text{and } \chi^{(1)} > \chi^{(2)} \\ 0, & \text{otherwise} \end{cases}. \tag{5.25}$$

This is simply stating the fraction of total intensity at the point $\left(\tau^{(2)}, \chi^{(2)}\right)$ that is a result of another incident at $\left(\tau^{(1)}, \chi^{(1)}\right)$. This also specifically incorporates the idea of unidirectional triggering. If we apply this to Eq. (5.5), letting $f(\tau, \chi) = \rho(t_i, x_i, \tau, \chi)\mathbb{1}_{\tau - t_i \in [t-\Delta t, t+\Delta t]}$, then we attain

$$\begin{aligned} &\sum_j \rho(t_i, x_i, \tau_j, \chi_j)\mathbb{1}_{\tau_j - t_i \in [t-\Delta t, t+\Delta t]} \\ &\approx \int_0^T \int_0^X \rho(t_i, x_i, \tau, \chi)\mathbb{1}_{\tau - t_i \in [t-\Delta t, t+\Delta t]}\lambda(\tau, \chi)d\chi d\tau \\ &= A\int_{t_i}^T \int_0^{x_i} \frac{g(\tau - t_i)h(\chi - x_i)}{\lambda(\tau, \chi)}\mathbb{1}_{\tau - t_i \in [t-\Delta t, t+\Delta t]}\lambda(\tau, \chi)d\chi d\tau \\ &= A\left[\int_{t_i}^T g(\tau - t_i)\mathbb{1}_{\tau - t_i \in [t-\Delta t, t+\Delta t]}d\tau\right]\left[\int_0^{x_i} h(\chi - x_i)d\chi\right]. \end{aligned} \tag{5.26}$$

If we then let $s = \tau - t_i$ we have

$$\begin{aligned}
&\sum_j \rho(t_i, x_i, \tau_j, \chi_j) \mathbb{1}_{\tau_j - t_i \in [t-\Delta t, t+\Delta t]} \\
&\approx A \left[\int_0^{T-t_i} g(s) \mathbb{1}_{s \in [t-\Delta t, t+\Delta t]} ds \right] \left[\int_0^{x_i} h(\chi - x_i) d\chi \right] \\
&= A \left[\int_{t-\Delta t}^{t+\Delta t} g(s) ds \right] \left[\int_0^{x_i} h(\chi - x_i) d\chi \right] \\
&\approx 2A\Delta t g(t) \left[\int_0^{x_i} h(\chi - x_i) d\chi \right] \\
&\propto g(t).
\end{aligned} \tag{5.27}$$

As a result

$$g(t) \propto \sum_i \sum_j \rho(t_i, x_i, t_j, x_j) \mathbb{1}_{t_j - t_i \in [t-\Delta t, t+\Delta t]} \tag{5.28}$$

and hence

$$\hat{g}(t) \propto \sum_{(i,j)} \rho_{i,j} \mathbb{1}_{t_j - t_i \in [t-\Delta t, t+\Delta t]} \tag{5.29}$$

with

$$\rho_{i,j} = \frac{A g(t_j - t_i) h(x_j - x_i)}{\lambda(t_j, x_j)} \; \forall \, (i,j) \text{ s.t. } t_i < t_j \text{ and } x_i > x_j. \tag{5.30}$$

As before, the estimator in Eq. (5.29) can be smoothed to give

$$\hat{g}(t) \propto \frac{\sum\limits_{(i,j)} \frac{\rho_{i,j} k_{\omega_g}(t - t_j + t_i)}{\int_0^{T-t_i} k_{\omega_g}(\tau - t_j + t_i) d\tau}}{\sum\limits_{i=1}^{N} \mathbb{1}_{t_i + t \leq T}}. \tag{5.31}$$

In Eq. (5.31), ω_g represents the smoothing bandwidth of the temporal triggering component, and we have normalised each kernel density estimate by the remaining time, and then finally dividing by the term $\sum_{i=1}^{N} \mathbb{1}_{t_i + t \leq T}$, which corrects for repetitions as detailed in [124].

In the same way, we can write the spatial triggering function, with smoothing bandwidth ω_h as

$$\hat{h}(x) \propto \frac{\sum\limits_{(i,j)} \frac{\rho_{i,j} k_{\omega_h}(x - x_j + x_i)}{\int_0^{x_i} k_{\omega_h}(\chi - x_j + x_i) d\chi}}{\sum\limits_{i=1}^{N} \mathbb{1}_{x_i + x \geq 0}}. \tag{5.32}$$

5.3.2 Determining background and triggering coefficients

In the model specification, we have two coefficients A and μ_0 that specify how much or little of the triggering and background component respectively enters

the conditional intensity function. In-particular, it should be noted that A can be interpreted as the proportion of the impact of the triggering functions on the total intensity, so a high A value suggests data is dominated by triggering and the converse for a small value. These parameters can be determined through maximum likelihood estimation. The log-likelihood function for a spatial-temporal point process model with triggering is given by

$$\log(L) = \sum_{i=1}^{N} \log\left(\lambda(t_i, x_i)\right) - \int_0^T \int_0^X \lambda(t, x) dx dt. \tag{5.33}$$

Using Eq. (5.4), we then see

$$\begin{aligned}
\log(L) = & \sum_{i=1}^{N} \log\left(\mu_0\mu_d(t)\mu_w(t)\mu_t(t)\mu_s(x) + A\sum_{\substack{t_i<t \\ x_i>x}} g(t-t_i)h(x-x_i)\right) - \\
& \int_0^T \int_0^X \mu_0\mu_d(t)\mu_w(t)\mu_t(t)\mu_s(x) + A\sum_{\substack{t_i<t \\ x_i>x}} g(t-t_i)h(x-x_i)dxdt \\
= & \sum_{i=1}^{N} \log\left(\mu_0\mu_d(t)\mu_w(t)\mu_t(t)\mu_s(x) + A\sum_{\substack{t_i<t \\ x_i>x}} g(t-t_i)h(x-x_i)\right) - \\
& \int_0^T \int_0^X \mu_0\mu_d(t)\mu_w(t)\mu_t(t)\mu_s(x)dxdt - \\
& \int_0^T \int_0^X A\sum_{\substack{t_i<t \\ x_i>x}} g(t-t_i)h(x-x_i)dxdt \\
= & \sum_{i=1}^{N} \log\left(\mu_0\mu_d(t)\mu_w(t)\mu_t(t)\mu_s(x) + A\sum_{\substack{t_i<t \\ x_i>x}} g(t-t_i)h(x-x_i)\right) - \\
& \mu_0 \int_0^T \int_0^X \mu_d(t)\mu_w(t)\mu_t(t)\mu_s(x)dxdt - \\
& A\sum_{i=1}^{N} \int_{t_i}^T \int_0^{x_i} g(t-t_i)h(x-x_i)dxdt.
\end{aligned} \tag{5.34}$$

We then denote

$$\begin{aligned}
U &= \int_0^T \int_0^X \mu_d(t)\mu_w(t)\mu_t(t)\mu_s(x)dxdt \\
G &= \sum_{i=1}^{N} \int_{t_i}^T \int_0^{x_i} g(t-t_i)h(x-x_i)dxdt
\end{aligned} \tag{5.35}$$

and attain the partial derivatives with respect to A and μ_0 as

$$\begin{aligned}\frac{\partial \log(L)}{\partial \mu_0} &= \sum_{i=1}^{N} \frac{\mu_d(t_i)\mu_w(t_i)\mu_t(t)\mu_s(x)}{\lambda(t_i, x_i)} - U \\ \frac{\partial \log(L)}{\partial A} &= \sum_{i=1}^{N} \frac{\sum\limits_{\substack{t_i<t \\ x_i>x}} g(t_j - t_i)h(x_j - x_i)}{\lambda(t_i, x_i)} - G.\end{aligned} \tag{5.36}$$

Setting these equal to 0 and solving, one attains the following iterative system of equations

$$\begin{aligned}\psi_i^{(\zeta)} &= \frac{\mu_0^{(\zeta)}\mu_d(t_i)\mu_w(t_i)\mu_t(t_i)\mu_s(x_i)}{\mu_0^{(\zeta)}\mu_d(t_i)\mu_w(t_i)\mu_t(t_i)\mu_s(x_i) + A^{(\zeta)} \sum\limits_{\substack{t_j<t_i \\ x_j>x_i}} g(t_i - t_j)h(x_i - x_j)} \\ A^{\zeta+1} &= \frac{N - \sum\limits_{i=1}^{N} \psi_i^{(\zeta)}}{G} \\ \mu_0^{\zeta+1} &= \frac{N - A^{(\zeta+1)}G}{U}.\end{aligned} \tag{5.37}$$

5.3.3 Constraining triggering functions

We note that this model aims to explain any residuals from the background process using the triggering component. As such, the model has significant freedom to adapt to the data. We can limit this freedom to ensure that the triggering component truly reflects increased rates of incidents on a short-time scale (compared to the background components) in the wake of a particular incident. A natural way to do this and extension of the original methodology is to enforce the triggering functions to be monotonic. This is a reasonable constraint that ensures interpretability of the triggering functions, whilst still allowing them to be constructed by the data. Triggering decaying in space and time suggests that we expect reduced influence of an incident as we move further from its location and as time passes. Existing work details how to ensure a smoothing estimate of data is monotonic, in-particular we use the methods described in [160]. Recall that one writes a smoothed estimate of some dataset (X,Y) as

$$\hat{\nu}(x) = \frac{1}{N} \sum_{i=1}^{N} k_\omega \left(x - X_i\right) Y_i \tag{5.38}$$

with all terms as previously defined. Specifically in our application, X represents differences in incident locations or incident times, and all Y values equal 1, however the technique works for more general cases. In Eq. (5.38), we place no constraints on the estimate $\hat{\nu}(x)$, however in practice there are many situations

where we may want to enforce some minimal structure. One such structure is monotonicity, with [160] introducing a generalised definition of Eq. (5.38), incorporating a weight p_i to each data-point used in the smoothing to 'adjust' the initial smoothed fit to be monotonic. One then writes this adjusted fit as

$$\hat{\nu}_{\text{mono}}(x|p_1,\ldots,p_N) = \sum_{i=1}^{N} k_\omega\left(x - X_i\right) Y_i p_i. \tag{5.39}$$

Our goal is now to choose a weight p_i for each data-point i used in the construction of the function, whilst altering the original estimate as little as possible. There are an infinite number of sets of $\{p_1, p_2, \ldots, p_N\}$ one could choose to enforce a monotonic function, and to identify a unique solution, we choose the set that is as close to the uniform distribution $\left\{\frac{1}{N},\ldots,\frac{1}{N}\right\}$ as possible. One distance measure used in [160] to compare the p_i's to a uniform distribution is

$$D_0(p_1,\ldots,p_N) = -\sum_{i=1}^{N} \log(Np_i). \tag{5.40}$$

Using this, we can then introduce a step in our model fitting where we solve a further optimisation problem, determining each p_i value to produce a monotonic triggering function. The optimisation problem is specified as

$$\begin{aligned} \min_{p_1,\ldots,p_N} \quad & D_0(p_1,\ldots,p_N) \\ \text{s.t.} \quad & \frac{d\hat{\nu}_{\text{mono}}(x|p_1,\ldots,p_N)}{dx} \leq \epsilon \\ & p_i \geq 0 \;\; \forall i \in \{1,2,\ldots,N\} \\ & p_i \leq 1 \;\; \forall i \in \{1,2,\ldots,N\} \\ & \sum_{i=1}^{N} p_i = 1. \end{aligned} \tag{5.41}$$

Since we are enforcing the triggering to decay with increasing time and distance, ϵ constrains the gradient of the triggering functions to be below some value. It should be noted that in practice, constraining the triggering function only alters the resulting functions in section 5.4.6, where we fit the model to small subsets of data. When fit to 13 months of data, the resulting functions are already monotonic.

5.3.4 Fitting algorithm

Our final reference back to the original work [124] is the description of the fitting algorithm used, with our additional components incorporated, detailed in algorithm 1. For computational speed, we assume triggering after 12 hours is 0, which is informed by considering time-scales similar to the worst recorded

Algorithm 1: Fit triggering point process model to data

Input: Initial guesses $\mu_d, \mu_w, \mu_t, \mu_s, g, h, A, \mu_0$

while *Not converged* **do**

  Compute $w_i^{(d)}, w_i^{(w)}, w_i^{(t)}, \psi_i, \rho_{i,j}$ for all valid (i, j) using Eq. (5.9), (5.15), (5.19), (5.24) and (5.30)

  Estimate $\mu_d, \mu_w, \mu_t, \mu_s, g, h$ using Eq. (5.13), (5.14), (5.18), (5.23), (5.31) and (5.32). Appropriate boundary correction should be applied to each.

  Determine $p_i \ \forall i \in \{1, 2, \dots N\}$ by solving Eq. (5.41)

  Estimate A and μ_0 using Eq. (5.37)

end

Output: Optimised components $\mu_d, \mu_w, \mu_t, \mu_s, g, h, A, \mu_0$

traffic jams in the UK, detailed in [161].

5.4 Results

5.4.1 Bandwidth selection

Selection of smoothing bandwidths is an open problem for models of this form, however the application discussed offers natural choices with time-scales inherent to the system. We choose the daily, weekly and trend bandwidths to be 60, 10×60 and $60 \times 24 \times 14$ minutes respectively. The value of daily bandwidth is selected due to the 'rush-period' behaviour in the UK typically varying on a time-scale of around an hour, whilst the weekly and trend components capture variation across larger time-scales. The spatial bandwidth is chosen as 5500 metres, which appears small enough to capture differing features across the M25, whilst not introducing superfluous oscillations. This is also larger than the uncertainty we would expect in incident locations, therefore accounting for potential uncertainty in the data. Finally, the temporal and spatial bandwidths for the triggering functions are chosen to be 30 minutes and 500 metres respectively. We considered variations on all of these values, finding those listed provided a reasonable compromise of model interpretability and identifying known components of traffic flow.

5.4.2 Model selection & prevalence of triggering

We first consider some measure of goodness of fit for models containing variations of the discussed components. The first is a simple, homogeneous Poisson process, used as the simplest reference model one could construct. We then compare models with: daily and weekly background components, daily, weekly and trend background components, daily, weekly background and triggering components, and daily, weekly, trend background and triggering components.

To compare these, we consider the log-likelihood, given by

$$\log(L) = \sum_{i=1}^{N} \log\left(\lambda(t_i, x_i)\right) - \int_0^T \int_0^X \lambda(t, x) dx dt, \tag{5.42}$$

with a larger value suggesting a better model. Note that in using log-likelihood to judge the goodness of fit, we ignore model complexity, the idea that we could continuously add components to any model and see increasingly marginal improvements as we attain a more complex model. However, our results throughout this section show the estimated functions present reasonable and interpretable behaviour that seems to catch marginal but still present features of the phenomenon of interest. To check the specification of the model we inspect the residuals and validate the results in section 5.4.5, while a successful attempt of out-of-sample validation is attained in section 5.4.6.

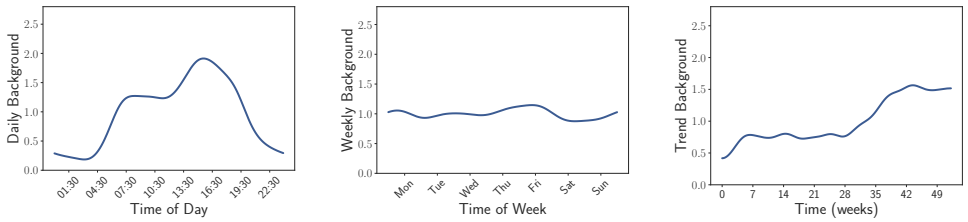
Model	A	Log-Likelihood
Fixed Rate Poisson Process	-	-28861.55
Daily + Weekly Background	-	-28028.05
Daily + Weekly + Trend Background	-	-27929.60
Daily + Weekly + Triggering	0.068495	-27864.48
Daily + Weekly + Trend + Triggering	0.065462	-27781.38

Table 5.1: Model log-likelihood values for models with various components. We see by far the worst model is a homogeneous Poisson process, as expected, and adding periodic daily and weekly components to this shows the largest improvement in log-likelihood. Including a trend and triggering component further improves the model log-likelihood.

Using this methodology, the parameter A can be interpreted as the proportion of the impact of the triggering function on the total intensity. For our optimal model incorporating all components, this means about 6.55% of incidents appear to be the result of triggering, in practical terms about 100 incidents. One may want to consider this as an upper bound, as discussed in section 5.4.4, along with an appropriate time-scale.

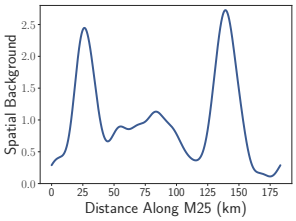
5.4.3 Background analysis

The background component of the model is strong, as seen when inspecting the changes in log-likelihood from table 5.1. In our model, this constitutes a daily, weekly, spatial and trend component. We visualise the temporal background components in Fig. 5.1. Inspecting Fig. 5.1a, we see that the daily background increases to an initial peak during the morning rush hour, then remains roughly constant, before rising again to a peak at around 4pm, and decaying after this. From Fig. 5.1b, there appears to be much less variation in the intensity across the week compared to all other identified components, but a slightly higher



(a) Daily background. (b) Weekly background. (c) Trend background.

Figure 5.1: Temporal background components fit to 13 months of data.



(a) Spatial background.

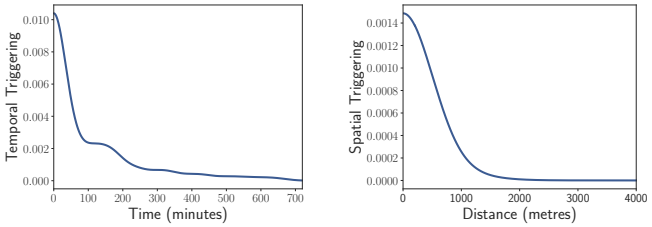
Figure 5.2: Spatial background components fit to 13 months of data. Recall that we have collected data for the clockwise direction of the M25, and as a result 'distance along the M25' is measured in the clockwise direction, starting below the gap for the Dartford crossing on the west side of the motorway.

intensity on Thursdays and Fridays, and the lowest on Tuesdays and Saturdays. Finally, we see a small increase in the trend during the first 7 weeks of the data, then it remains reasonably flat until week 28, where it begins to rise again. Around week 40, it stabilises again. This could be due to an increase in actual incident intensity, or more comprehensive reporting after a certain point, more operators and so forth, however no changes in reporting are known to us.

As well as the temporal background, where incidents are most common around the M25 is of interest. We show the spatial background in Fig. 5.2, from which a clear spatial structure is visible. We see two distinct peaks in Fig. 5.2a, and a smaller spread out peak in-between these two. Recall from chapter 3 that the largest peak, around 140 kilometres along the motorway, is located near the 'Potters Bar' junction. The second largest, around 25 kilometres into the motorway, is located between where the M25 meets the M26, and where the M25 meets the M23.

5.4.4 Triggering analysis

Triggering does appear to improve the log-likelihood of our model, and as we have seen in table 5.1 it explains around 6.55% of incidents in the data. We visualise the resulting functions in Fig. 5.3. From Fig. 5.3b, it is clear that spatial triggering is limited to around 2-kilometres, after which we do not see any non-zero values. However, temporal triggering, pictured in Fig. 5.3a,



(a) Temporal triggering. (b) Spatial triggering.

Figure 5.3: Triggering functions, fit to 13 months of data.

appears to have quite a 'long tail' in the sense that it decays over a relatively long time-range. However, this is a clear time-scale in this result of around 100 minutes. As a result of this, one should take 6.55% as an upper bounds of sorts, understanding this, combined with the identified time and length scales is an informative conclusion to draw.

5.4.5 Model validation

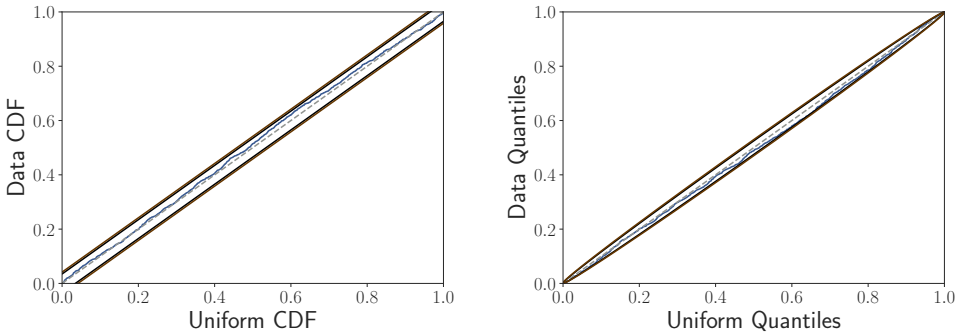
To validate if our model captures the relevant features of a process, one often follows the work of [125], [162] and [163]. Specifically, we use the transformed time-sequence given by

$$t_i \to \Lambda_i = \int_0^{t_i} \int_0^X \lambda(u, x) dx du, \tag{5.43}$$

then the resulting sequence of values $\Lambda_i \, \forall \, i \in \{1, 2, \dots N\}$ will follow a unit rate Poisson process if the model is correctly specified. To prove this, one can derive that the sequence of values $\Lambda_i - \Lambda_{i-1}$ are i.i.d. random variables that follow an exponential distribution with parameter 1, which is done in [163]. Given that we know the expected distribution of $\Lambda_i - \Lambda_{i-1}$, we can then transform this to follow a standard uniform distribution by computing

$$z_i = 1 - e^{-(\Lambda_i - \Lambda_{i-1})}. \tag{5.44}$$

Now, the computed z_i values should follow the simplest known distribution, which we can evaluate by comparing the measured and theoretical cumulative distribution functions (CDF), as well as the measured and expected quantiles. One can generate confidence bounds for the comparison of two CDFs by inversion of the Kolmogorov-Smirnov statistic, and for quantile-quantile (QQ) plots by using the fact that the order statistics of a uniform distribution follow a beta distribution. The distribution of the k-th order statistic has parameters k and $n + 1 - k$, with n being the number of sample points. Using this, we then generate CDF and QQ plots for our model, given in Fig. 5.4a and 5.4b. It is clear from Fig. 5.4 that the model is statistically defensible when inspecting



(a) CDF validation.

(b) Quantile-Quantile validation.

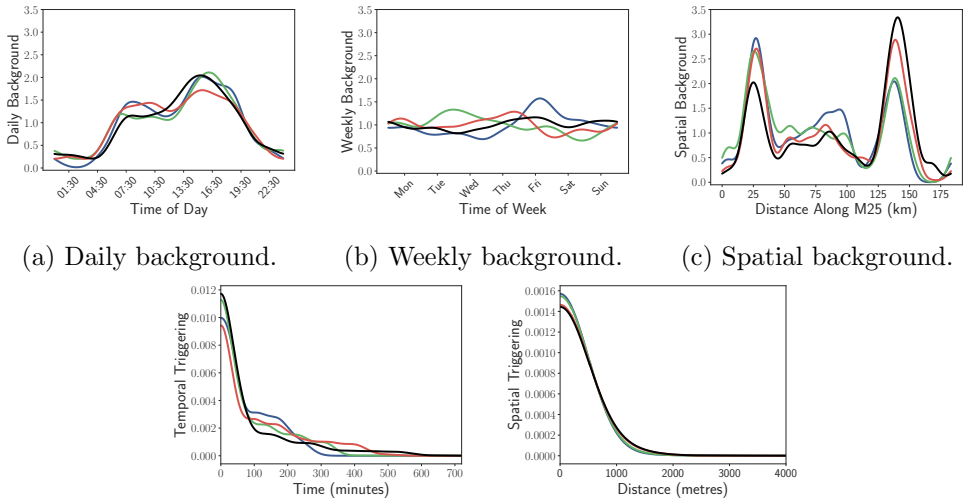
Figure 5.4: Validation results for 13 months of data. Model results are shown with —, 95% limits: —, 99% limits: — and a reference line: - - -.

both the CDF and QQ plots of the results. Some of the quantiles in Fig. 5.4b are just on the edge of acceptable, but do not deviate outside of the confidence bands. These results show that our model is well specified, however extra components could continue to be added to improve the fit. The compromise between model complexity and overfitting should always be considered when including additional features. Considering our application, traffic incidents are already rare in absolute terms as previously discussed. Further, the model has significant freedom to adapt to any data it is trained on. As a result, incorporating a large amount of features may quickly result in a model that is too specific to the training data, and does not generalise to unseen data. Hence, if further work does look at applying additional features to our model, one should always consider both in-sample and out-of-sample validation to ensure the model has not become too specific to the supplied training data. We detail how to do so at the end of the following section.

5.4.6 Do components change with season?

Given our results for the 13 months of data, we can question how resilient the background and triggering components are by inspecting subsets of the data. To test this, we partition our data into 3-month seasonal periods, and fit the model to each subset. We then overlay the components in Fig. 5.5. Summary statistics for these periods are given in appendices B.4.9, B.4.10, B.4.11 and B.4.12.

It is clear from our results in Fig. 5.5 that there is varying amounts of consistency in components of the model. Generally, we see in Fig. 5.5a that the daily background component is constant throughout the year, with a little variation present in the 3-months of data starting on 03/2018, where the evening peak is a little less pronounced than any of the other datasets. The morning peaks of the datasets starting at 09/2017 and 03/2018 are also a little



(a) Daily background. (b) Weekly background. (c) Spatial background.

(d) Temporal triggering. (e) Spatial triggering.

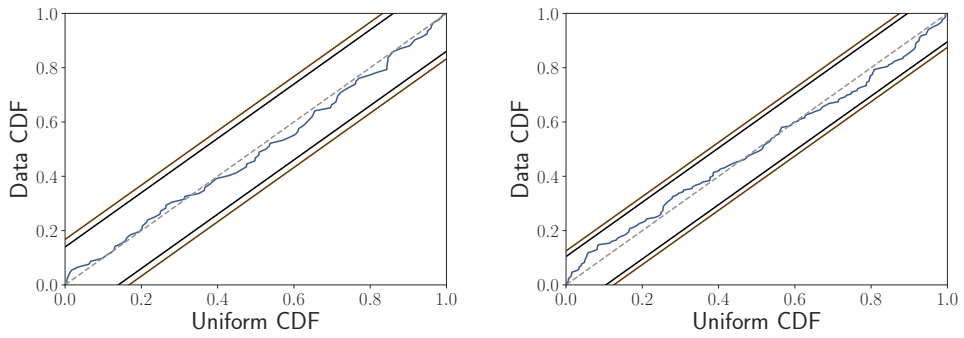
Figure 5.5: Background and triggering components compared across different 3-month datasets. Datasets are: 09/2017-11/2017: —, 12/2017-02/2018: —, 03/2018-05/2018: — and 06/2018-08/2018: —. Distance along the M25 is as defined in Fig. 5.2.

more pronounced than in the other two time-periods. It is difficult to make conclusions about the weekly component, when we have only 12 instances of each day of the week for a given 3-month subset. Of particular interest is the spatial background through time, shown in Fig. 5.5c, where we see that the peak around 140 kilometres along the M25 actually becomes more pronounced as time progresses. Both the peak at 25 and 140 kilometres are present throughout all subsets of data, but later periods appear to show that the peak around Potters Bar is more significant later in the dataset compared to earlier. It is unclear if a physical change occurred leading to this, but would be of interest to investigate with more data. Finally, the temporal and spatial triggering functions are generally quite consistent across all datasets. We still see somewhat of a long decay in the temporal triggering, but time-scales of around 100 minutes remain. Considering the triggering in each of these subsets, we attain A values of 0.034, 0.063, 0.068 and 0.075 for the periods starting 09/2017, 12/2017, 03/2018 and 06/2018 respectively. Without more data, we cannot say if these values vary throughout the year, or if they are increasing over time.

A further practical consideration when inspecting Fig. 5.5 is that in England, schools have a 6-week holiday in the summer. Whilst we expect this would alter urban and inner-city traffic behaviour significantly, we believe the impact on motorway traffic should generally be less substantial. This is mainly due to the amount of heavy goods vehicles that use motorways, and also that many motorway journeys are likely to be used by commuters to work. The lack

of traffic from students being dropped at schools may therefore have a minor effect, but we certainly do not expect it to be significant for our studied domain. The general consistency of the data subset '06/2018-08/2018' in Fig. 5.5 with the alternative subsets suggests our assumptions are reasonable, and whilst the spatial hotspot near Potters Bar does appear most significant during this time-period, it is clearly also a significant feature of the data in the preceding subset of data.

We note that, for small temporal subsets of data, the trend component becomes less impactful in the model. Recalling Fig. 5.1c, the trend is almost flat for long periods, suggesting we can omit it for some temporal subsets. As a result, we can perform out-of-sample model validation, something we are not aware of being done previously in the literature for this type of model. Since the trend is omitted, and we see the triggering components are consistent through time, we can train the model on some 3-month subset of data, then perform our validation on unseen data. We show two examples of this in Fig. 5.6a and 5.6b. It is clear from Fig. 5.6 that on short time-scales, one can



(a) Out of sample CDF Validation, fit to data 12/2017-02/2018, validated on data for 03/2018.

(b) Out of sample CDF Validation, fit to data 06/2018-08/2018, validated on data for 09/2018.

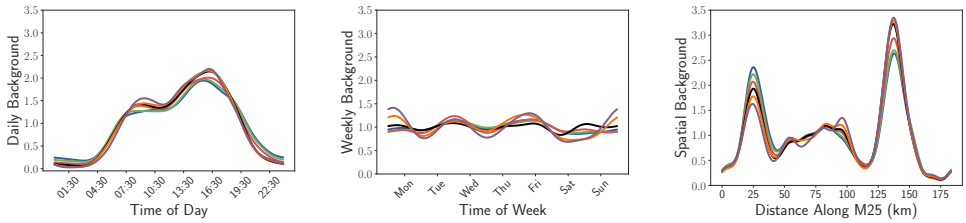
Figure 5.6: Out of sample model validation for two datasets. Model results are shown with —, 95% limits: —, 99% limits: — and a reference line: - - -.

use the model for acceptable performance out-of-sample. This is not possible over longer time-scales however, due to the clear trend and varying spatial background evident in the data.

5.4.7 Do components change for significant incidents?

Whilst all incidents flagged in NTIS should correspond to actual traffic incidents, many of them may not have had a significant impact on the traffic state. Consider the case where two vehicles have a minor collision, create no debris, and the drivers pull into the closed hard-shoulder to exchange insurance information. In such a case, there should be little impact on flow, travel

time and speed. Additionally, if a vehicle breaks down and pulls into the hard-shoulder, and the road is far from capacity, we should again see little drop in average speed. To consider only the behaviour of incidents that have some significant impact, we inspect the link-level data, which contains significantly less noise than the loop-level. For a given incident window, we consider the largest percentage drop in speed between a simple historical median segmentation profile and measured values across the entire window. If this percentage drop is above some threshold, then we say the incident caused a significant impact on the traffic state. As we raise this threshold, we isolate more extreme incidents but discard so much data that it is no longer reasonable to fit a model. To retain enough data for fitting, we consider only thresholds between 0 and 50%. We split our dataset into subsets containing only incidents that lead to a speed decrease of at-least 0%, 10%, ..., 50%, then re-run our model fitting on these subsets. The resulting background components are visualised in Fig. 5.7.



(a) Daily background. (b) Weekly background. (c) Spatial background.

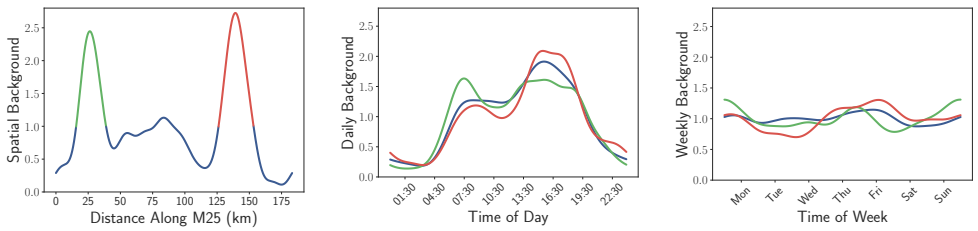
Figure 5.7: Background components compared across significant incidents, varying what speed decrease is required to define an incident as significant. Thresholds are: 0%: —, 10%: —, 20%: —, 30%: —, 40%: — and 50%: —. Distance along the M25 is as defined in Fig. 5.2.

Inspecting Fig. 5.7a, we see the daily background is reasonably stable across different thresholds of significance. The main difference is that as we raise the threshold, the morning and evening peak structure becomes clearer, and the periods very early and late in the day are lowered. From Fig. 5.7b, we observe that the weekend has a lower intensity than weekdays as we raise the threshold for significance. This is likely due to demand being significantly lower on weekends compared to weekdays, and hence when an incident does occur, there is less chance of an queue forming as the road is further from capacity than on a weekday. Finally, the spatial background in Fig. 5.7c clearly shows that the second peak, identified around 140 kilometres along the M25, appears to experience more significantly impactful incidents than the first peak, observed around 25 kilometres along the M25. This suggests that not only is Potters Bar a 'hot-spot' for incidents, but also that the incidents here are some of the more extreme on the network. Analysis of triggering

during different significance thresholds shows that the time-scales and A values remain consistent throughout.

5.4.8 Do components change around hotspots?

Whilst we have seen that the background component of our model has two clear peaks in it, we can also question if the daily and weekly background components differ significantly in the vicinity of these peaks, compared to their behaviour across the entire motorway. To investigate this, we consider each peak separately. Firstly, we define two 'hotspots' surrounding the two peaks in spatial intensity. These locations start where the spatial background component has a value above 1, and end where it then falls back below 1. We then isolate a set of incidents that occur in each spatial hotspot. For our model, this suggests we are isolating the spatial location where there is an increase in the rate of incidents compared to the average spatial location. We then re-fit our model using only this subset of incidents, and question what resulting daily and weekly background components arise. We visualise our results in Fig. 5.8. Specifically, we visualise the subsets by location in Fig. 5.8a, the daily background in Fig. 5.8b and the weekly background in 5.8c.



(a) Visualisation of hotspot locations.

(b) Comparison of daily backgrounds.

(c) Comparison of weekly backgrounds.

Figure 5.8: Identification of hotspot locations, and comparison of temporal background components around them. In 5.8a, we show the spatial locations of hotspot 1 (—) and hotspot 2 (—). In 5.8b and 5.8c, we show the results for the entire dataset (—), hotspot 1 (—) and hotspot 2 (—).

From Fig. 5.8, it is clear that the temporal background around each hotspot is reasonably similar to that across the entire M25. The daily background around the first hotspot has a slightly more pronounced peak in the morning and lower peak in the evening compared to the entire network. The second hotspot has a slightly higher peak in the evening. Both hotspots have a similar weekly component. Overall, it seems reasonable to conclude that behaviour at these two hotspots is not fundamentally different to that across the entire motorway.

5.5 Summary & conclusions

We have analysed the spatio-temporal variation in the incident rate on London's M25 motorway over a period of 13 months using a model that distinguishes between primary and secondary incidents. This variation is found to be strongly inhomogeneous. The temporal variation shows a strong daily double peak structure reflecting commuting patterns superimposed in a weaker weekly variation with a peak on Fridays and a trough on Saturdays. This pattern of temporal variation remains stable over the data period. The spatial variation shows two primary peaks in intensity. The first and largest is in the vicinity of the Potters Bar Interchange. The other is in the vicinity of Junctions 5 and 6 where the M26 and M23 join the M25. The peak at Potters Bar appears to increase in intensity during the data period, and is more pronounced when we condition on the most significant incidents in terms of impact on traffic speed. We find that 6-7% of the observed incidents are most probably secondary incidents under the assumptions of our model. Plausible time and length scales emerge for the range of the triggering effects: 100 minutes in the temporal triggering, and 1 kilometre for spatial triggering. From these figures we conclude that the effects of secondary incidents is a small but detectable feature of the M25 incident data set. Our analysis suggests that, on the M25 at least, the scope to further reduce accident rates by reducing the number of secondary incidents is limited compared to what could be achieved by reducing the peaks at specific times or 'hot-spot' locations.

Unlike other aspects of this thesis, this chapter does not present a real-time tool for operators to use. Instead, it presents retrospective analysis of the mechanisms that generate incidents on the network and an understanding of these mechanisms in space and time. Therefore, the main real-world application of our work would be to aid decision makers deciding where to focus on improvements to existing infrastructure. Clearly, our analysis of the background components of incidents on the M25 identify spatial locations where improvements should be made to reduce the background rate of incidents. Further, if changes are made to infrastructure, one could apply our modelling approach to data recorded before and after the change, and understand how these changes have impacted the different components that constitute the incident rate. Any analysis of this type would be subject to other variation in the traffic properties over the same time period, however if one can incorporate this, our work provides a useful way to judge the effectiveness of changes made to the network.

Similarly, if one repeats our analysis over the entire SRN, conclusions could be drawn regarding the impact of smart motorway features such as variable speed limits and message signs. The results from such an analysis can then

influence future infrastructure planning, as we can quantify the impact it has on different components on the incident generation mechanisms. Equally, one could investigate how altering the methodologies by which variable speed limits and message signs are set on current smart motorways impacts the time and length scales for secondary incidents. Doing so would involve close collaboration with the operators in control rooms, and extended periods of time where various methodologies for determining what speed limit to enforce or message to display are implemented. As a result, collecting enough data to confidently calibrate the model we have used would be expensive and time consuming, however it remains an option if there is enough commercial interest in evaluating alternative control strategies for smart motorway infrastructure.

For any of the discussed real-world applications, we believe our current approaches to displaying the model outputs throughout section 5.4 are both intuitive and allow for easy comparison across components. As all background components are scaled to have average value 1, it is simple to understand where increases or decreases to the average behaviour are observed, and one could overlay the model results before and after a change to clearly show differences in the components of the model. Equally, we believe the optimal way to visualise the time and length scales of secondary incidents is through visualisations similar to those we have shown, allowing direct comparison before and after any changes. Finally, since we can conclude that the parameter A is the proportion of secondary events in the data, we can in a single number quantify the prevalence of secondary incidents. Each of these visualisation would be informative and concise for decision makers to view.

In terms of incorporating more data, one needs to ensure that only components that could significantly impact the background rate are included in the model. As discussed, the model has significant freedom and specifying a huge number of inputs could quickly lead to overfitting. One candidate variable that could be investigated is the weather, specifically how best to incorporate this as it will likely influence driving conditions but there is no immediately obvious way to incorporate it into the model other than simply recording rainfall and ignoring other weather conditions. Another includes a measure of severity of an incident, for example the measure defined in chapter 4, however as events are already rare in absolute terms, providing a further segmentation of them may make fitting the model difficult. Inclusion of additional variables could be done through methods that consider marked point process, and this is a clear avenue for future extensions to our modelling.

There are two main of sources of uncertainty in our modelling. The first of these is the uncertainty in the start time of incidents and the locations determined by the localisation produced described in chapter 3. Delays in the reporting of incidents may result in the temporal background determined by

the model being slightly different to the true underlying background, however we have observed results similar to the literature so whilst this is an area of uncertainty, we do not believe it has significantly impacted our analysis. We have discussed that we cannot validate our localisation procedure, however have gone to lengths to ensure it is based on existing literature and expert knowledge. This does however mean our locations are somewhat uncertain and incidents may be placed incorrectly along a link. Since we apply kernel smoothing to each of the model components, some of the uncertainty is handled directly in the model construction, however we would like to repeat the analysis using the true incident locations. As a result of this we do not believe any particular changes should be made to the visualisations used when presenting results to decision makers based on these sources of uncertainty. Finally, since events are rare in absolute terms, collecting enough data to observe the underlying patterns in incident generation is difficult, however we believe the consistency we observe when comparing 13 months of data and 3 month subsets indicate there is sufficient data for our purposes.

Finally, missing data could bias the conclusions we have made, both through missing incident labels in the data, and missing sensor data making localisation impossible. There is little we can do regarding potential missing incident labels, and if there are specific locations or times where incidents are likely to be missed by operators, our background analysis will be biased against these times and locations. To avoid this impacting the applicability of our methodology in transportation planning, we suggest that both traffic operators and decision makers are involved in discussions of our model results, as the operators will have the deepest understanding of the potential for missing flags and when and where they are most likely to occur. This expert knowledge will augment our results, allowing decision makers to account for it alongside our analysis. Since we see a reasonably high number of missing speed values for links 63, 64 and 65 in our dataset during the studied period, we expect that the spatial background at these locations may be underestimated. This constitutes around 6 kilometres of road where we may see a lower contribution to the background than expected. Indeed, this may somewhat explain why the second peak in spatial intensity appears more pronounced in later subsets of data, when we observe less missing data at these locations. This should be directly mentioned when discussing our results with decision makers, and ideally results should be compared to future datasets when missing data at these locations is less prominent to asses any impact on the spatial background.

Chapter 6

Dynamic and interpretable hazard-based models of traffic incident durations

Understanding and predicting the duration or 'return-to-normal' time of traffic incidents is important for system-level management and optimisation of road transportation networks. Increasing real-time availability of multiple data sources characterising the state of urban traffic networks, together with advances in machine learning offer the opportunity for new and improved approaches to this problem that go beyond static statistical analyses of incident duration. In this chapter we consider two such improvements: dynamic update of incident duration predictions as new information about incidents becomes available and automated interpretation of the factors responsible for these predictions. We use the collected NTIS incident and time series data to train models that predict the probability distribution of incident durations, utilising both time-invariant and time-varying features of the data. The latter allow predictions to be updated as an incident progresses, and more information becomes available. For dynamic predictions, time series features are fed into the Match-Net algorithm, a temporal convolutional hitting-time network, recently developed for dynamical survival analysis in clinical applications. The predictions are benchmarked against static regression models for survival analysis and against an established dynamic technique known as landmarking and found to perform favourably by several standard comparison measures. To provide interpretability, we utilise the concept of Shapley values recently developed in the domain of interpretable artificial intelligence to rank the features most relevant to the model predictions at different time horizons. For example, the time of day is always a significantly influential time-invariant feature, whereas the time series features strongly influence predictions at 5 and 60-minute horizons. Although we focus here on traffic incidents, the methodology we describe can be applied to many survival

analysis problems where time series data is to be combined with time-invariant features.

6.1 Introduction & problem relevance

Traffic congestion can be broadly separated into two types: recurrent and non-recurrent. Recurrent congestion is simply the result of demand regularly exceeding capacity on busy sections of road during 'rush periods'. Non-recurrent congestion on the other hand is mainly caused by traffic incidents and rare incidents [44]. To better manage traffic during these incidents, traffic operators require reliable estimates of how long a particular incident will last.

Whilst there is significant existing work on modelling incident durations, many fundamental challenges remain that are both of practical interest to traffic management centres and remain active areas of research in an academic sense. A review of existing work on this problem is found in [44], where six future challenges for incident duration prediction are listed. Recalling the discussion in section 2.3, these are: combining multiple data-sources, time sequential prediction models, outlier prediction, improvement of prediction methods through machine learning or alternative frameworks, combining recovery times and accounting for unobserved factors. Cases by case, we detail how our work in this chapter addresses some of these challenges. Firstly, we combine multiple data-sources by using both operator reports, stating incident types, times and locations along with the high-resolution time series provided by the sensor network. Whilst we have seen other works directly input speed and flow measurements to a model, we first remove the seasonality from the series, then consider manually engineered features (levels and gradients of residual series) as well as automatically engineered features attained through applying temporal convolutions across the series. It is unclear a priori if these complex derived features will offer more explanatory power, however considering residuals from a baseline is a natural way to view the inputs.

Secondly, we consider time sequential prediction models both from classic survival modelling and using emerging machine learning methods. This is done by applying methods from the discipline of healthcare, specifically 'landmarking' for classic models, and a sliding window mechanism for machine learning methods, with more detail given in section 6.4.4. Thirdly, when outlier prediction is discussed in [44], it is specifically suggested that time-sequential methods could account for extremely long incidents as it would become clear as more information was available that a section of road was not recovering in typical times. Our dynamic approaches should do exactly this, allowing for some large predictions when the information suggests it. Fourthly, improvement of prediction methods through machine learning was listed as a challenge,

and this is clearly addressed by how we consider very recent work that has shown success in other disciplines and apply it to the problem of incident duration prediction. Further, using these methodologies allow us to consider a non-parametric form of incident distributions, based on work in [54], which is again a modelling advancement when applied to this discipline.

Finally, the review stated combining recovery times was a future challenge, by which it meant modelling when the traffic state on some link would return to some sense of normal operating conditions. We are explicitly able to measure this as we have a year of data, recording traffic speed, flow and travel time every minute. As such, we are able to generate data-driven, robust estimates of baseline behaviour for each link, varying by time of day and week. Using this, we can then measure when exactly the traffic state on a link returns to a level close to what would be normally be expected for that section at that time of day. This point is of particular interest to practical operators and road users.

We note from the outset that we want to ensure any model we make is of practical use, meaning as soon as an operator observes an incident, they can 'turn on' our algorithm to generate predictions of some form, feeding in new data as it arrives. We specifically do not want to use any features that would not be known either when the operator can see the incident location through a camera, or that are not provided by existing functioning sensors that are installed on the road. We overview the features used in our models in section 6.4.1, which are ultimately a combination of incident, temporal and spatial characteristics (time-invariant) along with traffic measurements taken from the sensor network (time-varying). One could include additional features such as injury reports as they become available however we do not consider that here.

The rest of this chapter is structured as follows. Firstly, we discuss the data, perform an initial exploratory analysis of it and detail how we determine data-driven speed baselines. We then describe the considered modelling approaches, both for static and dynamic predictions, and results for our dataset. Finally, we consider what variables are important for the models and end by summarising our main findings.

Note that throughout this chapter, an 'event' is defined as the traffic state on a section of road returning to a baseline behaviour. As such, when an event has 'occurred' we really mean the traffic state has recovered. This is just a note of terminology commonly used in the survival analysis literature.

6.2 Data details

For this work, we utilise the link level incident flags and time series from NTIS discussed in section 3.2.2. We use all data available to us, and detailed summary statistics and discussions of this dataset can be found in chapter

3.3.3 and table 3.1. However, we require further data processing to model the full duration of an incident, which we now discuss.

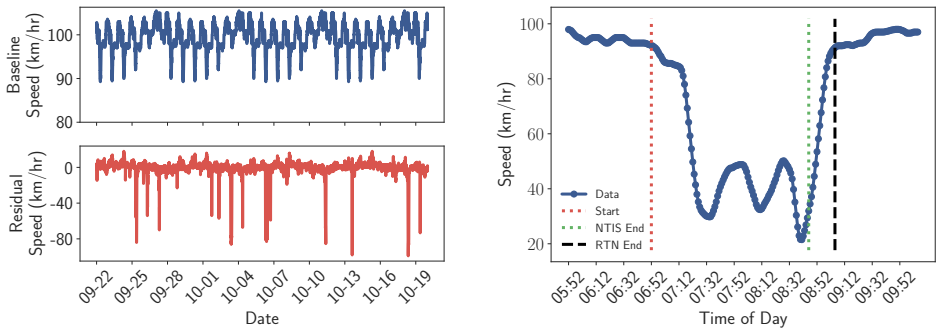
6.2.1 Establishing a data-driven baseline

As discussed in chapter 2, incident duration consists of 4 distinct phases, and an open problem highlighted in the literature is modelling the total incident duration. Like many works in this field, we consider an incident to have 'started' when it is identified in the system by the human operator. Our focus is to model the time from this point until both the incident has been physically cleared, and traffic behaviour on the road has returned to some sense of 'normal' behaviour. We use a speed profile to define this normal behaviour following the ideas in [49], where they define the 'total incident duration' to be the time from incident start until the speed has recovered to the profile.

How to determine an optimal speed baseline, and ensure it is robust to outliers whilst retaining important features of the data is an open problem. However, a natural way to approach it is to develop a seasonal model of the speed data. We do so by first taking the speed time series and pre-filtering it by removing the periods impacted by incidents. After, we account for any potential missing incident flags by further removing any remaining periods with a severity higher than 0.3. Severity is defined as in chapter 4. We then take this filtered series and extract the seasonal components, in our case daily and weekly components, to capture natural variability on the link. Note that inspection of the data shows no trend.

We try two methods of extracting this seasonality, the first is simple phase averaging, where for each minute in a week we take the median of all filtered data at that time of week and consider that to be the seasonal component of the series. The second is more complex, decomposing the series using the 'seasonal trend decomposition using loess' (STL) algorithm, first discussed in [164]. We see little difference between the two methods, so choose to use the phase average baseline for simplicity. We define one speed baseline for each link, establishing a robust profile describing the speed behaviour on a 'typical week', and replicate this over the entire data period. It is robust in the sense that we have pre-filtered the extreme outliers. It also captures the clear seasonality in the problem and has be applied to new test data without any difficulty. An example of this profile for a single link, along with the residuals from it are given in Fig. 6.1a, along with an example incident in Fig. 6.1. We specifically mark where the NTIS flag was raised, where it was closed, and where the speed behaviour returned to the baseline.

Using this methodology, we process our dataset such that we have a set of records with the start time of each incident and the time at which the link



(a) An example weekly baseline, and residuals for a single link, showing 4 weeks of data from 2017.

(b) An example comparison between the NTIS end time and the Return to Normal (RTN) end time. We are trying to predict the time until we return to normal operating conditions, rather than the time at which the NTIS flag is turned off.

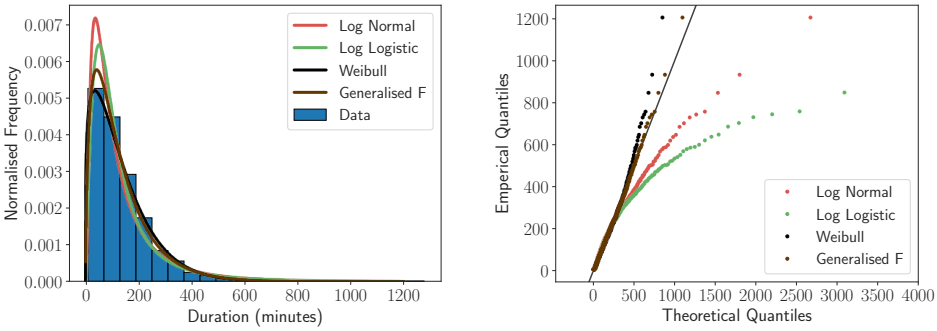
Figure 6.1: The baseline for a single link and an example incident. The baseline captures clear seasonality in the data. On a much shorter time-scale, we see a drop in speed in the wake of an NTIS incident flag, and a recovery after a sustained period of low speed.

returned to normal operating conditions. We include a safety margin in this baseline to account for any persistent but minor problems, shifting it down by 8 km/hr ($\approx$ 5 mph). A link is considered to have returned to normal when its speed is above this shifted baseline for at-least 3 consecutive minutes, acting as a persistence check. We then focus on predicting this return time given a variety of associated features.

6.3 Exploratory analysis

6.3.1 Incident duration analysis

As discussed in chapter 2, an initial step for many works in this area is to determine if the observed durations are well modelled by a particular statistical distribution. Our review of the literature suggested fitting log-normal, log-logistic, Weibull and generalised F distributions, so we do exactly this and show our results in Fig. 6.2. Specifically, Fig. 6.2a shows the probability density functions (PDFs) compared to the data and Fig. 6.2b shows a quantile-quantile plot. An overview of each distribution considered is given in appendix E.1. From Fig. 6.2, we see that the Weibull PDF follows the data most closely, but inspecting the quantiles we see only the generalised F distribution has reasonable estimates of the extreme values. When we apply a Kolmogorov-Smirnov with estimated parameters [165], we see no significant evidence that any of these candidate distributions are statistically valid fits. Note that in



(a) Comparison of probability density functions.

(b) Quantile-Quantile plot.

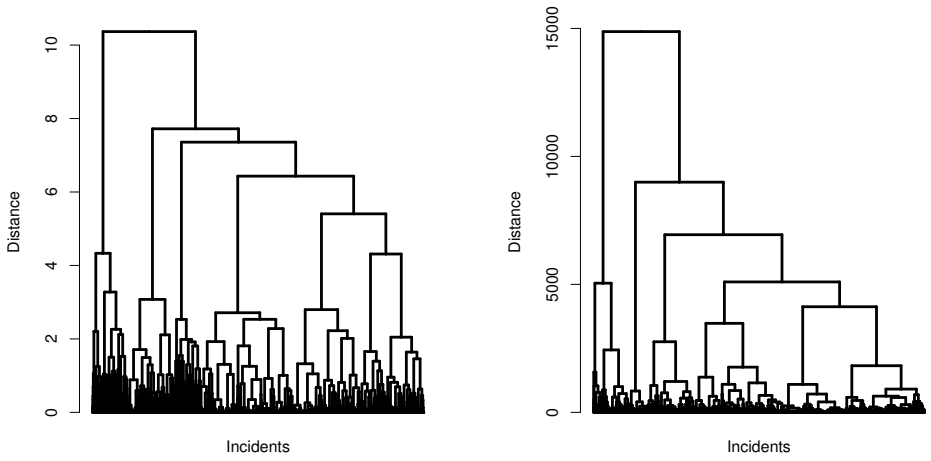
Figure 6.2: Analysis of various distributions fit to incident duration data across the M25. Overall, the Weibull distribution fits the bulk of the data most closely, but the heavy tail is best captured by the generalised F distribution.

[53], a chi-squared test was applied to determine the statical significance of distributional fits, however doing so requires binning the data, which we avoid here. These results suggest that to properly describe the data, we may want to consider more complex distributional representations, either non-parametric or mixtures of many components. Note also that the tail of the durations contains around 10 outliers, suggesting very rare incidents where the link was at a reduced speed for almost an entire day, however the bulk of the incidents attain a duration of less than 300 minutes.

6.3.2 Incident clustering analysis

We saw throughout the literature that time-invariant features were sometimes clustered to reveal different sub-populations of incidents, and provided useful features in modelling. This was particularly evident in [65] and [75]. We question if these clusters are also reflected in the time series for the data. To do so, we require a distance metric between time series. A common choice is dynamic time warping (DTW) distance, discussed in [166] and used in, for example, [167]. We specifically use the implementation in [168]. The idea is to stretch or squeeze a pair of time series such that they are as similar as possible. If one time series has the same shape as another, but is perhaps extended over a longer time period, they will have a low DTW distance, whereas if two are fundamentally different, they will have a large distance. We first standardise all link data by removing the seasonality and then subtracting any remaining mean and and dividing by the standard deviation, then compute the DTW distance between all pairs of series that represent a window where an incident occurred. From this, we attain a distance matrix, and we use hierarchical agglomerative clustering (HAC) with a ward linkage function [169] to construct

a dendrogram to visualise distances between elements. The results are shown in Fig. 6.3, where Fig. 6.3a clusters the data by their time-invariant features and Fig. 6.3b clusters the data by their time series. From Fig. 6.3, there



(a) Clustering of incidents using the time-invariant features.

(b) Clustering of incidents using the speed time series.

Figure 6.3: Clustering Dendrogram using all incidents in the NTIS data. We see clear structure, perhaps suggesting 4-6 distinct groups exist. Different distance metrics are used for each clustering, so the scales of the y-axis are different.

is clear structure in the distance matrix, suggesting that incidents cluster by both time-invariant features, but also by the observed traffic metrics on the links. The number of clusters is somewhat subjective, but if we inspect the average within cluster sum of squared distances, we see that the reduction in this begins to diminish after around 6 clusters are identified for both clustering instances. More evidence of this and discussion of the clustering is given in appendix E.3, along with more examples of time-series from incident windows in appendix E.2.

6.4 Methodology

Our modelling approach compares multiple methodologies to predict incident durations. First, we detail the features used for each of our models, then discuss at length how each of our proposed methodologies are implemented. A concise summary with all models considered and the main points of note about them is available in appendix E.4.

6.4.1 Incident features

There is clearly much work in the literature identifying the most influential features for incidents. However, we are restricted in two ways. The first is

that at any time-step of a prediction, we only incorporate features that would be known at that step. The second is that our dataset does not provide as comprehensive an overview of some features that are likely to be highly informative of the duration, for example we do not have in-depth injury reports or arrival time of police forces. The features we do use are separated into time-invariant and time-varying categories. The time-varying features are derived from time series of speed, flow and travel-time provided at a 1-minute resolution, which we take from the link an incident occurs on. We first remove the seasonality from each of these series by determining 'typical weeks' just as in the case of the speed baseline, then subtracting this from the time series to generate a set of residual series. We then hand engineer features that may be of use for some simple dynamic models, computing the gradient of the residuals at some time t using the previous 5-minutes of data from t, as well as simply recording the value of the series. These are used as initial features from the time series as they provide a sensible and intuitive summary of an incident from the sensor data, however of course more complex features can be derived from the series. We consider models that do just this by applying temporal convolutions across the residual series, and compare the modelling results in section 6.5.

A full list of the time-invariant features used are detailed in table 6.1, and are a combination of what an operator might know using existing camera and phone coverage of the SRN. Time-varying features are detailed in table 6.2 for completeness. Whilst not exhaustive, the features in tables 6.1 and 6.2 offer a combination of contextual information in time, space and specific to the incident. The main omission from our features that might be useful are details of injuries sustained (if any) by motorists involved in the incident. We did not record this, and do not know when it would be available to operators, so do not consider it for our dynamic cases. Weather information and potentially highly specific geometry information about the road surfaces could also be useful. Our choice of bins for time of day reflects typical commuting patterns in the UK. Some authors use day and night time as separation [52], whereas others account for peak times in their binning [55, 63], as we have. Whilst one could consider the binning used for time of day as a parameter to tune in the model, we believe this choice is reasonable for an initial exploration and consistent with general traffic patterns in our dataset. One could consider if starting the night time bin slightly later improves model performance, but with the current value we believe a significant amount of commuter traffic should have left the motorway at the time the night time bin begins. The same exploration of optimal encoding could be performed for many of these features, however our initial choices are consistent with the wider literature and are reasonable for an initial exploration.

Feature	Variable Type	Description
Binned daily time	Categorical	An indicator of the time of day. Bins: Morning Rush (6am-9am), Afternoon (9am-3pm), Evening Rush (3pm-6pm) and Night (6pm-6am).
Capacity reduction	Categorical	The fraction of all lanes that are blocked due to the incident, binned into 0-25%, 25-50%, 50-75% and 75-100%.
Incident type	Categorical	Specified incident type, coded as Accident, Vehicle Obstruction, Non-Vehicle Obstruction, and Abnormal Traffic.
Link length	Continuous	Length of the link the incident occurred on in metres
Link downstream atypical?	Binary	Is the link downstream perturbed to some atypical state at the time of the incident flag?
Link upstream atypical?	Binary	Is the link upstream perturbed to some atypical state at the time of the incident flag?
Number of vehicles	Count	How many vehicles are involved in the incident?
Has cascade?	Binary	Did the incident occur immediately after another incident nearby in space and time?
Has roadworks?	Binary	Did the incident occur on a link with roadworks active?
Spatial location	Categorical	Network is split into 8 sections (North, North East, East, ..., North West) and the incident location is specified with this encoding
Season	Categorical	What season did the incident occur during?
Vehicle types involved	Binary	Binary variables indicating if an incident involved a car, motorcycle, lorry, trailer and articulated vehicle.
Weekend indicator	Binary	An indicator specifying if an incident occurred on a weekend.

Table 6.1: Overview of the time-invariant features considered for incident duration modelling. These would be recorded by an NTIS operator when an incident is declared in the system and observed on the network. The presence of roadworks is determined using roadwork flags provided by NTIS, giving start times, end times and locations of any work performed. A weekend is defined as all times between (and including) 00:00 on a Saturday to 23:59 on the following Sunday.

Feature	Variable Type	Description
Time series residual	Continuous	Speed, flow and travel time residuals from their weekly baselines (used only in landmarking models)
Gradient of time series residual	Continuous	Gradient of speed, flow and travel time residuals from their weekly baselines (used only in landmarking models)

Table 6.2: Overview of the time-varying features considered for incident duration modelling. The time series used to measure the time-varying features are reported at the link level each minute. We further consider machine learning models that engineer features from the time series automatically.

6.4.2 Survival analysis methods

We first offer more detail on models in the vein of classic survival analysis that we will consider for our problem. We remind the reader that, following the convention in the survival analysis literature, we use the word "event" to mean the occurrence of the outcome of interest. In our case, this is the end of a traffic incident as determined by the return of the speed to within some threshold difference from the the profile value. Survival analysis methods aim to model some property of the duration distribution. Let $f_{\text{ID}}(t)$ be the PDF of incident durations, and $F_{\text{ID}}(t)$ be the CDF. A key component in survival analysis is the survival function $S_{\text{ID}}(t)$, which in our context describes the probability an incident has not ended by time t, where time is measured from the start of the incident. Denoting the event time (the incident end time) by T, we formally write

$$\begin{aligned} S_{\text{ID}}(t) &= \mathbb{P}\left(T > t\right) \\ &= 1 - F_{\text{ID}}(t) \\ &= \int_t^{\infty} f_{\text{ID}}(x)dx. \end{aligned} \tag{6.1}$$

Further, many survival analysis methods are concerned with the hazard function $\lambda_{\text{ID}}(t)$, describing the instantaneous rate of occurrence of events. One can show that

$$\begin{aligned} \lambda_{\text{ID}}(t) &= \lim_{dt \to 0} \left(\frac{\mathbb{P}\left(t \leq T < t + dt \mid T \geq t\right)}{dt} \right) \\ &= \frac{f_{\text{ID}}(t)}{S_{\text{ID}}(t)}. \end{aligned} \tag{6.2}$$

In practice, this means that the instantaneous rate of events is equal to the density of events at that time divided by the probability of surviving to that time. A final concept of note is the cumulative hazard function $\Lambda_{\text{ID}}(t)$, which

is the integral of the hazard function between time 0 and t, written as

$$\Lambda_{\mathrm{ID}}(t) = \int_0^t \lambda_{\mathrm{ID}}(s)ds. \tag{6.3}$$

Using these concepts, the first model we apply is a Cox regression model, reviewed in [170]. Suppose some 'individual' i (incident in this application) has covariate vector $\boldsymbol{x}_i$. A Cox model specifies the hazard function for individual i as

$$\lambda_{\mathrm{ID}_i}(t \mid \boldsymbol{x}_i) = \lambda_{\mathrm{ID}_0}(t)e^{\boldsymbol{x}_i'\boldsymbol{\beta}} \tag{6.4}$$

where $\lambda_{\mathrm{ID}_0}(t)$ is some baseline hazard at time t, and $\boldsymbol{\beta}$ is a vector of regression coefficients. The baseline hazard describes the hazard function for an individual with covariates all equal to 0, and then it is adjusted for a particular individual with the exponential of the regression term. In this original formulation, the covariate effect is constant in time, but the baseline hazard varies in time. Various methods exist for estimating a baseline hazard function, with more details found in [171]. In short, the baseline hazard is assumed to be piecewise constant and determined without any distributional assumptions, allowing the data to construct an approximation. One can determine $\boldsymbol{\beta}$ by optimising the partial likelihood

$$PL(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left[\frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{\sum_{j \in \mathcal{R}(\tau_i)} e^{\boldsymbol{x}_j'\boldsymbol{\beta}}} \right]^{\delta_i} \tag{6.5}$$

where δ_i is 1 if the event time is observed and 0 if it is censored and $\mathcal{R}(\tau_i)$ is the set of at-risk individuals at time τ_i. Here τ_i represents the recorded incident duration for incident i. The baseline hazard function can be determined using the Breslow estimator

$$\lambda_{\mathrm{ID}_0}(\tau_i) = \frac{\delta_i}{\sum_{j \in \mathcal{R}(\tau_i)} e^{\boldsymbol{x}_j'\boldsymbol{\beta}}}. \tag{6.6}$$

We use the implementation of Cox models provided in the R package 'survival' [172]. Ties in event times are handled using the 'Efron' method, detailed in [173], altering the likelihood in Eq. (6.5).

The next model we apply is an accelerated failure time model (recall AFT from the literature review). Such a model supposes relationships between the survival and hazard functions of the form

$$s_{\mathrm{ID}_i}(t) = s_{\mathrm{ID}_0}\left(te^{\boldsymbol{x}_i\boldsymbol{\beta}}\right), \;\; \lambda_{\mathrm{ID}_i}(t) = \lambda_{\mathrm{ID}_0}\left(te^{\boldsymbol{x}_i\boldsymbol{\beta}}\right)e^{\boldsymbol{x}_i\boldsymbol{\beta}}. \tag{6.7}$$

Here, $s_{\mathrm{ID}_0}(t)$ and $\lambda_{\mathrm{ID}_0}(t)$ represent assumed baseline survival and hazard forms, and covariates 'accelerate' or 'decelerate' the survival time of particular individuals. Given an assumed form, for example Weibull, log-normal or so on

with parameters $\boldsymbol{\theta}$, one can then fit this model through maximum likelihood, optimising

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \left[f_{\mathrm{ID}}(\tau_i)\right]^{\delta_i} s_{\mathrm{ID}}(\tau_i)^{1-\delta_i}. \tag{6.8}$$

A common way to interpret the AFT model is as a regression on the log of the durations, writing

$$log(\tau_i) = -\boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i \tag{6.9}$$

where ϵ_i is noise, with some assumed form. We implement the models using the R package 'flexsurv' detailed in [174].

As Cox and AFT models involve, in some way, a linear regression on covariates of interest, they are unable to account for potential non-linear, complex interactions and effects of variables without manual investigation and specification. One way to account for this is to instead use random survival forests (recall RSF from the review in chapter 2), which are non-linear models based on an ensemble of individual tree models. The basic idea is as follows. Firstly, one takes a training set and generates B bootstrap samples from it, that is samples with replacement. Each of these samples is used to grow a decision tree, however randomness is introduced in the growing of the tree, by selecting a set of potential split variables at any point when the tree needs to be split. The optimal split variable from this set is chosen to optimise some survival criterion. One of the most commonly used is the log-rank splitting rule [70]. The tree is then grown until some condition is met, either a maximal size or minimum number of cases remaining, and the output at the end of any branch is the cumulative hazard function for all data-points that are placed into that branch when passed through the tree. This process is repeated several times and the collection of trees is referred to as a forest. Each decision tree is a non-linear mapping from input covariates to the output cumulative hazard function, and the collection of many trees acts as an ensemble learner. Ensemble models are known to show promising performance in a range of tasks, and this in addition to the non-linear decision tree models suggests such models may improve upon Cox and AFT models for certain datasets. For our work, we use the R implementation found in [175].

6.4.3 Deep learning methods

Alternative approaches in survival analysis have focused on applying methods from deep learning to incorporate non-linear covariate effects and behaviours. One of the first such methods was in [176], where a Cox model is considered, but the term $\boldsymbol{x}_i'\boldsymbol{\beta}$ is replaced with $g(\boldsymbol{x}_i)$, the output from a neural network given input $\boldsymbol{x}_i$. Similarly, [80] and [177] extend the Cox model to a neural network setting, however fundamentally such models are still somewhat restrictive in

that they assume a form of the hazard function. More recently, [54] suggested to make far fewer assumptions, and instead train a feed-forward network to directly model the function $F_{\text{ID}}(t \mid \boldsymbol{x}) = 1 - \mathbb{P}(T > t)$, referred to as the failure function. To avoid specifying any particular form of this function, the output space is treated as discrete, defined on times $\{t_1, t_2, \ldots t_{max}\}$. We suppose a single output value in this discrete space at time t_j gives $\mathbb{P}(t_j \mid \boldsymbol{x}_i)$ and hence we derive $F_{ID}(t_j \mid \boldsymbol{x})$ as

$$F_{\text{ID}}(t_j \mid \boldsymbol{x}_i) = \sum_{t=t_1}^{t_j} \mathbb{P}(t \mid \boldsymbol{x}_i)\,. \tag{6.10}$$

However, we still need to enforce that the output vector actually defines a discrete probability distribution. A natural way to enforce this is apply a softmax function on the output layer, normalising the sum of the values to 1 though

$$\sigma_{\text{NN}}(\boldsymbol{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{t_{max}} e^{z_k}}. \tag{6.11}$$

Note that these values will also all be positive. In particular, [54] considered an application with competing risks, where individuals experienced one of many possible events. Here we consider a simpler case, having only one event (traffic state returning to normal), however the methodology remains consistent in this application. As we are only able to measure if an event has occurred each minute from our sensor data, the discrete nature of the model is not restrictive in our context, yet the non-parametric output is appealing as we have seen in section 6.3.1 that the data does not appear to be generated from any particular, simple closed form distribution. Such a model for the distribution also adheres to the increasing complexity considered for distribution modelling in the literature, as discussed in chapter 2. We adapt our implementation from the implementation found in [178].

For our implementation, we specify a t_{max} value equal to the longest duration, plus a 20% margin as in the original implementation, and define the output grid at a 1 minute resolution. However doing so leads to a very large output space for the model, and could potentially lead to over-fitting. To combat this, we apply dropout after every fully connected layer in the network, elastic net regularisation on the weights and early stopping based on holdout data. We further consider if a parametric distribution may be sufficiently flexible when attached to a neural network model to perform well in the prediction task. To test this, we build another model as described above, but remove the softmax output layer and replace with with a mixture distribution layer. The choice of a mixture distribution rather than a single distribution is influenced by [52], in which they found success with such distributions. We

choose our mixture components to be log-normal, avoiding the other specified distributions for numerical stability. This alters the output size to be $3 \times N_m$ where N_m is the number of mixtures, a hyper-parameter to tune.

A final alternative that compromises between the non-parametric discrete output and the parametric mixture is to allow the output layer of the network to define a set of weights, and construct a probability distribution from these weights using kernel smoothing. As discussed in the previous chapters, this is a non-parametric way to determine a distribution, however the output has some degree of smoothness unlike the discrete non-parametric approach. We write this weighted smoothing estimate as

$$\hat{\nu}(x) = \sum_{i=1}^{N} q_i k_\omega \left(x - X_i\right). \tag{6.12}$$

Here q_i are the weights for each kernel centre, N is the number of kernel centres and as before $k_\omega(x)$ is the kernel taken to be Gaussian with bandwidth ω, and the resulting estimate essentially builds a distribution as a weighted smoothed sum over all kernel centres. All kernels are normalised to avoid the leaking of mass problem. A point with high weighting will result in a significant amount of mass near this location, and a large bandwidth will smooth this mass out to the surrounding area. Applying this to our problem, it allows us to avoid treating the output space as discrete, and instead we place a kernel centre at each point in the formerly discrete grid, and treat the neural network output (including having a softmax applied) as values of the weights q_i. Doing this also enforces some amount of smoothness in the output distribution, determined by the choice of ω. We choose to use a bandwidth of 3 minutes, which still allows significant freedom to the distribution.

The actual function proposed in [54] to optimise in order to train the network is a combination of two loss values, the first accounting for the likelihood of the observed data and the second enforcing ordering. The likelihood loss function is given as

$$\mathcal{L}_1 = -\sum_{i=1}^{N} \left[\delta_i \log\left(\hat{f}_{\text{ID}}(\tau_i \mid \boldsymbol{x}_i)\right) + (1-\delta_i)\log(1 - \hat{F}_{\text{ID}}(\tau_i \mid \boldsymbol{x}_i))\right] \tag{6.13}$$

where $\hat{f}_{\text{ID}}$ is the PDF (or PMF in the discrete case) implied by the model output, and $\hat{F}_{ID}$ is the CDF or (cumulative mass function in the discrete case) implied by the model output, given a particular input $\boldsymbol{x}_i$. This is exactly as in Eq. (6.8), but taking logs, describing the likelihood of survival data, then taking the negative as we want to minimise the loss during training. The

second loss function is written

$$\begin{aligned} \mathcal{L}_2 &= \sum_{i \neq j} \mathbb{1}_{\tau_i < \tau_j} \eta \left(\hat{F}(\tau_i \mid \boldsymbol{x}_i), \hat{F}(\tau_i \mid \boldsymbol{x}_j) \right) \\ \eta(x, y) &= e^{-\frac{x-y}{\eta_\sigma}} \end{aligned} \tag{6.14}$$

where $\mathbb{1}_{\tau_i < \tau_j}$ includes pairs where $\tau_i < \tau_j$. This loss penalises the incorrect ordering of pairs in terms of the cumulative probability at their event time. If an incident i ends before j, then we would expect $\hat{F}(\tau_i \mid \boldsymbol{x}_i)$ to be larger than $\hat{F}(\tau_i \mid \boldsymbol{x}_j)$, and if so this pair is considered correctly ordered. Large deviations from correct ordering are penalised by $\eta(x, y)$, with η_σ controlling how rapidly the penalisation increases. The total loss function is then the sum of $\mathcal{L}_1$ and $\mathcal{L}_2$. The hyper-parameter grids used for all machine learning models can be found in appendix E.6, and all models are trained using 100 instances of random search in these grids.

6.4.4 Dynamic methods

To this point, all models discussed in this section have been static, that is an individual i has a covariate vector $\boldsymbol{x}_i$, it is passed through some model, and an estimate of its hazard function, failure function or alternative is attained. However, in-practice our specific application contains a significant amount of information that may be useful in determining the duration that is not available at the start of the incident. Such information in the traffic domain is a police report made on the scene, recovery information, and specifically of use to us, the time series provided by the sensors along the road. A significant incident on a road network could lead to speed drops, flow breakdown and travel time spikes, all of which will be evident when we inspect the time series as the incident progresses. However, the recovery of the link to normal operating conditions is closely tied to these time series, firstly through the level of the speed series (as this defines how far from a baseline we are), but one could imagine much richer indicators of traffic state can be mined from them. Recall Fig. 6.1, we see significant structure in the series, having a linear drop near the start of the incident, an unstable oscillation period at lower speeds, then a recovery to normal conditions. Further examples of these series for incident periods can be found in appendix E.2. A number of methods have been suggested to handle dynamic predictions in a survival analysis setting, incorporating this data as it becomes available and conditioning predictions based on this. We overview this work in the following section.

Landmarking

With any dynamic prediction approach, the goal is to provide estimates of a hazard function, survival function or similar at some time t, conditioned on the fact that the individual has survived to time t and any covariates they provide. A simple method to do this is known as 'landmarking' and is discussed at length in [179]. We note from the outset that landmarking is similar to truncated regression discussed in chapter 2, however this terminology is consistent with the wider survival analysis literature. To carry out landmarking, one first specifies a set of 'landmark times' $\{t_{\mathrm{LM}_1}, t_{\mathrm{LM}_2}, \ldots, t_{\mathrm{LM}_K}\}$ at which we want to make dynamic predictions. One then chooses some survival model, for example a Cox model, and the hazard function at landmark time t_{LM_j} becomes

$$\lambda_{\mathrm{ID}_i}(t \mid \boldsymbol{x}_i(t_{\mathrm{LM}_j}), t_{\mathrm{LM}_j}) = \lambda_{\mathrm{ID}_0}(t \mid t_{\mathrm{LM}_j}) e^{\boldsymbol{x}_i(t_{\mathrm{LM}_j})' \beta(t_{\mathrm{LM}_j})} \tag{6.15}$$

with $t_{\mathrm{LM}_j} \leq t \leq t_{\mathrm{LM}_j} + \Delta t$, for some Δt defining how far ahead we are interested in looking. Notice how, compared to Eq. (6.4), the covariate values $\boldsymbol{x}_i$ are replaced with those known at time t_{LM_j} and the regression coefficients and baseline hazard can vary based on landmark time. At each landmark time, only incidents that are still ongoing are retained, so the model is therefore conditioned on surviving up to this landmark time. To account for potential time-varying effects and avoid misspecification of the regression parameters, events that occur after $t_{\mathrm{LM}_j} + \Delta t$ are administratively censored, that is they are marked as censored if they survived past the look-ahead time of interest. Such a model is simple to implement as one can refine the dataset at different times to produce dynamic models. However, as the landmark time grows and less data becomes available, some power may be lost when drawing statistical conclusions. To implement these models, we choose landmark times t_{LM} of $\{0, 15, 30, 45, 60, 120\}$ minutes, and horizons Δt of $\{5, 15, 30, 45, 60, 120, 180, 240\}$ minutes and display results throughout section 6.5. Finally, the landmarking framework can be applied with models other than a Cox model, so we consider both Cox and RSF landmarking models as two candidate dynamic prediction models, with RSF offering a non-linear alternative. The same was done to compare to dynamic models in [180].

MATCH-Net based sliding window model

The models considered so far require us to manually engineer features from the time series variables to incorporate them as covariates. As discussed, we use the level and gradient of the residual series, as these will indicate both how close the link is to reaching standard behaviour, and if the situation is getting better or worse. However, gradients computed in short windows can be noisy,

yet computed on large windows may lead to significant delay in identifying features. Instead, we would like some automated method that, given a time series, is able to learn meaningful features from them and incorporate them into predictions. Such a method is proposed in [81], where the authors detail a sliding window model which they name MATCH-Net. We note that the algorithm is designed to make dynamic predictions accounting for missing data, although in our application we do not have missing data so are interested in the dynamic prediction aspect only. In this model, a window of longitudinal measurements is fed through a convolutional neural network (CNN), with the convolutions learning features from the time series that are then used for prediction of risk in some look-ahead window. The model slides across the data, shifting the window each time, meaning features are updated as time progresses and predictions are also shifted forward. It is upon this we base our sliding window methodology.

Specifically, we take a historical window of length w, and at time t feed the time series from time $t - w$ to t through a CNN, where the filters in the CNN aim to derive features from the series without any manual specification of what they should be. We then concatenate the features output by the CNN with the time-invariant features, and then pass these through a series of fully connected layers. In [81], the output layer was a discrete space upon which a softmax activation was applied, and we again consider this, a mixture of log-normal distributions and a kernel smoothed output distribution. At each input time, we consider a window ahead of the the same length as in the static case, and treat w as a hyper-parameter to optimise. We then optimise the network by minimising the negative log-likelihood. Finally, since this model is more complex and has far more parameters than in the former case, we actually consider the discrete distribution to be piecewise constant for 5 minute intervals. As a result, the output space decreases in size by 80% without sacrificing too much freedom. A schematic of the network architecture is given in Fig. 6.4, with the output layer left intentionally vague to be clear that we consider multiple different forms of output.

6.5 Results

A point infrequently discussed in the context of traffic incidents, is that there are multiple criteria that define a 'good' survival model, and multiple ways to measure this in the dynamic setting. We discuss some of these ways in the text below. We note that elastic net regularisation is applied to all deep learning methods, and the optimal Cox and AFT models are selected though inspection of sample-size adjusted Akaike information criterion (AIC) to avoid over-fitting. The data is split into 70% training, 30% testing sets, with performance metrics

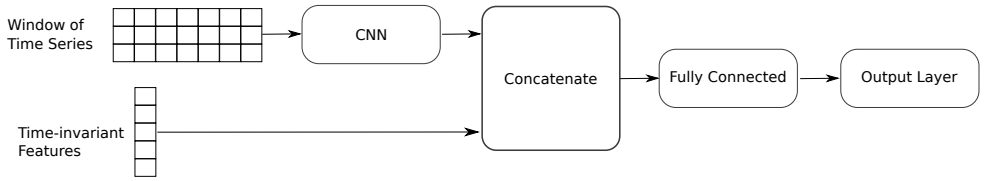


Figure 6.4: Network schematic for the sliding window model. We pass filters across the residual time series to engineer features, then concatenate these with the time-invariant features to create a feature vector, which is passed through a series of fully connected layers, and some output layer is applied to the result. The example shown is for a single traffic incident being passed through the network. The number of boxes for features is not to scale. The window of time series represents 3 variables and a window size of 7 in this simple example. The different time series variables are considered channels in the CNN context.

shown being computed only on the test set. The training data is further split, using 30% of it as hold-out data to judge when to stop training the neural networks.

6.5.1 Discriminative performance - concordance index

The concordance index (C-index) has different definitions in the static and dynamic setting. In the static setting, we write it as

$$C_{\text{index}} = \mathbb{P}\left(\hat{F}_{\text{ID}}(\tau_i \mid \boldsymbol{x}_i) > \hat{F}_{\text{ID}}(\tau_i \mid \boldsymbol{x}_j) \mid \tau_i < \tau_j\right). \tag{6.16}$$

Eq. (6.16) is the so called 'time dependent' definition used in [54], accounting for the fact that we care about the entire function $\hat{F}_{\text{ID}}$ and not a single point in the output distribution. In the dynamic setting, it is written given prediction time t and evaluation time Δt as

$$C_{\text{index}}(t, \Delta t) = \mathbb{P}\left(\hat{F}_{\text{ID}}(t + \Delta t \mid \boldsymbol{x}_i(t)) > \hat{F}_{\text{ID}}(t + \Delta t \mid \boldsymbol{x}_j(t)) \mid \tau_i < \tau_j, \tau_i \leq t + \Delta t\right) \tag{6.17}$$

which is the definition given in [180]. The only difference here is now we are specifically computing the C-index at a given prediction time and horizon rather than over the entire dataset. In computing this, we are taking the $\hat{F}_{\text{ID}}$ values at some time $t + \Delta t$ and compare pairs where an incident i actually ended in the horizon.

As described in [180], such a measure compares the ordering of pairs. If individual i experienced an event before individual j, then we would expect a good model to correctly assign more chance of an event to individual i than j. A model with perfect C-index, given N traffic incidents, will perfectly predict the order in which the incidents will end. This idea stems from viewing survival analysis as a ranking problem, and since we compare the CDF for two incidents, we see that it incorporates the entire history from a prediction time up to an

evaluation time, not just a single point measurement. A random model will achieve a C-index on average of 0.5, and a perfect model will attain a value of 1.0, so these are reference values to consider when interpreting this measure.

We formally compute Eq. (6.17) given our dataset by evaluating

$$C_{\text{index}}(t, \Delta t) \approx \frac{\sum_{i \neq j} \Psi_{i,j} \cdot \mathbb{1}_{\hat{F}(t+\Delta t \mid \boldsymbol{x}_i(t)) > \hat{F}(t+\Delta t \mid \boldsymbol{x}_j(t))}}{\sum_{i \neq j} \Psi_{i,j}} \tag{6.18}$$
$$\Psi_{i,j} = \mathbb{1}_{\tau_i < \tau_j,\, \tau_i \leq t+\Delta t}$$

where we simply evaluate empirically how often the ordering is correct conditioned on the requirements. The same is true for the static case. If two incidents happen to give exactly the same CDF values when evaluating, we take the convention of adding 0.5 to the total rather than 0 or 1, following [181].

6.5.2 Calibration performance - Brier score

Brier score measures how well calibrated a model is, and compares the binary label (1 if an event has happened at some time, 0 otherwise) with the model prediction at that time. Formally, we write

$$BS(t, \Delta t) = \frac{1}{N} \sum_{i=1}^{N} \left(\mathbb{1}_{\tau_i \leq t+\Delta t} - \hat{F}_{\text{ID}}(t + \Delta t \mid \boldsymbol{x}_i(t)) \right)^2 \tag{6.19}$$

where we sum over all incidents still active at time t, and ask did incident i end by $t + \Delta t$. If it did, then we would expect a good model to have a high CDF value at this point, with 1 being perfect (i.e. predicting the incident would end by $t + \Delta t$ for certain). On the other hand, if the incident did not end before $t + \Delta t$, then we would expect a low CDF value. This definition is that proposed in the supplementary material of [180]. In a sense, this measures the mean square error of a probabilistic forecast of a binary outcome. In terms of reference values, a model that outputs a survival function value equal to 0.5 at a particular time will have a Brier score of 0.25, so lower values than this are desirable, and a perfect model will achieve a score of 0.

6.5.3 Point-wise performance - mean absolute percentage error

Whilst C-index and Brier score are used throughout the survival literature, we also note that mean absolute percentage error (MAPE) is used throughout the traffic literature and has some practical relevance to our application. Since we have no censoring, we know the true duration for each traffic incident. As such, we can evaluate for every data-point, what is the error between a

point-prediction and the true duration. A natural choice for such a point prediction is the median of each models output distribution, [83, 182, 183] as the distribution of traffic incident durations is known to be heavy tailed. We can then ask what is the point-wise error for each model. Note that, C-index and Brier score asked about the accuracy of the output distribution, whereas this asks about a single point taken from the distribution. Highways England currently states that NTIS are measured on their prediction of the 'likely delay associated with an event'. Specifically, NTIS is scored as follows. One aggregates all incidents that have a predicted return to profile time made at their half way point and lasted over 1 hour. The MAPE between these predictions at the half-way point and the true values is computed. The target for NTIS predictions is for this to be below 35%, and it is stated in [184] that the current value in practice is 35.49%.

There is of course a problem with this criterion, in that a 'perfect' model by this standard would just always predict double the current duration, which would optimise the prediction at the mid-point, but be of no practical use aside from this. Regardless, this rough measure allows us to frame our work in the context of the practical considerations traffic operators are currently working towards.

6.5.4 Static prediction models

We begin by considering how a range of models perform on the data in a static sense. For this, we use only the time-invariant covariate information available at the start of the incident to fit the models. We apply each of the discussed models, and show results for all metrics in table 6.3. Generally,

Model	C-Index	Point-Wise MAPE	Brier Score							
			Prediction Horizon (minutes)							Mean
			5	15	30	45	60	120	180	240	
AFT (LN)	0.624	40.677	**0.000**	0.052	0.110	0.146	0.185	0.231	0.168	0.102	0.124
AFT (W)	0.624	38.543	**0.000**	0.052	0.113	0.148	0.186	0.226	0.164	0.100	0.124
Cox	0.626	38.545	**0.000**	0.052	0.113	0.149	0.186	0.226	0.164	0.100	0.124
RSF	**0.676**	39.961	**0.000**	**0.048**	**0.103**	**0.134**	**0.167**	**0.210**	**0.149**	**0.093**	**0.113**
NN (LN)	0.666	41.401	**0.000**	0.049	0.106	0.141	0.178	0.221	0.157	0.094	0.118
NN (NP)	0.647	**37.416**	**0.000**	0.050	0.108	0.142	0.179	0.223	0.163	0.101	0.121
NN (Kernel)	0.659	39.332	**0.000**	0.049	0.104	0.137	0.173	0.218	0.157	0.097	0.117

Table 6.3: Performance measures for models in a static setting, where we make only a single prediction per incident using a set of static covariates. Optimal values are highlighted in bold. AFT (LN) - An accelerated failure time model assuming a log-normal distribution of incident durations. AFT (W) - An accelerated failure time model assuming a Weibull distribution of incident durations. Cox - A linear Cox regression model. RSF - Random survival forest. NN (LN) - A feed-forward neural network model with an output layer that parametrises a mixture of log-normal distributions. NN (NP) - A feed-forward neural network model with a non-parametric output layer. NN (Kernel) - A feed-forward neural network with a kernel smoothed output.

we see the ordering of incident durations, measured by the C-index, attains values between 0.624 and 0.676. Recalling that a random model will achieve on average a C-index of 0.5, we see all models are informative, and the biggest gains in C-index are seen when we go from the linear to non-linear modelling frameworks. The RSF achieves the optimal C-index, followed by the neural network with a mixture of log-normals.

In terms of point-wise error, we do not make predictions at the half-way point of incidents, we only make them at the start of an incident in this setting. Doing so and measuring for all incidents longer than 60 minutes, we see that all models achieve a MAPE between 37% and 41%, with the best model being the non-parametric neural network. No model achieves an MAPE of less than 35%. Finally, the optimal Brier score is always achieved by the RSF method, with the most noticeable differences observed at horizons of 120 and 180 minutes. There is not much to distinguish many of these models, and ultimately one might suggest that in a static setting, a RSF offers a good compromise between performance measured by C-index, Brier score and MAPE, however if MAPE is the single desired criterion, a non-parametric neural network model would be preferred. It is likely that we would need more features that distinguish incident durations, for example injury details, police and recovery information and weather details to significantly improve these models, rather than more parameter searches.

6.5.5 Dynamic prediction models

We now consider the models in a dynamic setting, that is we use landmarking or sliding windows to provide more details for the model and condition our predictions on surviving to particular times. We now consider C-index as defined in Eq. (6.18), and show results for it at various prediction times and horizons in table 6.4. We see that initially, the RSF achieves optimal C-index across all horizons when predicting at $t = 0$. As time of prediction increases, we see a strong favouring of neural network models, specifically the sliding window model with a kernel smoothed output achieves the optimal C-index in most cases. There is a systematic preference for the non-parametric sliding window models compared to others at all prediction horizons when considering prediction times of 30 minutes or greater. Even at a prediction time of 15 minutes, the non-parametric sliding window models are preferred when considering 180 and 240 minute horizons. As a general summary of table 6.4, one should see that out of all 47 prediction time, prediction horizon pairs considered, the optimal model in terms of C-index is the RSF roughly 36% of the time, the sliding window neural network with kernel output 43% of the time, and the sliding window neural network with non-parametric output the

remainder of times.

Averaged over all horizons, we see that one would prefer the RSF model when initially making predictions, but all prediction times after 15 minutes favour the kernel smoothed output, with the non-parametric neural network often similar in performance. The neural network model parametrising a mixture of log-normals achieved the highest C-index of the neural network models in the static case, following the RSF model, however it never wins in the dynamic case, suggesting that when we provide the time-varying features, the RSF makes better use of them initially, and then the other neural network models make better use of them as time-passes. All models achieve C-index values higher than the reference value of 0.5, across all prediction times and horizons showing their predictions remain informative.

One point of note from table 6.4 is that the Cox model has quite poor C-index compared to the alternatives considered when predicting at a horizon of 5 minutes. We believe this is due to the amount of administrative censoring introduced at such a short horizon. If we look to [185], an assumption of the Cox landmarking model is that there is not too much censoring at the horizon time. For a very short horizon of 5 minutes, almost all incidents last longer than this, and hence, when we are applying our administrative censoring, this assumption becomes invalid, and we suspect this is why the Cox model has poor results at this horizon.

We further show the Brier scores for each model in table 6.5. Again, we observe that initially, the random survival forest achieves optimal scores across all horizons, however as time of prediction increases we gradually see the sliding window neural network with kernel smoothed output start to achieve better Brier scores for short prediction horizons. This is again systematic, and we see for a prediction time of 120 minutes that the optimal model is the sliding window neural network with kernel smoothed output at horizons up to and including 45 minutes, but for a prediction time of 45 minutes it is only optimal for horizons up to and including 15 minutes. One could postulate that initially, time series provide less useful information than the fixed features, that is at the very start of an incident, we see only the state of the link before the incident, which might have been reasonably seasonal. However, as time progresses, we will attain more informative features specific to single incidents, and in this case the fact the sliding window method engineers its own features, rather than the noisy gradient and level values we manually input to the RSF model, may prove more useful. Despite this, if a duration is very long, say 4 hours, and make a prediction 60 minutes into it, how much do we truly expect to gain from inspecting the time series so far? It may be that there is just simply no sign of recovery and all we can really conclude is that speed has been slow for a long time, and shows no other clear features.

Prediction Time (minutes)	Model	Prediction Horizon (minutes)								Mean Over Horizons
		5	15	30	45	60	120	180	240	
$t = 0$	Cox	-	0.851	0.799	0.754	0.712	0.654	0.642	0.638	0.721
	RSF	-	**0.870**	**0.832**	**0.803**	**0.774**	**0.698**	**0.667**	**0.651**	**0.756**
	SW (LN)	-	0.766	0.709	0.673	0.650	0.606	0.599	0.597	0.657
	SW (NP)	-	0.798	0.743	0.705	0.682	0.642	0.637	0.634	0.692
	SW (Kernel)	-	0.823	0.757	0.717	0.689	0.641	0.634	0.630	0.699
$t = 15$	Cox	0.513	0.864	0.784	0.732	0.698	0.648	0.639	0.637	0.689
	RSF	**0.956**	**0.893**	**0.826**	**0.788**	**0.751**	**0.686**	0.662	0.653	**0.777**
	SW (LN)	0.927	0.851	0.773	0.732	0.694	0.644	0.633	0.632	0.736
	SW (NP)	0.947	0.884	0.811	0.772	0.731	0.678	**0.669**	**0.666**	0.770
	SW (Kernel)	0.953	0.891	0.815	0.774	0.733	0.679	0.667	0.662	0.772
$t = 30$	Cox	0.529	0.861	0.770	0.731	0.702	0.662	0.652	0.648	0.664
	RSF	0.921	0.880	0.795	0.760	0.738	**0.699**	0.676	0.662	0.766
	SW (LN)	0.947	0.867	0.774	0.735	0.707	0.662	0.653	0.653	0.750
	SW (NP)	0.960	0.905	0.803	0.761	0.733	0.689	**0.681**	**0.677**	0.776
	SW (Kernel)	**0.971**	**0.907**	**0.810**	**0.768**	**0.739**	0.691	0.679	0.675	**0.780**
$t = 45$	Cox	0.504	0.859	0.796	0.764	0.724	0.679	0.667	0.661	0.707
	RSF	0.970	0.884	0.817	**0.787**	**0.757**	**0.706**	0.684	0.654	0.783
	SW (LN)	0.950	0.852	0.778	0.750	0.718	0.673	0.663	0.662	0.756
	SW (NP)	0.967	0.880	0.813	0.784	0.749	0.703	**0.692**	**0.688**	0.785
	SW (Kernel)	**0.974**	**0.893**	**0.821**	**0.787**	0.750	0.701	0.687	0.683	**0.787**
$t = 60$	Cox	0.578	0.872	0.796	0.755	0.723	0.684	0.670	0.668	0.718
	RSF	0.954	0.887	0.811	0.782	0.742	0.700	0.674	0.654	0.776
	SW (LN)	0.928	0.871	0.796	0.759	0.726	0.681	0.672	0.671	0.763
	SW (NP)	0.953	0.903	0.830	0.787	0.755	0.711	**0.702**	**0.699**	0.793
	SW (Kernel)	**0.969**	**0.912**	**0.834**	**0.793**	**0.758**	**0.712**	0.701	0.698	**0.797**
$t = 120$	Cox	0.522	0.850	0.804	0.781	0.750	0.706	0.692	0.683	0.724
	RSF	0.961	0.889	0.839	0.807	0.777	0.731	0.697	0.688	0.799
	SW (LN)	0.944	0.878	0.822	0.799	0.769	0.718	0.715	0.713	0.795
	SW (NP)	0.968	0.896	0.852	0.824	0.791	**0.744**	**0.739**	**0.735**	0.819
	SW (Kernel)	**0.986**	**0.904**	**0.853**	**0.825**	**0.793**	0.743	0.737	0.732	**0.822**

Table 6.4: C-Index values for considered models, across a range of different prediction times (when predications are made) and prediction horizons (at what time after the prediction time they are evaluated). Higher values show a better model. Optimal values for each prediction time - prediction horizon pair are shown in bold. Cox - A linear Cox landmarking model. RSF - Random survival forest. SW (LN) - Sliding window with a log-normal mixture output. SW (NP) - Sliding window with a non-parametric output. SW (Kernel) - Sliding window with a kernel smoothed output.

Another point of note when considering Brier score is that a RSF model appeared to perform favourably compared to a non-parametric neural network model in existing work. If we look to the supplementary material of [180], we see that a neural network with a non-parametric output did not consistently improve upon the Brier score achieved by a RSF model (see table VI in the supplementary material of the cited reference). It is unclear therefore if there is some fundamental reason for this in the modelling framework, as two entirely different datasets and applications appear to have observed the same behaviour. Despite this, all models generally achieve Brier scores below the reference value of 0.25.

Finally, we show the error in a point prediction made at various times throughout incidents in table 6.6. For reference we also include the value achieved by the fixed model to get an idea of what we are gaining from making dynamic predictions. From table 6.6, we see that when making a prediction after 30% of the duration of an incident has passed, we can expect between 30%

Prediction Time (minutes)	Model	Prediction Horizon (minutes)								Mean Over Horizons
		5	15	30	45	60	120	180	240	
$t = 0$	Cox	**0.000**	0.046	0.096	0.130	0.171	0.224	0.164	0.099	0.116
	RSF	**0.000**	**0.041**	**0.089**	**0.120**	**0.154**	**0.205**	**0.149**	**0.091**	**0.106**
	SW (LN)	0.006	0.096	0.197	0.260	0.315	0.317	0.204	0.115	0.189
	SW (NP)	0.007	0.082	0.160	0.213	0.262	0.290	0.195	0.113	0.165
	SW (Kernel)	0.006	0.079	0.159	0.211	0.261	0.292	0.197	0.114	0.165
$t = 15$	Cox	0.027	0.062	0.113	0.152	0.192	0.218	0.149	0.088	0.125
	RSF	**0.018**	**0.059**	**0.107**	**0.141**	**0.177**	**0.204**	**0.138**	**0.083**	**0.116**
	SW (LN)	0.020	0.077	0.157	0.212	0.266	0.280	0.176	0.097	0.161
	SW (NP)	0.019	0.069	0.134	0.180	0.230	0.259	0.170	0.096	0.145
	SW (Kernel)	**0.018**	0.070	0.136	0.181	0.230	0.259	0.171	0.097	0.145
$t = 30$	Cox	0.025	0.068	0.127	0.165	0.198	0.204	0.138	0.082	0.126
	RSF	0.020	**0.064**	**0.121**	**0.157**	**0.186**	**0.190**	**0.128**	**0.076**	**0.118**
	SW (LN)	0.019	0.074	0.154	0.204	0.250	0.255	0.161	0.090	0.151
	SW (NP)	0.018	0.067	0.138	0.183	0.223	0.236	0.155	0.088	0.139
	SW (Kernel)	**0.017**	0.066	0.137	0.180	0.218	0.235	0.156	0.089	0.137
$t = 45$	Cox	0.028	0.073	0.131	0.162	0.196	0.194	0.132	0.082	0.125
	RSF	0.018	**0.070**	**0.125**	**0.153**	**0.185**	**0.185**	**0.125**	**0.077**	**0.117**
	SW (LN)	0.021	0.082	0.157	0.201	0.245	0.239	0.154	0.090	0.149
	SW (NP)	0.019	0.074	0.138	0.176	0.217	0.221	0.148	0.089	0.135
	SW (Kernel)	**0.016**	**0.070**	0.136	0.173	0.213	0.219	0.149	0.089	0.133
$t = 60$	Cox	0.034	0.083	0.139	0.168	0.196	0.185	0.129	0.083	0.127
	RSF	0.023	0.079	**0.135**	**0.161**	**0.190**	**0.177**	**0.119**	**0.077**	**0.120**
	SW (LN)	0.025	0.082	0.154	0.199	0.240	0.228	0.145	0.090	0.145
	SW (NP)	0.023	0.073	0.136	0.176	0.212	0.210	0.139	0.089	0.132
	SW (Kernel)	**0.019**	**0.070**	**0.135**	0.172	0.208	0.208	0.140	0.090	0.130
$t = 120$	Cox	0.034	0.090	0.146	0.168	0.184	0.169	0.117	0.083	0.124
	RSF	0.024	0.084	0.132	0.157	**0.174**	**0.159**	**0.109**	**0.078**	**0.115**
	SW (LN)	0.026	0.087	0.148	0.179	0.213	0.199	0.137	0.091	0.135
	SW (NP)	0.023	0.082	**0.128**	0.157	0.189	0.183	0.130	0.089	0.123
	SW (Kernel)	**0.018**	**0.076**	**0.128**	**0.155**	0.186	0.183	0.132	0.090	0.121

Table 6.5: Brier scores for considered models, across a range of different prediction times (when predications are made) and prediction horizons (at what time after the prediction time they are evaluated). Lower values indicate a better model. Optimal values for each prediction time - prediction horizon pair are shown in bold. All keys are as defined in table 6.4.

and 33% MAPE in that prediction. This is around a 5-10% improvement from the prediction made from the corresponding static models at the start of the incident. If we make a prediction half way through an incident, we see that the neural network models all now achieve quite a significantly better MAPE value than the landmarking models, with an optimal MAPE of 21.6% achieved by the mixture of log-normals model, followed by 22% for the non-parametric model. The discrepancy between the sliding window and landmarking models grows as we make predictions later and later, with the sliding window models achieving an MAPE of between 16.5% and 17.3% compared to a value of 26.5% for the optimal landmarking model (RSF). Additionally, the prediction error shows very little improvement moving from the 50th to 70th and 90th percentiles of an incidents duration for the RSF model, and actually increases for the Cox model, suggesting that they are not sufficiently capturing signs in the time series that indicate the end is near. A key point of practical interest is that with the dynamic models, we do indeed achieve a MAPE value below 35% as desired by Highways England. Of course, we would need to attain data for all incidents across the UK to truly ensure that we are able to maintain this on a

wider scale, but as far as we are able to measure we achieve what would be considered industrially satisfactory error rates with the dynamic models.

Whilst the landmarking models appear to plateau in point-wise performance here, we note that it is partly due to the fact that very few incidents remain active at large landmarking times, and as a result the model is fit to smaller subsets of data and prediction performance is limited. We visualise how the average error evolves minute by minute in appendix E.10, which shows this. The plots within appendix E.10 are in general agreement with those one can find in other dynamic traffic incident prediction work, for example figure 2 in [83].

Model	MAPE Static Model	MAPE Dynamic Model			
		Percentile Into Incident Prediction Made at			
		30th	50th	70th	90th
Cox	38.545	32.513	31.002	32.180	35.923
RSF	41.607	**30.286**	27.707	26.478	25.319
SW (LN)	37.416	32.839	**21.576**	**16.506**	11.319
SW (NP)	41.401	31.660	21.998	17.069	10.399
SW (Kernel)	39.332	31.056	22.432	17.319	**10.040**

Table 6.6: MAPE at various points for incidents, all of which are at-least 60 minutes long. The optimal model for each prediction point is shown in bold. The point prediction is generated as the median of the output distribution from each model. All keys are as defined in table 6.4.

6.5.6 Do temporal convolutions improve predictions?

As we have applied methods from [81] to formulate a sliding window model, we have naturally included temporal convolutions to generate information from the time series. However, a valid question one could ask is are these required in our application, or do the simple levels and gradients retain enough information to make informative predictions? In essence, we are asking is the improvement observed in some metrics for the neural networks due to the non-linear feed-forward structure, or do the temporal convolutions also contribute? We test this by implementing a model without the CNN structure and instead feeding an input vector consisting of the time-invariant features and the level and gradients computed as in the landmarking case to a feed-forward network. We see that such a model achieves a worse Brier score and C-index at every considered prediction time and horizon compared to the equivalent sliding window model, and a worse error at the half way point of an incident, so the temporal convolutions do indeed appear to be an informative feature engineering method for our data. Interested readers can find the results for this in appendix E.7.

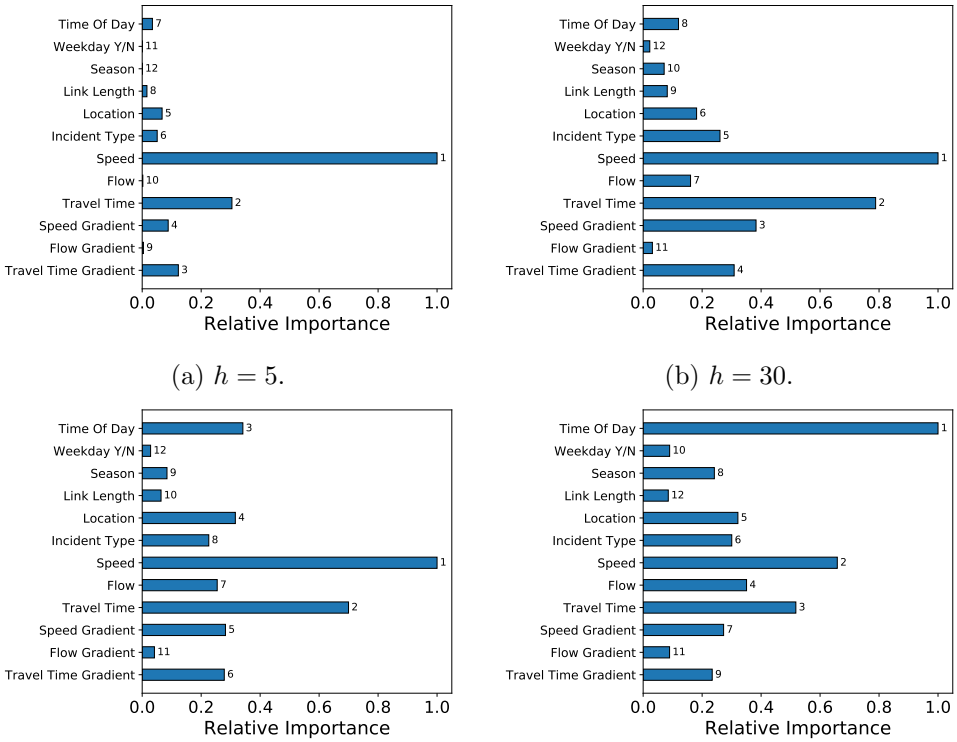
6.6 Variable importance

Variable importance has been studied for static models of incident duration many times in the literature, however here we consider variable importance in both our novel dataset and in the dynamic sense. We choose to evaluate variable importance for two dynamic models: RSF and the sliding window non-parametric neural network. Doing so for the non-parametric model makes interpretation of impacts simpler, as the output neurons represent the actual distribution at particular times, rather than some distribution parameters as in the log-normal case or weights as in the kernel case.

6.6.1 Random survival forest variable importance

As RSF are adaptations of random forest methods, standard variable importance metrics are well explored and readily implemented. Recall that trees are trained based on a bootstrap sampled dataset, meaning a set of observations remains for each tree that are 'out of bag' which will be used for measuring variable importance. Given a trained forest and some variable of interest x, we drop the out of bag data for each tree down the tree and whenever a split on x is encountered, we assign a daughter node at random instead of evaluating based on the value of x. We then compute the estimates from the model doing this, and the variable importance for variable x is the difference in prediction error between the original ensemble and the new ensemble ignoring the value of x. A large variable importance suggests that a variable is highly useful in accurately predicting the output. We compute the importance for all features, then scale the importance values by diving by the largest. This yields variable importance on a scale from 0 to 1 and we plot particularly important variables in Fig. 6.5.

From Fig. 6.5, we see that the speed value is the most important variable for horizons of 5, 30 and 60 minutes. At a horizon of 180 minutes, the most important variable becomes the time of day. This makes intuitive sense, as we expect the time-varying features to provide more useful information about the immediate future rather than times far into the future. Similar reasoning applies to the importance of travel time and flow. It is interesting to note that flow is often less important than other, time-invariant features, across all horizons. Flow is the tenth most important variable at a horizon of 5 minutes, seventh most important at horizons of 30 and 60 minutes, and fourth at a horizon of 180 minutes. We expect speed to be important as this directly defines what a normal state is, however clearly other features for an incident are more useful in predicting durations than the flow time series. The gradients are always less important than the residual values themselves, which may be a consequence of noise when estimating them, or the fact that the we can



(a) $h = 5$. (b) $h = 30$.

(c) $h = 60$. (d) $h = 180$.

Figure 6.5: Variable importance, as measured for the random survival forest model, for a subset of the important variables. Each plot is a different prediction horizon, all shown are at a prediction time of $t = 30$. The importance at each prediction horizon has been normalised such that the largest importance at any time is 1, and all others are relative to this. The rank of each variable at a given time is written beside each bar. Due to the scaling, one should focus on the ranking and relative difference between bars in each plot.

see short term rises and falls in the traffic variables that do not indicate the incident is actually near ending, but rather traffic state is just unstable as in Fig. 6.1. The location of an incident is always somewhat important, ranking between sixth and fourth across all horizons. This suggests clear heterogeneity in durations by location. Note that with a location and length, a model should be able to identify specific single links in the network, so predictions can be specific to these if it improves performance. However, the importance of length decays over horizons, and is always less than the location itself, so the coarse segmentation we have introduced for location seems more important than the specific link an incident occurs on. We see that the season becomes increasingly important as horizon increases. The type of incident is generally of medium importance, being the sixth, fifth, eighth and sixth most important variable going from horizons of 5, 30, 60 and 180 minutes respectively. This suggests that it is always somewhat important to know what type of incident the network is experiencing, however other factors can be more useful for predictions, such

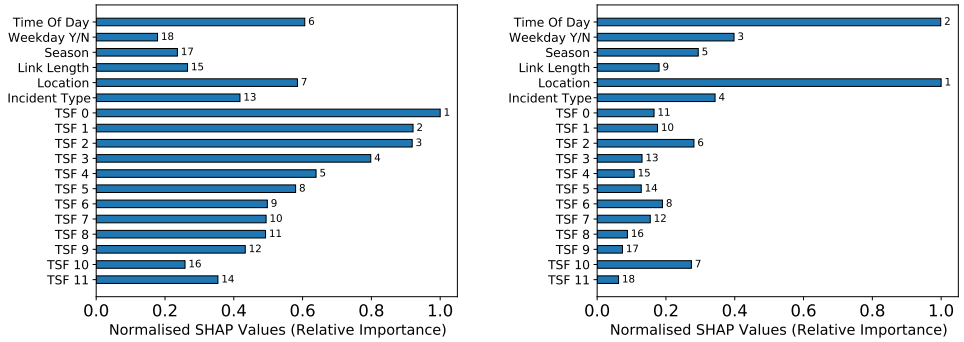
as location, time of day and the speed values. Finally, we might have expected more influence from the weekday indicator, however this is between the tenth and twelfth most important variable across all horizons for this specific model.

6.6.2 Neural network variable importance

Recently, there has been a significant effort to improve the interpretability of prediction models, both those involving neural networks and more general frameworks. As discussed in section 2.5.3, an emerging unified approach to this is to consider SHAP values, seeing how features shift a model output up or down from some reference value. It is upon this that we base our feature importance exploration for the neural network model. We compute the SHAP values for the network, for the time-invariant features and the features output from the CNN deriving information from the time series. Recall that computing the SHAP values for a neural network requires a background dataset, which we set to the training data, and we explain a random sample of 1000 instances from the hold-out set due to computational demand. We use the implementation[1] provided by the original authors of [131], specifically the permutation method for computational speed and the incorporation of structured inputs. A more extensive discussion of SHAP values for neural networks is given in appendix A.

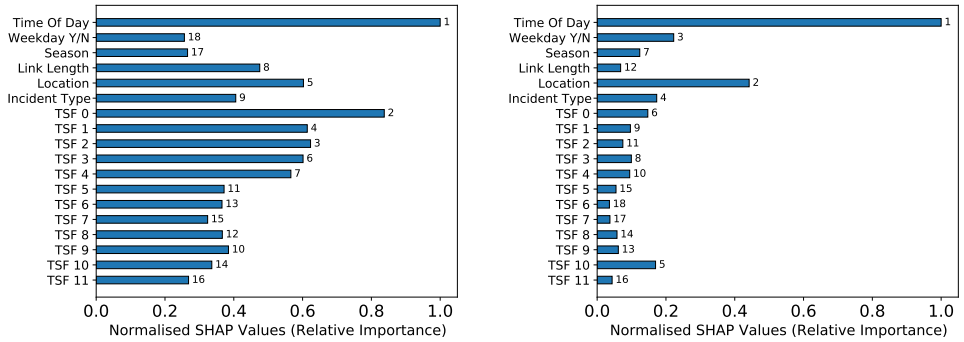
A point of note for this method of feature importance is that we are computing values for each output neuron in the network that correspond to particular horizons, and how this output value changes, not how some performance metric changes. As a result, we can question if different features have more or less impact on different parts of the output distribution for a single input data-point. First, we consider raw feature importance, that is does a variable have a large or small impact the output of the model at particular horizons, showing results in Fig. 6.6. From Fig. 6.6, we see that for very short horizons (5 minutes) there are a large number of time series features with high importance. This makes intuitive sense, as the time series are most likely going to be highly informative of short-term behaviour. After the time series features, we see the time of day, location and incident type are the features with the highest impact on the model output. Moving to a horizon of 30 minutes, we see the time series features become less important, and location and time of day dominate the other features. Note that at the 5 minute horizon, there were lots of features with quite high importance, showing quite a number influenced the model output, but at a horizon of 30 minutes we see two with large importance and many others with far less. At a horizon of 60 minutes, we again see the importance of the time series features increase, but the time of

[1] Implementation and examples given in `https://github.com/slundberg/shap`



(a) $h = 5$.

(b) $h = 30$.



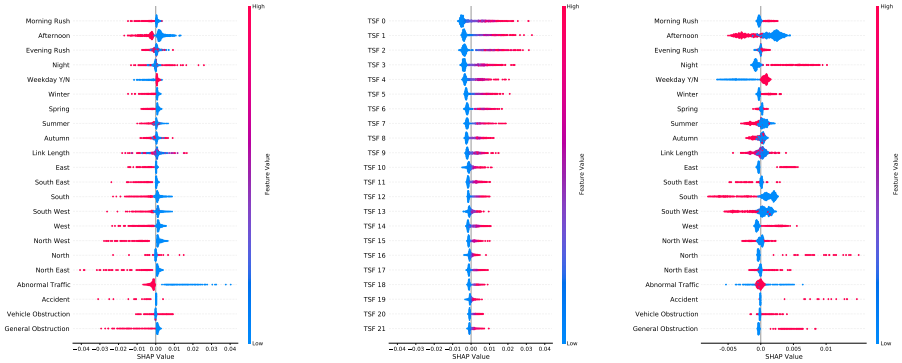
(c) $h = 60$.

(d) $h = 180$.

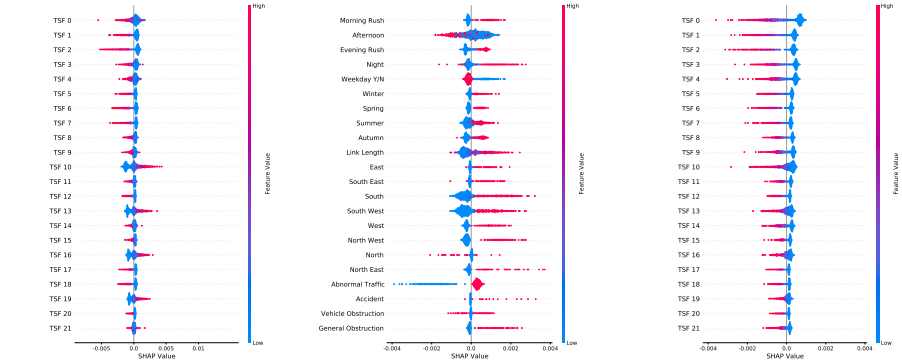
Figure 6.6: Variable importance, as measured for the sliding window neural network model, for a subset of the important variables. Each plot is a different prediction horizon, all shown are at a prediction time of $t = 30$. The importance at each prediction time is computed by taking the average of the absolute SHAP values for each feature across all query instances, and then has been normalised such that the largest importance at any horizon is 1, and all others are relative to this. The rank of each variable at a given time is written beside each bar. Due to the scaling, one should focus on the the ranking and relative difference between bars in each plot. TSF stands for time series feature, extracted by passing temporal convolutions across the data. As SHAP values are linear, one-hot encoded variables such as time of day have had their corresponding ϕ_i values summed, yielding a single result for that feature.

day and location are still the two most important of the time-invariant features, ranking first and fifth respectively. One might question why the time series resurge in importance here, and we explore this further in Fig. 6.7 and the analysis of it. At a long horizon of 180 minutes, the time of day is by far the most important feature, and the location is second. A natural interpretation of this might be that for long horizons into the future, knowing if an incident will overlap with rush hour or go into lunch time or the night is a good indication of if we believe it might last a long time.

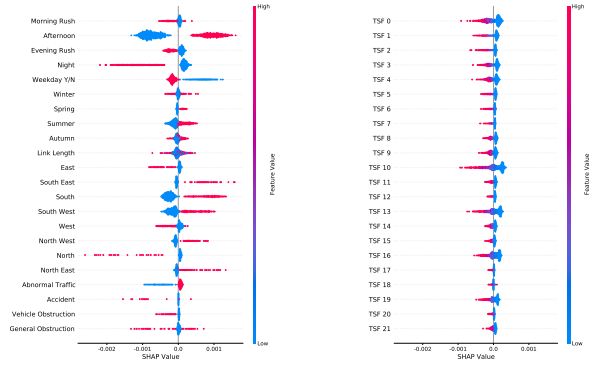
Having inspected the magnitude of SHAP values, we now question how do actual feature values shift the network output, either increasing or decreasing it from some reference output ϕ_0. Visualising this is more complex due to



(a) $h = 5$, static. (b) $h = 5$, series. (c) $h = 30$, static.



(d) $h = 30$, series. (e) $h = 60$, static. (f) $h = 60$, series.



(g) $h = 180$, static. (h) $h = 180$, series.

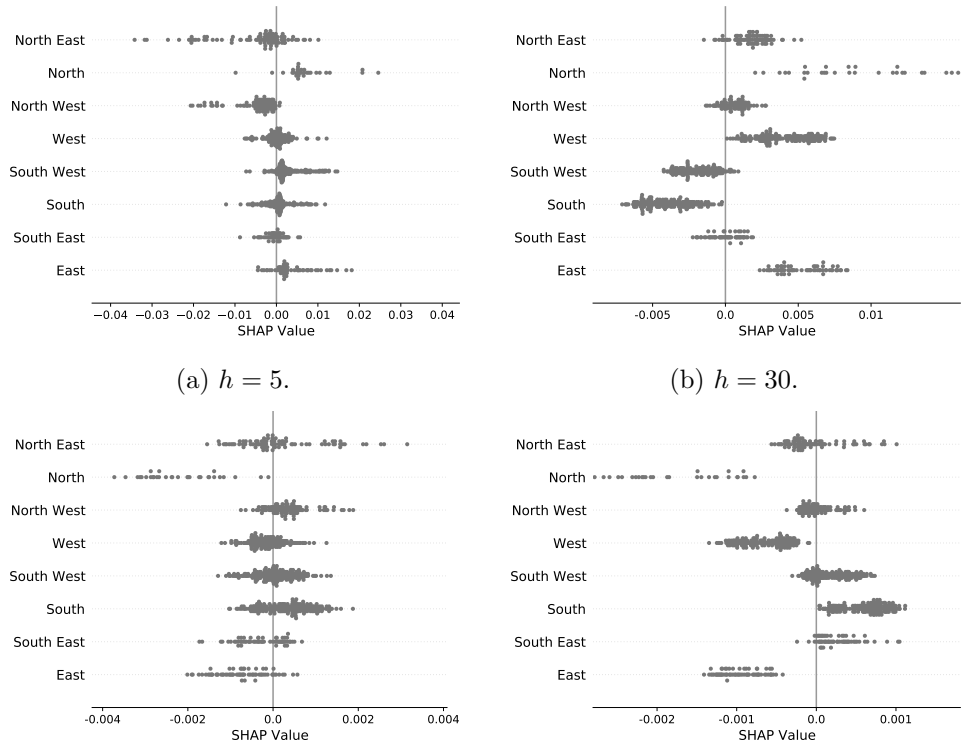
Figure 6.7: SHAP values for various features, plotted to visualise how feature values shift the output of the network at particular horizons. One should read these plots as follows. The y-axis shows model features, where categorical ones have been split into their one-hot encodings. The x-axis displays the raw SHAP values, showing how a particular feature shifts the model output up or down. Feature values are indicated by colour, with high binary features indicating 'yes'. Where many data-points had similar SHAP values for the same feature, points are expanded outwards in the y-axis, so a large vertical strip of points for a single feature indicates a high density of points at that SHAP value. Horizon is indicated by h in each sub-caption, corresponding to particular output nodes of the network. Time-varying and time-invariant features are split to aid readability. Pushing the model output 'up' (positive SHAP value) indicates increasing the probability mass at this horizon. TSF stands for time series feature, extracted by passing temporal convolutions across the data.

the fact that we want to plot the impact of various features, the value they attain, and if this shifted the prediction up or down. A standard way to do this for SHAP values is to make a so called 'beeswarm' plot, in which each data-instance is plotted as a single dot, once per each feature. Examples of making such plots for our dataset, and descriptions of how to read them are given in Fig. 6.7.

Whilst there is a large amount of information in Fig. 6.7, we breakdown the main points here. Firstly, we see that at a horizon of 5 minutes (Figs. 6.7a, 6.7b), there is clear coherence in the features, shown by there not being a random assortment of colours across single features. Earlier, we saw time series features, the time of day and location were influential factors at this horizon. Considering the time of day more finely, we see from Fig. 6.7a that when incidents are in the morning rush period, the value of the PMF at this time is decreased, suggesting the model has learnt it is unlikely incidents at this time of day will end very quickly. Inspecting horizons of 30 and 60 minutes, in Figs. 6.7c, 6.7e, we see that that when incidents occur in the morning rush, this increases the output values at these horizons. Finally, there appears to be more complex behaviour at a horizon of 180 minutes, as incidents occurring in the morning rush sometimes increase, and sometimes decrease the output at this time.

If we turn instead to view how a location impacts the result, for example inspecting the SHAP values for 'West' we see that attaining a value of 1 here decreases the model output at a horizon of 5 minutes (Fig. 6.7a), increases it at horizons of 30 and 60 minutes (Figs. 6.7c, 6.7e) and then decreases it again at a horizon of 180 minutes (Fig. 6.7g). Note however that since some of these features are in-fact categorical and have been one-hot encoded for use with a neural network, care must be taken in interpreting the impact of such values. In doing this analysis, we attain a SHAP value for each feature, however every data-point has as single location value equal to 1, and the rest equal to 0. So each location feature here will alter the neural network output, but the total impact of a data-point having a particular location will be the sum of the SHAP values for that data-point over all encoded categories. As such, we can better visualise the impact of categorical variables, for example location, by first summing the SHAP values for a data-point for all encoded groups of a particular feature, then visualising how the overall feature impacted predictions. We do so in Fig. 6.8.

The more intuitive description offered in Fig. 6.8 allows one to see the overall impact of the location variable. An example interpretation one could read from Fig. 6.8 is: for data-points where the spatial location was 'East', the overall impact of having this location on the model output was an increase in the output at horizons of 5 and 30 minutes, and a decrease at 60 and 180



(a) $h = 5$.

(b) $h = 30$.

(c) $h = 60$.

(d) $h = 180$.

Figure 6.8: SHAP values for the location feature. In each plot, we have summed the SHAP values for each one-hot encoded value, and then only plotted the resulting value in the row corresponding to each data-points true feature value. This shows the overall impact of the location feature, and allows one to view this impact separately for each value it attains.

minutes. From this more refined view, we can see that incidents with a location of 'North East' are quite varied, as their location sometimes increases and sometimes decreases the model output across all horizons. Incidents with a location of 'North' typically have the output increased at horizons of 5 and 30 minutes (Figs. 6.8a, 6.8b) and then decreased at horizons of 60 and 180 minutes (Figs. 6.8c, 6.8d). This suggests that, for example, the model has learnt incidents in the north are of shorter duration than incidents in the south, however this can then be adjusted further by other observed features. Locations can be compared in this way for all possible pairs.

A further question of interpretability relates to the features engineered from the time series: what effect do these have on the model? We previously saw that they were highly important at 5 and 60 minute horizons, but we can now use Fig. 6.7 to consider what impact they are having. If we start with a horizon of 5 minutes, we see from Fig. 6.7b that when the time series features attain a high value, the model output at this horizon is increased. We do not have an interpretable explanation of these features, but we see that

they can provide quite significant shifts up in the output if their values are high. Moving to a horizon of 30 minutes, we see from Fig. 6.7d that the series features become less coherent. Some features attain a high value and increase the output, others decrease it, and the overall impact is small for many features compared to the impact of the time-invariant features. This does indicate that the features learnt from the series are distinct in some sense, providing different impacts on the model output. At a horizon of 60 minutes, we see from Fig. 6.7f that the features attaining a high value indicates a decrease in the model value at this point. From this we can interpret that when high values of features from the are attained, they are making the model put more mass in the immediate future, and less at horizons of 60 minutes or longer. This is perhaps an intuitive result, that the time series are providing information that can significantly increase the model output at short term horizons, and attaining these same values shifts down the predictions at long horizons.

Of course, the sheer amount of data available here is somewhat overwhelming, however using SHAP values one can gain a significant understanding of why the machine learning model is outputting particular values. We further show plots for the overall impact of categorical features in appendix E.8, for interested readers, omitted for brevity here. However, care must always be taken in ensuring that feature importance and causality are not assumed, rather we are able to question why the model gives a set of outputs for a particular set of inputs.

6.7 Summary & conclusions

In this chapter, we have addressed a number of issues raised in the literature regarding traffic incident duration prediction. Firstly, we considered a method to determine incident duration accounting for both when an operator declared it cleared, but also when the traffic state at the location had returned to some typical behaviour. This ensured that our predictions reflected when commuters could expect normal traffic conditions to resume if an incident had a significant impact even after it was cleared. Secondly, we considered a range of models, some based on classic survival analysis and others based on machine learning and assessed how they performed on our dataset in both a static and dynamic setting. In-particular, we took inspiration from work in the domain of healthcare and applied emerging methods used there to problems in traffic incident analysis. We saw that in a static setting, there was little to choose between the models but generally either a neural network or random survival forest method would be preferred to the others considered. We moved into the dynamic setting by utilising landmarking and sliding window neural networks that applied temporal convolutions to the time series, both of which

were inspired by success in healthcare applications. We saw non-parametric methods that made no distributional assumptions were preferred to methods that parametrised mixture distributions, and saw some benefit in applying kernel smoothing to enforce some minimal structure on a non-parametric output. Utilising models with non-parametric distributional forms was influenced by the increasingly complex distributions being considered for this problem in the traffic literature, and showed significant promise in our results.

We assessed how each model performed using three different scoring criteria and in the dynamic sense, we saw clear structure in the results. The kernel smoothed neural network model generally achieved optimal C-index when making predictions on or after 30 minutes, and the random survival forest was optimal at times before this. Optimal Brier score was dependent on the prediction time and horizon considered, with the random survival forest generally being preferred, however at long prediction times and short prediction horizons, the non-parametric neural network models were optimal. Finally, we saw the neural network models showed much better performance in terms of point-wise error than the comparison models. We also related our results back to practical considerations, noting that the existing industrial specification is to achieve a prediction error at the half-way point of an incident of less than 35%. We showed the dynamic models considered here were able to do this, with the optimal model achieving an error of 21.6% at this point. We further questioned if one gained anything from applying temporal convolutions to the time series data, or if feeding manually engineered features to the network achieved the same performance, as had been done in previous traffic studies [78, 82]. We saw that the C-index, Brier score and error at half-way point were all improved by applying these temporal convolutions, suggesting future work should continue to explore optimal methods of deriving features from sensor data rather than inputting manually engineered features.

After, we considered variable importance for our model in the dynamic sense, assessing how the random survival forest and neural network models were influenced by the derived features. Whilst we are aware of variable importance being studied in the traffic literature previously, we are not aware of SHAP values being applied to neural network models for incident duration analysis. Time of day and location were generally important across models, particularly at long horizons, and the time series features were shown to have significant impact on the neural network output at 5 and 60 minute horizons. Specifically for the random survival forest, we saw the in addition to time of day and location, the time series residuals and incident type strongly influence predictions. Ultimately we hope that our work highlights a range of potential ways to address pressing issues in the field of incident duration analysis, and leads to more cross-disciplinary consideration of models and approaches that

have shown success in other applications.

Finally, our suggestion to use the output of a neural network to define kernel weights, and from this construct a non-parametric distribution through smoothing was novel in that we are not aware of this being done before to the considered model. It has relevance to other applications, specifically if one requires a continuous distribution to be output, but does not want to make strong parametric assumptions about what form this will take. Further work could be done to consider alternative methods to select a bandwidth when applying this type of model, but the freedom it provides appeared promising on our data.

To take this work forward into a real-world application, new data-sources would first need to be incorporated and the model retrained appropriately. The data that we believe should be included before the model is used in a traffic control room are information on injuries to drivers and passengers, emergency service responses and arrival times, clearance notifications for disabled vehicles and weather conditions. These could realistically be incorporated by the current control room operators, recording into the system what emergency services are dispatched to the scene and when they arrive, transcribing injury reports and noting when obstructions have been cleared. Weather may be more complex, as we would ideally want to incorporate information relating to visibility, rainfall and details on any surface water. If each of these data sources were available, it would however be a simple extension to incorporate them into some of the modelling frameworks we have described. The sliding window neural network specifically in the original implementation has a missing mask designed to enable the network to account for missing values at different time points, so we suggest future work implements this for a transport setting. One would then simply need to retrain the model with the new inputs and recompute the variable importance scores.

After doing so, we suggest displaying the output of our model to operators by printing estimates of the median duration predicted at a given time, along with the 25th and 75th percentiles of the output distribution. Doing so would, in three numbers, summarise the expected behaviour and associated uncertainty in this prediction. By displaying the associated uncertainty, an operator could then take this into account when acting upon the model output, altering their response based on the variability identified by the model. It may be that when tested, alternative uncertainty estimates are preferred, for example the 5th and 95th percentiles, however this is a simple aspect to adjust. The operator may then be able to expand the explanation for a prediction, where a plot similar to Fig. 6.6 could be displayed, acting as a sense check for why the model is outputting a particular result.

We believe that the main source of uncertainty in this chapter relate directly

to the unobserved data that should be collected and incorporated before practical use of our model. These features will likely be highly informative of the duration of an incident, and hence training a model without them does not account for their potentially significant influence. Since we expect these to be incorporated before practical use of our methods, this should not impact an operators use of our methods. A second source of uncertainty is fundamental to the underlying problem. It is seen both in our work and the wider literature that there is fundamentally a large amount of variation in the duration of traffic incidents. As a result, predictions will always be subject to some amount of uncertainty, even as more features are incorporated into models. Our suggestion to present a measure of this uncertainty to operators during practical use of our method is based upon this observation, and we feel this is a fundamental aspect of our work. This further provides justification to apply survival analysis methodologies rather than those for point-predictions alone.

Missing data, as with the previous two chapters, could create a bias in our results. Firstly, if there is a systematic process or reason resulting in incidents not being logged more often for particular locations or times, then our model will be trained on only a subset of the true range of incidents the network is subject to. This may result in a bias as the performance is optimised over this subset, rather than a representative sample of the true population. Since we see our tool being used directly by operators, we expect them to be aware of any potential mechanisms that would cause such a recording pattern. Further, since events that are not input into the system are mostly likely those with a minor impact on the network, we do not expect them to be the focus of applying the models we have considered. Rather, it is likely operators are most interested in predictions for the most significantly impactful events in terms of causing congestion and reducing flow.

The other potential bias in our methodology is that locations with a significant amount of missing data may have behaviours and incident behaviour that is not observed on other sections of the domain. As a result, the model will not have optimal performance on these incidents if they exist. This could create systematic errors in the output if locations start to report data after the model is put into practical use. To overcome this, we believe retraining the model periodically may be useful, and potentially weighting more recent observations higher than older ones. This should ensure that the model is constantly adapting to the potential inputs it will receive. Along with this, displaying the most recent training period to operators can give them a sense of what input data the model has been optimised for, and they can then use their judgement on a case by case basis to decide if the model is appropriate to use. We expect both this and the previously discussed bias to be most

prominent for links 60-69 along the M25, giving a location to specifically focus on in future analysis.

Chapter 7

An emergent behaviour approach to path planning for autonomous vehicles

Path planning is a fundamental problem in autonomous driving. Methods that address this problem need to incorporate the complexities of human decision making, vehicle dynamics, and some sense of safety, comfort and progress to choose the optimal actions in a given scenario. In this chapter, we consider an emergent behaviour approach to path planning, designing a fully interpretable cost function with components expressing concepts of safety, progress and comfort. After suggesting functional forms to measure each component, we include minimal hard-constraints, representing legal requirements, and then consider what behaviours can emerge in typical driving scenarios from such a model. We show our model naturally produces vehicle following and overtaking with and without the presence of oncoming vehicles. So-called advanced driving techniques emerge from the model, in-particular driving off the centre line of the road and allowing greater visibility for sensing. Further, when our predictions of actor behaviour prove to be incorrect, we show that the model still recovers safe behaviour, and obeys legal and comfort limits. This chapter is an exploratory analysis of what emergent behaviours can be seen from such a model, with clear examples of interpretability and control.

7.1 Introduction & problem relevance

Path planning for autonomous vehicles, in-particular vehicles that will use public roads, is a complex task. The challenge of encoding human driver behaviour, refined over years of training and development, into an algorithmic set of rules is a task that is subject to extensive ongoing development [186]. There are a huge number of variables that could effect the 'optimal' path in a

scenario, and indeed numerous ways one could measure optimality of a path. The first factor we highlight is uncertain behaviour of other drivers, pedestrians, cyclists and so forth (actors) in a scenario (or scene). This is comprised of two components: what is physically possible from a kinematic standpoint for a given actor, and what is likely to occur in practice. A second factor, highly coupled to this first is some measure of safety, and exactly what safety means in the context of driving. A further minimal requirement would be progress from some origin to some goal over the course of the path. Other factors determining the optimality of a path include the level of comfort passengers experience, abiding by legal regulations, travel-time, emissions, fuel-usage and various other considerations.

Machine learning provides one way to approach this task, attempting to imitate an expert driver with training data providing examples of what such a driver would do in each scenario an autonomous vehicle encounters. At the other extreme is a full rules-based system, trying to encode desired behaviour into a rigid set of rules that can be executed and checked in all scenarios. The approach discussed in this chapter is between these two extremes. We formulate an optimisation problem to determine the optimal actions up to some look-ahead horizon, minimising a cost function that accounts for some of the different components of desirable path planning. Our model ultimately presents a model predictive control approach to solving the coupled planning and control problem, from which we show common driving behaviours are re-created and controlled by varying components of the cost function. We retain some level of interpretability through manually designing a cost function, yet we have no rules dictating what to do in a given scenario and how to do it, rather behaviours are emergent from the scenario and choice of cost function. In-particular, we control the optimal steering rate and longitudinal jerk, which in turn determine a set of vehicle states into the future.

Our original contributions are as follows. Firstly, we explicitly incorporate distributional forecasts of actor behaviour into our planning, which naturally incorporates the uncertainty in how a situation may develop. Secondly, we avoid explicit specification of how an overtake should look, how far to follow vehicles and so forth, instead letting the minima of our cost function control what behaviour emerges. We are therefore not attempting to define different sub-systems for different manoeuvres, instead we design a single cost function that allows a range of desired behaviours to emerge. Thirdly, we propose to weight cost components using a distance dependent scheme, allowing components of the vehicle state to vary significantly from the target state when far away from the target location. In practice, this means that one can drive at high speed towards a distant target, and only slowdown to some lower, target velocity when near to the target location. Finally, we retain interpretability of our

model despite it allowing for various complex emergent behaviours. If one were to train a machine learning model to imitate an expert driver, then show it a new scenario, it may result in a particular manoeuvre, however it extremely difficult in the autonomous driving setting to truly determine why such a black box model resulted in that driving behaviour. This is a consequence of the huge number of features that such a model would use. By explicitly representing practical engineering concerns relating to driving, we are able to interpret the behaviour of our model, and indeed show that one could raise practical questions and adjust the model to meet these. Examples of this can be seen throughout section 7.4. A specific example could be an expert saying 'we are following vehicles too closely' and explicitly being able to control this by adjusting the cost. It is unclear how one would do the same with a machine learning approach without simply collecting significantly more data demonstrating the desired behaviour.

Comparisons between a machine learning approach and the one we pursue encapsulate compromises made by each. To utilise the machine learning framework, one has to collect significant amounts of data demonstrating the behaviour of an expert driving in a range of practical scenarios that will be encountered. Doing so is time consuming and carries a high financial cost, however the policy learnt has enormous freedom in its power to approximate the training data. On the other hand, one could formulate a cost function that incorporates the components expert engineers value most, and embed a significant amount of prior knowledge into the model itself. Doing so reduces the freedom of the model, however it will also lead to orders of magnitude fewer parameters to fit, and hence require significantly less data to tune. We see our method as a balance between a full rules based approach, in which the model has no freedom, and the complex machine learning approaches. We limit the freedom compared to a machine learning approach, but we never specify any rules for manoeuvres. A significant drawback of such a method however is that solving an optimisation problem at every step of a drive is far more computationally demanding than passing data through a neural network. One could envision a scenario where a machine learning model is trained on results from a well calibrated emergent behaviour model, avoiding the computational complexity of solving a full optimisation problem and requiring less real world data than a standard machine learning model. There is additional scope for methods to augment each other, and there are many future directions to investigate.

The cost function we formulate makes explicit use of distributional predictions of other drivers behaviours. This is not often used in the literature. There is some discussion of it and application in [103], constructing a distribution from data and then using it to design a reference path for an overtake. We argue

that if one trains such models for every type of object one might encounter, these allow for emergent behaviours regardless of specific manoeuvre. Further consideration of other vehicles uncertainty is considered by NVIDIA in [187], where a 'safety potential' is designed in order to determine occupied sets of space and trajectories. The idea is then to choose some subset of state space which the controlled vehicle can occupy that will ensure some safety margin to other vehicles and avoid any collisions. Similar work on evaluation of safety is conducted by Mobileye in [188], where 5 'common sense' rules are used to formulate some sense of safety. This is an extensive work considering sensing errors, different road geometry and right of way, which we do not consider in our model. We also have no concept of common sense rules, instead our actions are purely determined by the predictions of other vehicles and our cost function minima.

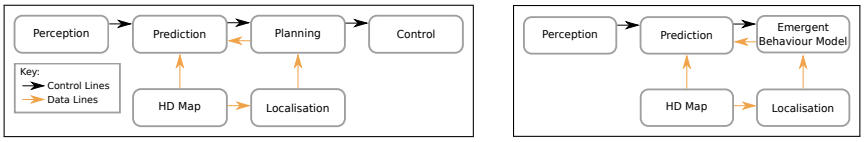
The remainder of this chapter is structured as follows. Firstly, we discuss the components of autonomous driving, detailing where our model would fit into the existing framework. After, we detail our methodology, costs and constraints to formulate an optimisation problem. We then show how our method performs in a number of typical driving tasks: vehicle following, passing and overtaking, and how it copes with abnormal actor behaviour. To end we summarise our findings.

7.2 Components of autonomous driving

Autonomous driving is typically separated into distinct components, each with clear functions. Those relevant to our work are often referred to as 'perception', 'prediction', 'planning' and 'control.' Perception deals with the identification of actors in a scene using data from on-board sensors: cameras, light detection and ranging (LIDAR), radar and so forth. In our work, we take the output of the perception module to be object level data, that is identification of explicit cars, pedestrians, cyclists and other objects in a scene. Prediction then takes this object data and outputs a multivariate probability distribution for each object, specifying the probability an actor will be in a particular state at a given future time. This distribution should be defined over all feasible states from a kinematic perspective, however of course many of these will be extremely unlikely due to social factors.

After prediction, a planning step is typically performed, where a set of coordinates are generated that is deemed the path to follow into the future. We detailed how such paths can be generated in section 2.4, however in short it is common to separate generating a path and following the path into two tasks. Given a reference path, the goal of control is to select actions that ensure the vehicle follows this path as closely as possible. Such a work-flow is

typical in this domain, with a major example being the open-source 'Apollo' platform [189]. In Fig. 7.1, we show how the relevant parts of the software overview in Apollo relates to our work. In such a framework, 'HD Map' refers to an assumed available map of the domain a vehicle is driving on, along with information such as speed limits, lane markings and instructive signs. 'Localisation' is the process of determining the current location of the vehicle in this mapped domain.



(a) Relevant components of the Apollo software overview.

(b) Where our model fits in the autonomous work-flow.

Figure 7.1: Example of the Apollo software overview relevant to our work, and a depiction of where our model fits in this framework. Control lines show the flow between stages of the autonomous driving platform, and data-lines show how environmental, intended actions and positional information are used by different components of the framework.

Our model sits after prediction, combining planning and control into one task. We take the prediction output, along with knowledge of the domain, and instead of generating a reference path, we choose the control actions to optimise some criteria. These actions are dynamically updated over a receding horizon in time using model predictive control (recall MPC from section 2.4). As a result, our path is emergent and only exists up to some look-ahead time. However, we will see that this allows us not to plan any particular manoeuvre, for example an overtake, but instead have them occur if they are optimal.

One point to note is that we envision some hardware 'watchdog' will exist between our autonomous interface and the vehicle, ensuring a last-line of safety against any problems that may emerge in the system. This is common and included in the full referenced Apollo framework. From this point forward, we shall use the common terminology in autonomous vehicles of referring to our own controlled vehicle as the 'ego' vehicle and all others as actors. The final point relevant to autonomous domain knowledge is that hard constrains on jerk, acceleration and steering angle used in section 7.3.2 are informed by ECE79-01 regulations found in [190] and by reasonable values from an experienced engineering perspective.

7.3 Methodology

Our approach to the problem of path planning relies on two novel aspects, the first of which is utilising distributional predictions of future actor states. A

fundamental assumption we make is that the prediction module will yield a set of PDFs that describe actor behaviour at all time-steps we look into the future. One such PDF should be available for each actor in a scene, or a single combined density function, the choice is entirely up to the modeller and can therefore be an expression of the actors individual intents, or a combined distribution considering how each will impact the other. We shall denote the probability density functions for actor $\mathcal{A}$, at time t as $\mathbb{P}(x, y, \mathcal{A}, t)$. Various models for this can be used, however for proof of concept we shall use multi-dimensional Gaussian distributions throughout this chapter. We keep the framework general to allow for future incorporation of more complex distributions. Notice that whilst we specify this distribution in space (x, y), it can also incorporate other attributes such as speed and acceleration.

Secondly, we do not specify a reference trajectory to follow and measure our deviation from this. Instead, we consider if placing minimal physical constraints on a vehicle, along with costs representing progression, safety and comfort are sufficient to reproduce intuitive driving behaviour when optimising our actions using MPC. As a result, the path we will take only ever exists up to some prediction horizon. Doing so however ensures we do not constrain ourselves to following the shortest route along a domain, rather we choose one that is a compromise between each of the components we consider important. Finally, our main goal is to question what behaviours are emergent from a model in which we enforce minimal rules, and incorporate practical engineering concerns. We aim for such a model to be useful in a range of scenarios rather than ever designing an overtake, following behaviour and so forth.

Developing a model for this problem involves a number of steps, and we detail each in the remainder of this section. We describe the model of the vehicle used in section 7.3.1, the hard constraints imposed on the optimisation in section 7.3.2, the cost function constructed in 7.3.3 and the optimisation problem in section 7.3.4.

7.3.1 Vehicle model

There is a wealth of literature on modelling vehicle behaviour. For our purposes, we require a model that takes some action inputs $\boldsymbol{u}$, an initial state $\boldsymbol{x}_0$ and generates a new state $\boldsymbol{x}$. Formally this is

$$\boldsymbol{x} = f_{\mathrm{VM}}(\boldsymbol{u}, \boldsymbol{x_0}). \tag{7.1}$$

A large number of the works reviewed in chapter 2 choose a specific form of f_{VM}, and general models of vehicles are reviewed in [191]. We choose to use the kinematic bicycle model in our work, specifically the model tested in a real-world setting in [192]. Further application of kinematic models is given in

[193] and they are used in the Apollo framework in [194]. Despite the simplicity, it is stated in [195] that such models are preferred for autonomous driving control design due to their computation speed and that the error from using one is not hugely detrimental to the final optimisation result. Whilst we ignore significant amounts of complexity relating to the vehicle dynamics, we propose this is sufficient for an initial exploration of our modelling approach. The key assumptions the model makes are: the vehicle is steered from the front wheels, slip angle is negligible, the front wheels are collapsed into a single steering point, the rear wheels are collapsed into a single point and the vehicle only moves in x, y space. Such assumptions are reasonable at low-speeds and in typical conditions, for example where no sudden large steering commands are input. Whilst some of the speeds we consider in scenarios may mean tyre forces become relevant from a practical stand point, we do not consider scenarios with high-speed, sharp turns where they could dominate the dynamics.

The system of equations for such a model is

$$\begin{aligned}
\frac{dx(t)}{dt} &= v_x(t)\cos(\phi(t)) \\
\frac{dy(t)}{dt} &= v_x(t)\sin(\phi(t)) \\
\frac{d\phi(t)}{dt} &= \frac{v_x(t)}{L_V}\tan(\delta(t)) \\
\frac{dv_x(t)}{dt} &= a_x(t) \\
\frac{da_y(t)}{dt} &= \frac{v_x(t)}{L_V}\left(2a_x(t)\delta(t) + v_x(t)\Delta\delta\right) \\
\frac{da_x(t)}{dt} &= j_x \\
\frac{d\delta(t)}{dt} &= \Delta\delta.
\end{aligned} \tag{7.2}$$

This is the model and notation used in [192]. Given a start state and set of actions to take, we can determine the state of the vehicle at any given time, being the position in space $x(t)$, $y(t)$, the heading $\phi(t)$, longitudinal velocity $v_x(t)$, steering angle $\delta(t)$, longitudinal acceleration $a_x(t)$ and lateral acceleration $a_y(t)$. We control the longitudinal jerk j_x and steering rate $\Delta\delta$, and the goal is to choose these values to create some optimal path through a domain. The final variable in this model is the vehicle length L_V. We note that one could choose to control the steering angle, steering rate, or its higher derivatives, and similarly one could control the longitudinal acceleration, the longitudinal jerk or higher order derivatives. Throughout the literature, it is common to choose either steering angle or rate to control direction, and longitudinal acceleration or jerk to control longitudinal velocity. We are informed by industry experts that going to any higher order derivatives presents a problem that from an

engineering standpoint would be very difficult to actually execute. To check our results are not highly dependent on the model used, we also run all of our tests but assume we directly control the steering angle and longitudinal acceleration, and see the same emergent behaviours and conclusions hold throughout all parts of this chapter. Details of this are in appendix F.

7.3.2 Hard constraints

When planning a path, clear physical and legal requirements are present and must generally be abided by. The first such constraint is that the vehicle must always occupy drivable space. To do this, we first define the section of our domain that is legal to drive on, and then constrain the vehicle positions to lie inside of this. For simple domains, this is just a linear constraint on $x(t)$ and $y(t)$, however for more complex domains one can utilise methods such as the 'point in polygon' algorithms to check arbitrary shaped domains. The second obvious constraint to consider the the speed limit. This is practically implemented by placing an upper bound on the state component representing longitudinal velocity, and we also place a lower bound representing the physical reversing capabilities of the vehicle. The drivable region and speed limit is assumed to be known from the HD map at all times.

Further hard constraints are required due to legal regulations specifying allowed states on the vehicle and an absolute minimum level of comfort for passengers. These take the form of limits on lateral and longitudinal acceleration and jerk, detailed in table 7.1. We note that in [196] it is stated that kinematic

Quantity	Unit	Lower Value	Upper Value
Longitudinal acceleration - a_x	$\mathrm{m/s^2}$	-0.4×9.81	0.2×9.81
Longitudinal jerk - j_x	$\mathrm{m/s^3}$	-3	2
Lateral acceleration - a_y	$\mathrm{m/s^2}$	-3	3
Lateral Jerk - j_y	$\mathrm{m/s^3}$	-5	5
Steering angle - δ	radians	$-\frac{3\pi}{18}$	$\frac{3\pi}{18}$

Table 7.1: Constraints enforced on the vehicle. These are informed by legal requirements and experienced engineering perspectives. The steering angle constraints represent 'full lock' on a vehicle, being roughly ±30 degrees. One might consider these constraints as a 'comfort profile' in general terms.

models retain validity when lateral acceleration values are kept below $0.5 \times 9.81 = 4.905\,\mathrm{m/s^2}$ which is enforced by these constraints. For each quantity in table 7.1, we can compute the model values using Eq. (7.2).

The final hard constraint placed on the model relates to a minimum expected safety throughout the journey. At any point throughout a journey, we should

not expect to collide with another vehicle. To enforce this, we choose to set some area of space to be impassable, changing through time. At any horizon, we set the impassable space to be an ellipse placed over the largest mode of the probability distribution of actor positions at that time. The dimensions of this ellipse are chosen such that if the actor vehicle is in the position we believe most likely, we will not be in-contact with it when outside of the ellipse. We note that this component is purely a consequence of our model being point-based, and to enforce a realistic dimension of the vehicle and other actors. The safety component it provides is very local, and the cost function determines the majority of the safety, discussed further in section 7.3.3. If the prediction distribution is a bivariate Gaussian, quantiles of this distribution will be an ellipse, so one can choose to specify a quantile level α of the forecast distribution instead of a physical size for this constraint. However, care must be taken that, when predictions are very certain, for example at very short horizons, the hard constraint should still be sufficiently large to cover the physical size of the vehicles. Finally, for more complex situations with multi-modal distributions, one could assign a hard constraint only to modes above some threshold in probability, which is more akin to choosing a particular quantile of the distribution.

7.3.3 Costs

Whilst our constraints express required properties of a trajectory, our costs express desired properties. These are incorporated as a weighted combination of multiple separate components, many of which are often in conflict with each other. We now detail the form of the cost function and the practical and engineering intuition for each choice.

The most basic requirement of a trajectory, before any of the environment is considered, is to progress towards some target or goal state. We consider a target state to constitute a 4-element vector, specifying a position $(x_{\text{target}}, y_{\text{target}})$, a heading ϕ_{target} and velocity v_{target}. All of our testing is done on a highway and A road scenario, so we consider x to progress along the road, and y to describe movement between lanes. If we set x to be the major direction of travel, we want to progress as much as possible along this direction in each time-step to minimise journey time. At any time t, we will be at position $x(t)$, and if we made optimal progress in a time-step, the largest distance we could cover is $v_{\max}\Delta t$. As a result, the lowest possible distance from ourselves to a target in a given time-step will be

$$x_{\text{target}} - (x(t-1) + v_{\max}\Delta t)\,. \tag{7.3}$$

However, in practice we will have ended up at a distance of

$$x_{\text{target}} - x(t). \tag{7.4}$$

Therefore, we know that the difference between our actual progress and the optimal progress is

$$\begin{aligned} \Delta\text{prog}_x &= [x_{\text{target}} - (x(t-1) + v_{\max}\Delta t)] - [x_{\text{target}} - x(t)] \\ &= x(t) - x(t-1) - v_{\max}\Delta t. \end{aligned} \tag{7.5}$$

If we square and sum over all times, and divide by some characteristic value $\tilde{x}$, we have a progress function in the major direction of travel as

$$C_{p_x} = \sum_{t=1}^{N} w_{x,t} \left(\frac{x(t) - x(t-1) - v_{\max}\Delta t}{\tilde{x}} \right)^2 \tag{7.6}$$

where $w_{x,t}$ are weights placed on the progress at each time point t, and we look ahead N time points. If we are able to reach the target in a time-step Δt, we can adjust this accordingly. The future positions can be computed by simulating the vehicle model forward in time given a set of input actions over the horizon.

For all other target components, we can express simpler functions, of the form

$$\begin{aligned} C_{p_y} &= \sum_{t=1}^{N} w_{y,t} \left(\frac{y_{\text{target}} - y(t)}{\tilde{y}} \right)^2 \\ C_{p_\phi} &= \sum_{t=1}^{N} w_{\phi,t} \left(\frac{\phi_{\text{target}} - \phi(t)}{\tilde{\phi}} \right)^2 \\ C_{p_v} &= \sum_{t=1}^{N} w_{v,t} \left(\frac{v_{\text{target}} - v(t)}{\tilde{v}} \right)^2. \end{aligned} \tag{7.7}$$

Here, we simply compute the difference between the desired value and actual value of our state components, normalise by some characteristic values $\tilde{y}, \tilde{\phi}, \tilde{v}$ and then apply some weighting to each of these in time.

The need for weights is clear when we consider a situation far away in x, y that we want to arrive at a greatly reduced speed, for example starting at 30 miles per hour on a road, and ending parked on a driveway. If we equally weight all components of this progress, we may enter a solution where we slow down far earlier than necessary, as we would achieve the target velocity, whilst still progressing in all other values. Instead, we can specify that as we move closer to our target, we care more about particular components, namely the velocity, heading and being in a particular lane, whereas when we are far

away, we prioritise moving quickly towards the target. For an example of this phenomena, view the results in section 7.4.1. Whilst it is common to use weight matrices in other MPC works for autonomous driving, we propose to make these weights functions of distance to the target. In-particular, they are either constant functions, or increase as distance to the target decreases. To compute this distance at any time-step, we take the current optimal set of actions and determine the distance to the target position $(x_{\text{target}}, y_{\text{target}})$ over the horizon. When we solve the optimal control problem at the following time, we use these expected distances to determine weights. Denoting the expected distance at time t as $d(t)$, we specify weights of the form

$$w_{v,t} = \alpha_v e^{-\beta_v d(t)} \tag{7.8}$$

for some constants α_v, $\beta_v > 0$ and $d(t) > 0 \, \forall \, t$. The parameters α_v, β_v can be tuned according to the preparation time we believe should be taken to slow down, the conditions around the vehicle and the situation. One may also want to limit these by specifying a maximum value any weight can take to avoid issues when very close to the target.

The second aspect we impose a cost on is comfort, which we measure through lateral and longitudinal jerk values. We can limit jerk values by specifying costs of

$$C_{j_x} = \sum_{t=1}^{N} w_{j_x,t} \left(\frac{j_x}{\tilde{j}_x} \right)^2, \;\; C_{j_y} = \sum_{t=1}^{N} w_{j_y,t} \left(\frac{j_y}{\tilde{j}_y} \right)^2 \tag{7.9}$$

These supplement the hard constraints on comfort in section 7.3.2, minimising jerk if possible to improve passenger comfort.

The third aspect vital to journeys is safety, and whilst we have enforced a minimal level of it, we can further refine it by including a cost based on unsafe actions. Given we have a probability density function describing where we expect actor $\mathcal{A}$ will be over time, we can incorporate this directly through standard probability arguments. Considering any time t and actor $\mathcal{A}$, we can compute the probability the ego vehicle is at the same location as an actor with

$$\int_{x_{\min}}^{x_{\max}} \int_{y_{\min}}^{y_{\max}} \mathbb{P}\left(x, y, \mathcal{A}, t\right) dxdy, \tag{7.10}$$

where we are integrating over some bounds that describe the physical space occupied by the ego vehicle. This alone, summed over all time and actors is a candidate safety cost function. However, we can also penalise 'large' collisions, in the sense of avoiding hitting a vehicle with a very large velocity difference, by considering our own expected velocity and a prediction of the actors expected velocity. If we take another output of prediction to be expected velocities, then

we have some estimate of an actors velocity at each time, which we will denote $\tilde{\boldsymbol{v}}$. We can then compute the expected velocity difference between the ego and actor, and for any path can specify

$$C^{*}_{\text{Safety}} = \sum_{\mathcal{A} \in Actors} \sum_{t=1}^{N} ||\boldsymbol{v}(t) - \tilde{\boldsymbol{v}}(\mathcal{A}, t)||^2 \int_{x_{\min}}^{x_{\max}} \int_{y_{\min}}^{y_{\max}} \mathbb{P}\left(x, y, \mathcal{A}, t\right) dxdy. \quad (7.11)$$

In Eq. (7.11), we see that this safety cost avoids occupying the same area that we expect other actors to be in at any time. Note the square on the velocity differences, representing the kinetic energy involved in collision dynamics. Here, the region we are integrating over would change in time as we simulate where the ego vehicle would be after applying control actions. In practice, evaluating this integral slows down the optimisation, and for sufficiently well behaved density functions we can replace it with the actor PDF at some time t evaluated at the ego's expected position $x(t), y(t)$, reducing to

$$C^{**}_{\text{Safety}} = \sum_{\mathcal{A} \in Actors} \sum_{t=1}^{N} w_{\text{saftey},t} ||\boldsymbol{v}(t) - \tilde{\boldsymbol{v}}(\mathcal{A}, t)||^2 \mathbb{P}\left(x(t), y(t), \mathcal{A}, t\right). \quad (7.12)$$

Here, we have also multiplied by a weighting factor for each component of the safety function in time that incorporates characteristic values for the components. Note however that we have a minimum of this function when we have equal velocities. This means, we could find ourselves in a situation where we occupy a region of space that we believe is very likely to also be occupied by an actor. However, if we have the same velocity as we expect them to, we would consider it safe. Of course, this is not desired, and to alleviate this, we can write

$$C_{\text{Safety}} = \sum_{\mathcal{A} \in Actors} \sum_{t=1}^{N} w_{\text{saftey},t} \left(\zeta + ||\boldsymbol{v}(t) - \tilde{\boldsymbol{v}}(\mathcal{A}, t)||^2\right) \mathbb{P}\left(x(t), y(t), \mathcal{A}, t\right) \quad (7.13)$$

where we perturb the minima by some $\zeta > 0$, allowing us to pull away from an actors expected position in a rare but extreme case. Here there is now always a penalty for both passing actors at high speeds and occupying the same areas they are likely to occupy.

The final cost function is then given by a linear combination of components,

written as

$$C_{\text{total}} = \begin{bmatrix} w_{p_x},\, w_{p_y},\, w_{p_\phi},\, w_{p_{v_x}},\, w_{j_x},\, w_{j_y},\, w_{\text{safety}} \end{bmatrix} \begin{bmatrix} C_{p_x} \\ C_{p_y} \\ C_{p_\phi} \\ C_{p_{v_x}} \\ C_{j_x} \\ C_{j_y} \\ C_{\text{Safety}} \end{bmatrix} \tag{7.14}$$

where we can further weight the overall contribution of a component with the values w_{p_x}, w_{p_y}, w_{p_ϕ}, $w_{p_{v_x}}$, w_{j_x}, w_{j_y} and w_{safety}.

7.3.4 Model predictive control formulation

Given our model, constraints and costs, we formulate our optimisation problem as

$$\begin{aligned} \underset{\boldsymbol{u}}{\text{minimise}} \quad & C_{\text{total}} \\ \text{with} \quad & \boldsymbol{x} = f_{\text{VM}}\left(\boldsymbol{u}, \boldsymbol{x_0}\right) \\ \text{s.t.} \quad & v \in [v_{\min}, v_{\max}] \\ & (x, y) \in \text{Domain} \\ & (x, y) \notin \text{Actor Constraints} \\ & a_x \in [-0.4 \times 9.81, 0.2 \times 9.81] \\ & a_y \in [-3, 3] \\ & j_x \in [-3, 2] \\ & j_y \in [-5, 5] \\ & \delta \in \left[-\frac{3\pi}{18}, \frac{3\pi}{18}\right]. \end{aligned} \tag{7.15}$$

We solve the above problem using MPC in Matlab 2019a. In the MPC framework, we look some horizon ahead, and optimise the control actions $\boldsymbol{u}$ over that horizon. We then apply the action for one time-step and re-solve the optimisation problem using the new step, new predictions from the actor models and new cost function. In a receding horizon fashion, this dynamically updates our path to adjust the the varying uncertainty we might experience in actor behaviour. Constraints are required to be satisfied across the entire horizon. Numerically, we solve the optimisation problem at any time by using an interior-point (IP) method [197]. A review of such methods is found in [133], and an open source implementation is available in [198]. We offer some discussion of the method in appendix A.

7.3.5 A note on the safety cost component

A point of obvious note is that we are considering a safety cost that takes values from some distributional prediction of actor behaviour, and assume that the prediction component of the autonomous work-flow would output these. Recent work focusing on predicting a distribution of potential future states in autonomous driving is found in [199] and [200], and more generally for distributional trajectory prediction in [201].

7.3.6 Modelling summary

One can incorporate each of the steps we have described into a single model and, given some driving scenario, iteratively determine a route through the scenario. We write the pseudo-code for applying this model in algorithm 2. We determine the distance between the current state and target state using some $\mathcal{D}$, which could be the euclidean distance, or some other choice if different components are more important for a particular scenario. We of course note

Algorithm 2: Iterative path planning

Input: Prediction model(s) $\mathcal{M}_{\mathbb{P}}$, Cost function with associated weights and parameters C_{total}, Initial state $\boldsymbol{x}_0$, Target state $\boldsymbol{x}_{\text{target}}$, Stopping threshold ϵ, Time-step Δt, Hard constraints $\mathcal{H}$

Set current state: $\boldsymbol{x} = \boldsymbol{x}_0$
while $\mathcal{D}\left(\boldsymbol{x} - \boldsymbol{x}_{target}\right) > \epsilon$ **do**
    Get actor states from perception module: $\mathcal{A}_s$
    Get future state predictions: $P_{\mathcal{A}} = \mathcal{M}_{\mathbb{P}}\left(\mathcal{A}_s\right)$
    Determine actions $\boldsymbol{u}$ by solving Eq. (7.15) subject to $\mathcal{H}$
    Apply control actions $\boldsymbol{u}$ to vehicle for time-step Δt
    Localise vehicle to determine $\boldsymbol{x}$
end

that the prediction models $\mathcal{M}_{\mathbb{P}}$ may depend on some window of historic actor states, and take many more inputs from the environment than just the current actor state.

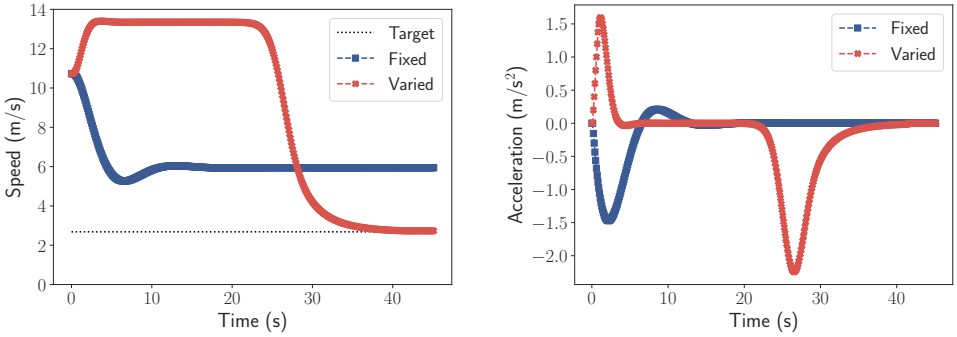
7.4 Results

Given our problem and methodology, we now specify a set of tests that show key features of the approach. We specify a time-varying uncertainty in all actor predictions, aligning with the common attribute of forecasting models that as we increase the horizon, we become more uncertain of predictions. Initially, the uncertainty is simply measurement error, but as we look further ahead in the MPC framework, we grow the variance to suggest reasonable uncertainty. Note that each of the examples shown are qualitatively similar for various choices of

horizons, but choosing a horizon so small (reasonably, less than 3 seconds) can yield situations where one can move into states that are actually detrimental in the long term.

7.4.1 Scenario 1: speed change

In section 7.3.3, we noted that distance-dependent weights were particularly important in the case of velocity progress costs. An important point in path planning is to ensure we are not too conservative, slowing down far too early and compromising progress for matching to a target final speed. Suppose we take a scenario where we drive in a straight line, with no other actors present. The ego vehicle starts at some (x_0, y_0) and $v_0 < v_{\max}$. The aim is to reach some $(x_{\text{target}}, y_{\text{target}})$ at a lower speed, $v_{\text{target}} < v_0$. We now set up such a simulation and run with two different weighting schemes in our progress function. The first is a fixed weighting on all components, whereas the second has a distance dependent weighting on velocity. Our results are differing profiles of acceleration and velocity, shown in Fig. 7.2.



(a) Speed profiles.

(b) Acceleration profiles.

Figure 7.2: Comparison of uniform and distance dependent velocity weighting in the progress function. We see the distance dependent scheme allows the ego vehicle to first speed up travel at a higher speed, then slow down towards the end of the simulation window and when the ego vehicle is nearer the target. The uniform one by comparison immediately moves to some comprise velocity, where there is a balance between optimal progress and matching the target velocity, leading to a situation where the ego vehicle does not reach the target speed and makes slower progress. Of course, one can change how quickly the target speed is reached with a weighting change.

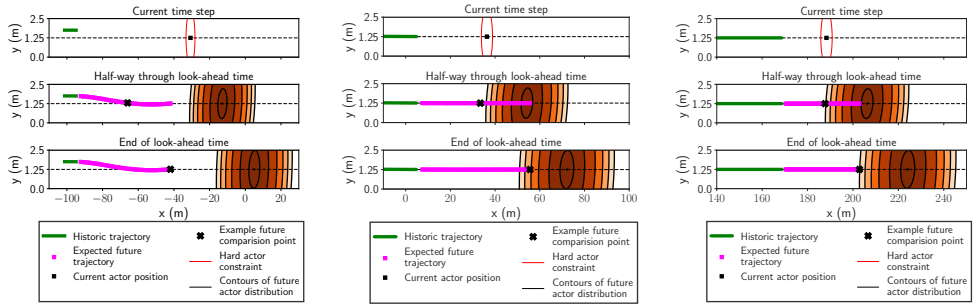
Of particular note in Fig. 7.2 is the fact that the constant weighting choice leads to a solution where the ego vehicle neither makes optimal progress in terms of distance to the $(x_{\text{target}}, y_{\text{target}})$, or matches the target speed throughout the entire simulation. This is due to the fact that the minima of the cost function coincides with some 'compromise' velocity, where some amount of progress is made and the speed of the ego vehicle is somewhat close to the target speed,

but a better solution for both of these components simultaneously cannot be found. The distance dependent case does not suffer from this, as we encode sensible restrictions on the model that in turn allow the ego vehicle to travel at higher speeds when progress towards $(x_{\text{target}}, y_{\text{target}})$ is most important, then brake when near the target and progress towards $(x_{\text{target}}, y_{\text{target}})$, relatively speaking, is less important. Whilst we observe a larger deceleration in Fig. 7.2b for the time dependent case compared to the uniform one, this can be dampened by appropriate weighting choices.

7.4.2 Scenario 2: vehicle following on a single lane road

The most basic test of any path planning method is to reproduce existing work that is more limited in scope. Consider the simple scenario of a single lane straight road, with some speed limit, say 13.41 m/s (30 mph). The ego vehicle drives down this road (for practical purposes, of a length that will take far longer to traverse than we simulate) and encounters a vehicle travelling at some lower speed, say 8.94 m/s (20 mph). In such a scenario, there is no opportunity to overtake the actor, so the ego vehicle simply has to slow down and ensure a safe distance to the vehicle in front. It can already be seen that such a scenario is simply longitudinal control, something accomplished in simple scenarios though adaptive cruise control. Our existing methodology should do the same, even if the basis of it may take more computation power. We test such a scenario, and present results in Fig. 7.3. Note that the only hard constraint we have relating to the actor is that we cannot occupy the ellipse surrounding its mode, and we have no other explicit concept of a 'safe following distance'. Instead we allow the probability distribution, or more specifically its variance, to yield a safe following distance.

When viewing each plot in Fig. 7.3, it is important to note that we have pictured 3 states of the actor, one at the current time-step, one half way through the look-ahead time and one at the end of the look-ahead time. Each of our future points is aware of a different forecast of the actor position and variance, and hence concept of safety. Note that here, and in all future plots of the same type we plot a fixed set of contours for the forecast distribution, which demonstrates the concept of a growing variance in time. If we compare the contour lines half way through the look-ahead time and at the end of the look-ahead time, we see those in the latter are wider than in the former. This shows that the variance is growing as we predict further ahead in time. Firstly, in Fig. 7.3a, the ego vehicle begins to move from the starting point towards the road centre and along the road. In Fig. 7.3b, it approaches the actor and starts to incur a safety cost by occupying regions of space that the actor may be in at future times. We see from Fig. 7.3c that later in the simulation,



(a) Example trajectory 0.5 seconds into the simulation.

(b) Example trajectory 8 seconds into the simulation.

(c) Example trajectory 25 seconds into the simulation.

Figure 7.3: Optimisation result for a vehicle following example. The green markers show the historic ego states, based on previous actions we have optimised and executed. The magenta markers show, at that current time-step, the positions that result from what we believe to be the optimal actions for each time-step in the future, extending from 1 time-step to the specified look-ahead time. In the top panel of each, we plot the current situation at the given time-step. We physically 'see' the actor through sensors and have a hard constraint which we cannot occupy. In the centre panel of each, we consider half-way through the look ahead time. At this point, marked with a black $\times$, our optimisation considers the distribution represented through the contours on the given plot, meaning the further the ego vehicle is from the centre of this distribution the safer it is. The dark zones should be considered the most unsafe. In the bottom plot, we consider the same but for the last point in our look-ahead time. As expected, the ego vehicle slows down as it approaches the actor. Note how the distance between the ego vehicle and actor is not 2 seconds of travel time, but instead relates to the probability distribution assigned to the actor. This is an example of emergent safety, following distances as a result of uncertainty.

the ego vehicle has settled into a steady state where it follows the actor and occupies regions of space that are highly unlikely to contain the actor both half way and at the end of the look-ahead time. Clearly, a particular following distance has emerged from the model and choice of weights, without a rule specifying it. Note that the ego vehicle has intentionally started off the road centre line in Fig. 7.3a to show a lateral deviation and emphasise this is not simply a reinvention of adaptive cruise control, but rather this is a subset of our methodology.

The associated acceleration and jerk values for this simulation are given in Fig. 7.4, showing how passengers in the ego vehicle would have felt during the drive. From Fig. 7.4, the ego vehicle clearly slows down and avoids a collision with the actor, and from Fig. 7.3 see that this corresponds to staying some distance away that stabilises as we slow down and begin to match the actor speed, hence giving a low safety cost. Both the decrease in velocity and the

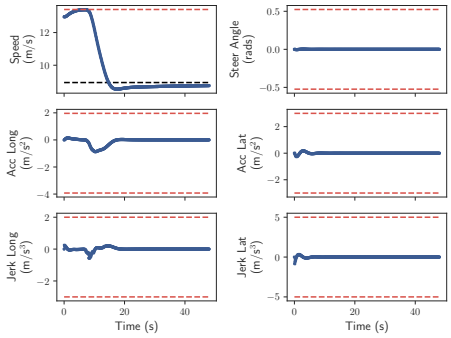


Figure 7.4: Constraints summary for a vehicle following example. The red dashed lines indicate hard constraint values, and the black-dashed line indicates the actor speed. The ego vehicle brakes when the actor is seen in the look-ahead time, and then slowly matches its velocity which results in very low jerk and acceleration values.

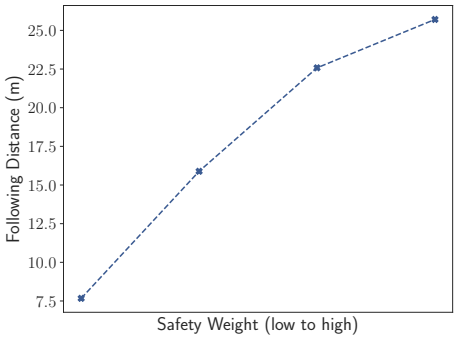


Figure 7.5: Comparison of following distances based on varying the safety weight from some low value to a high value.

distance to the actor contribute to the safety cost.

We further demonstrate the following distance as a function of the safety importance and variance by re-running the same simulation but varying the safety weight from some low value to a higher value, waiting for the ego vehicle to have reached a stable speed after encountering the actor, and plot the average distance between the ego vehicle and the actor for the remaining time in the simulation. Results for this are shown in Fig. 7.5. From Fig. 7.5, we see that a particularly low safety weight might lead to a following distance of 7.5 metres, whereas a particularly high one might lead to distance of 25 metres. Of course, coupled to this is the variation associated with the distributional prediction models. This also raises reasonable questions relating to models of autonomous driving: if we are very confident of a particular future state of an actor, we can potentially drive closer than a human naturally would, whilst still retaining some confidence level that we will not collide. Further, the opposite is true, if an actor is acting very unusually, we may hold further back because we simply are not confident where they will be in the future.

A final point of note from this vehicle following scenario is that some advanced driving behaviour is observed for particular weighting choices. Specifically, we observe a lateral deviation when following vehicles, pulling slightly off centre on the road, when we have a low weight on aligning to some y_{target} relative to our weight on safety. This is demonstrated in Fig. 7.6, where we show a frame of a simulation where the ego vehicle has reached a sustained behaviour driving slightly off the road centre-line. This behaviour is due to

the distribution we have placed over the actor position being Gaussian but decaying far faster in y (lateral direction) than in x (longitudinal direction). Essentially, we achieve a lower safety cost by moving lateral slightly. In doing so, we would then have a longer sensor range and it mimics so called 'advanced driving' techniques where one can see further ahead in traffic by driving off the road centre-line, but not so much so as to obstruct another lane. Note that

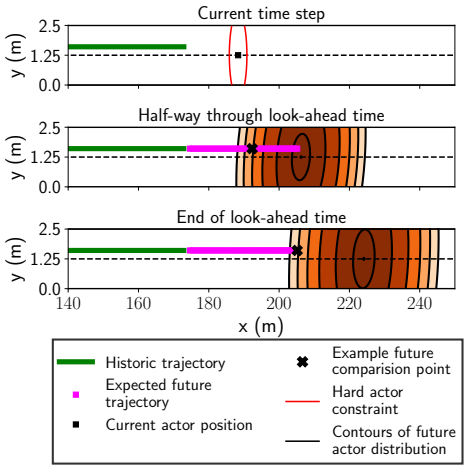


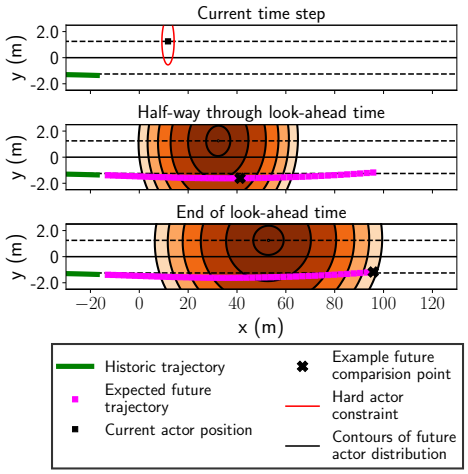
Figure 7.6: Optimisation result showing a concept of advanced driving. All details in these plots should be interpreted as explained in Fig. 7.3.

now we expect to be somewhat closer to the vehicle than in the previous case, due to the decay laterally of the PDF.

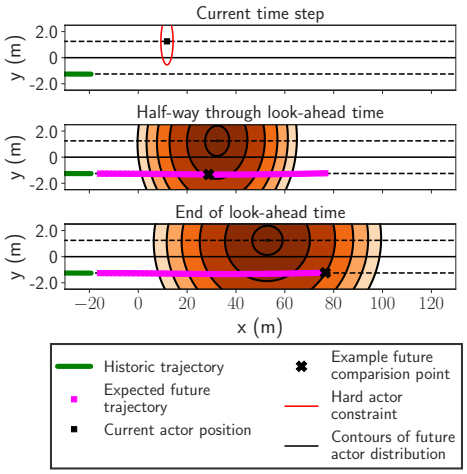
7.4.3 Scenario 3: vehicle passing on a dual carriageway

The next demonstration of our methodology is passing a vehicle. The domain is two straight lanes with the same direction of travel (imagine an A road in the UK) with the ego vehicle approaching a slow-moving actor (imagine a tractor in this example) in the left hand lane. We aim to continue along this road which will take far longer to traverse than the time of the simulation. The ego vehicle starts at some position x, y behind the actor in the right hand lane, and wants to pass the actor safely. Intuitively, we expect the ego vehicle to approach the actor, start to slow down creating a small velocity difference, pass the vehicle, then speed up once we are far from it to make good progress. We might also expect a movement away from the actor to maximise the passing distance, and then to return to the point aligned with the target y coordinate (in this case, the right-hand lane centre line). We set up such a simulation, again using a multivariate Gaussian distribution with growing uncertainty in time for the actor predictions. We choose the covariance matrix to cover both lanes as it grows but clearly favours the actor staying in the same lane. In Fig. 7.7 we visualise the solution of the resulting optimal control problem for

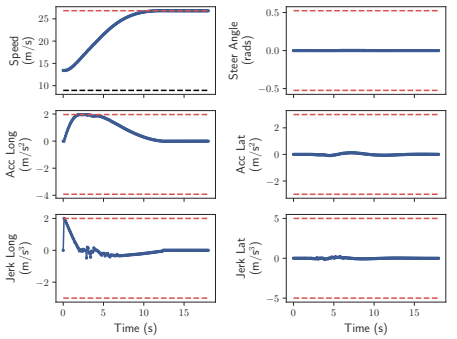
two weighting choices, and show a summary of the constraints after the entire simulation is complete.



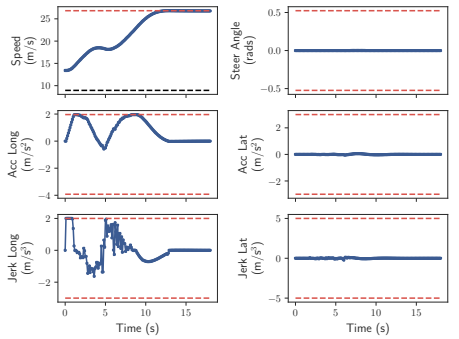
(a) Simulation 1 trajectory example, at the point at which a slight lateral deviation from the road centre-line is planned for the safest passing.



(b) Simulation 2 trajectory example. A lateral deviation is no longer planned but the actor is still passed.



(c) Simulation 1 constraints summary.



(d) Simulation 2 constraints summary.

Figure 7.7: All details in these plots should be interpreted as explained in section 7.4.2. We see a small lateral deviation during the pass in (a), keeping more of a distance when we expect the actor may move slightly across the lane. We see none in (b). The difference between these two trajectories is of course very subtle, but more is gained considering the corresponding constraint plots in (c) and (d). Inspecting these we see that in simulation 1, the ego vehicle does not slow down at all but use the slight lateral deviation to improve safety, whereas in simulation 2 it slows down to reduce the safety cost, and then does not laterally deviate at all.

From Fig. 7.7, we see how our simple model can recreate a very subtle but important component of human driving. The ego vehicle has a slight deviation from the road centre line, giving a small lateral jerk, acceleration and progress-y cost, but decreasing the safety cost in doing so. We see the slight bend is planned prior to being level with the actor, meaning using the current look-ahead horizon, the ego vehicle has a plan of the entire pass, but

can update this if the scene develops differently to that which is expected from the prediction models. Alternatively, if one considers passing speed to be more important, then the model can eliminate the turn and instead pass the vehicle with a lower velocity difference, and speed up after. Doing so incurs different penalties on lateral and longitudinal values.

Finally, we note that behaviours in-between these two shown are possible. One can deviate laterally and still slow-down for particular weightings, or ignore the influence of the actor entirely if progress and comfort is significantly more important. We plot four different trajectories across the road in Fig. 7.8, which gives an illustration of this. The distinction to make from Fig. 7.8 is

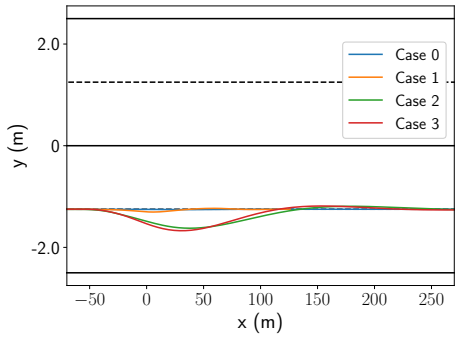


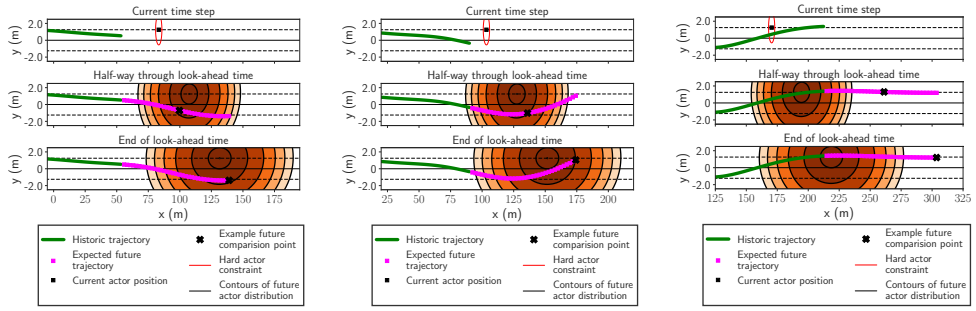
Figure 7.8: Comparison of trajectories for the ego vehicle during different weighting choices. We see two cases that stay on the centre line and two that deviate laterally.

that case 2 and 3 both deviate laterally, however the curve in case 2 does this after case 3, as the ego vehicle slows down in this instance as well as moving laterally to reduce the safety cost.

7.4.4 Scenario 4: vehicle overtaking on a two lane road

There are a number of papers that develop methods of vehicle overtaking and systems to support this. Often, this is a purpose built system, for example in [202], [203] and [204]. Indeed, it is stated in [202] that 'automation of the overtaking maneuver is considered to be one of the toughest challenges in the development of autonomous vehicles'. We argue that an overtake is a natural consequence of balancing costs and a distributional prediction of actor positions, showing an example here. In this case, the ego vehicle starts in the left lane of a two-lane domain, with a speed of 17.9 m/s (40 mph). Ahead is an actor with a speed of 13.4 m/s (30 mph) also in the left lane, and the speed limit on the domain is 26.8 m/s (60 mph). We have a target to be in the left hand lane, level with the centre line, at a large distance up the road, so naturally a human would approach the actor, move into the right hand to overtake and then return to the left lane. We show exactly this happening in

our model in Fig. 7.9. We further show a second weighting scheme that allows for a later cut-in back the original lane of travel in Fig. 7.10, valuing a safer drive.

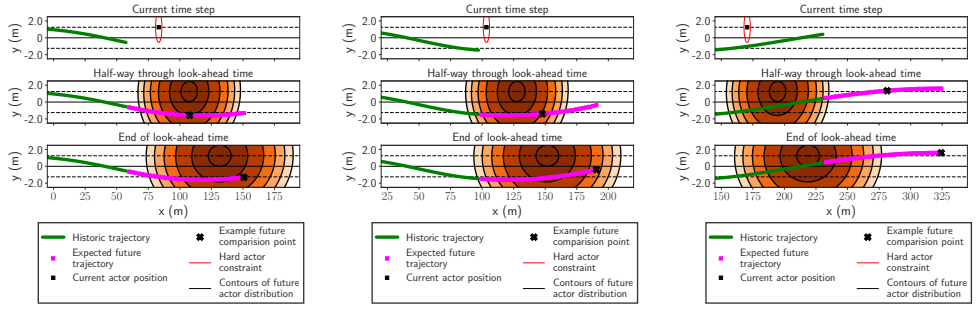


(a) Example trajectory. The ego vehicle plans to move alongside the actor during the look-ahead time.

(b) Example trajectory. The ego vehicle plans to cut back into the left lane ahead of the actor.

(c) Example trajectory. The ego vehicle has completed the overtake and continues to drive along the road towards a target.

Figure 7.9: Optimisation result for a vehicle overtake. We place a time-varying weight (as a function of distance) on y progress such that we care little about it initially, allowing the overtake. We move into the right hand lane, pass the vehicle and return to the left hand lane.



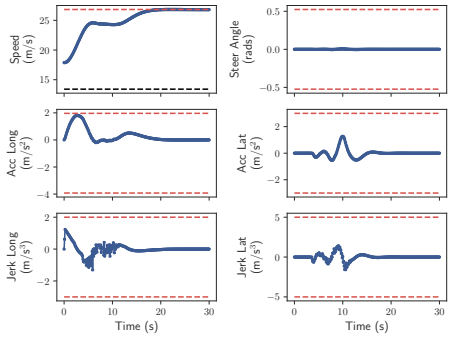
(a) Example trajectory. The ego vehicle plans to drive alongside the actor for a prolonged period of time.

(b) Example trajectory. We see the beginning of a plan to cut back into the left lane.

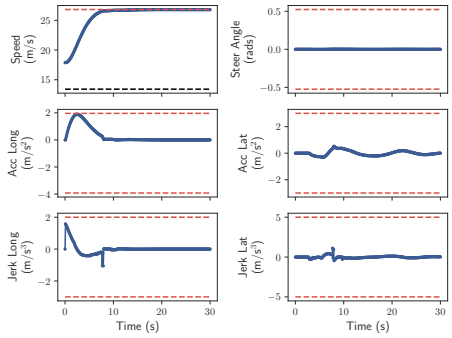
(c) Example trajectory. We see the end of the overtake is now planned.

Figure 7.10: Optimisation result for a vehicle overtake, showing the ego vehicle waiting longer to cut back into the original lane of travel after passing the actor compared to Fig. 7.9.

Whilst overtaking is a complex behaviour, we see from Fig. 7.9 that such a behaviour has emerged from our model. Altering the relative importance of safety, jerk and progress yields two distinct behaviours, changing the time the ego vehicle drives beside the actor for and when it returns to the original lane of travel. This, as well as the previous results for emergent following distances and passing behaviour, shows the potential of emergent behaviour models to capture complex driving behaviour with a minimal description of the scenario.



(a) Constraints summary shown for the overtake in Fig. 7.9.



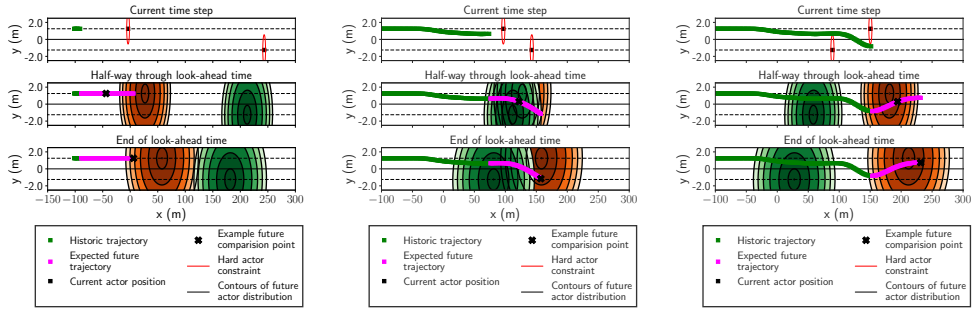
(b) Constraints summary shown for the overtake in Fig. 7.10.

Figure 7.11: Constraints summary shown for two weightings of overtake manoeuvres. All values remain inside the specified limits. Although we do apply a steering angle, we are no where near full lock and hence stay far from these constraints.

A summary of the constraints for the overtake scenarios are given in Fig. 7.11. A point to note from Fig. 7.11a is that the ego vehicle initially shows an increase in velocity to make good progress, but when the actor is encountered there is a clear time period where velocity stops increasing. Doing so not only increases the comfort of the overtake but also increases the safety, as the ego vehicle passes the actor with a lower velocity difference. The lateral distance moved during the overtake is controlled both by the weighting on the safety components and the progress on y, which both alter the compromises made in the cost function. One can therefore see how the model can be fine-tuned to account for practical concerns, for example passing vehicles too closely.

7.4.5 Scenario 5: multi-actor vehicle overtaking on a two lane road

A more complex version of an overtake is when there may be actors approaching in the opposite lane, in which case one has to plan a comfortable and safe overtake of a slow vehicle, but also avoid collisions with oncoming vehicles. Our framework generalises to this example, with our results shown in Fig. 7.12. When viewing Fig. 7.12, we see a clear waiting period where the ego vehicle allows the oncoming vehicle to pass before moving into the right-hand lane. Whilst we see quite a small lateral deviation for the overtake, this can be explained by the fact that we have placed distance dependent weights on the y progress. As the ego vehicle waits for the oncoming vehicle to pass, it moves closer to the target, and as a result the distance-dependent weight increases. When the vehicle has passed, there is a larger weight on y progress than in the previous example, still allowing the ego vehicle to overtake but not to



(a) Example trajectory, before the overtake begins.

(b) Example trajectory. The ego vehicle plans the overtake after it expects the oncoming vehicle to have passed in the look-ahead window.

(c) Example trajectory. The ego vehicle has pulled into the right-hand lane and plans to overtake and cut back into the left lane.

Figure 7.12: Optimisation result of an overtake, with another actor in the right-hand lane delaying our overtake. Imagine this as a B-road scenario in the UK.

deviate significantly laterally. As a result the ego vehicle also cuts back into the left-hand lane in a higher region of the prediction distribution than it did in the single vehicle case, however again tuning the weights to adjust for this is a possibility.

7.4.6 Scenario 6: abnormal actor behaviour

Throughout this chapter, we have used distributional predictions of actor states to determine what we consider to be safe behaviour in the future. However, there may be instances were forecasts are incorrect, and we realise at some time that an actor is doing something that we thought was highly unlikely just a short time prior. Suppose we take the vehicle following example in section 7.4.2, but part way through the simulation, the actor brakes suddenly, without warning and reduces their speed from 8.94 m/s (20 mph) to 2.24 m/s (5 mph). We allow this slow-down to be a $0.8g$ deceleration, and recall that our comfort profile specifies a $0.4g$ deceleration constraint (taking g to be 9.81). Results for this are shown in Fig. 7.13. For such an example, we require a prediction model, yet for this case we take the simplest possible model. At time t, we assume we can measure the velocity and acceleration of an actor, a sensible assumption for a vehicle equipped with a modern radar system. Then, our prediction model is simply to assume the actor will continue with their current acceleration profile into the future. We do not assume an actor will ever travel backwards, instead they will stop rather than reverse on a major road. Such a model is very simple and invalid in many scenarios, however it is sufficient to illustrate the models adaptation in abnormal circumstances.

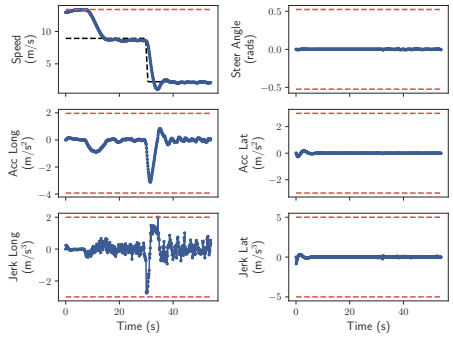
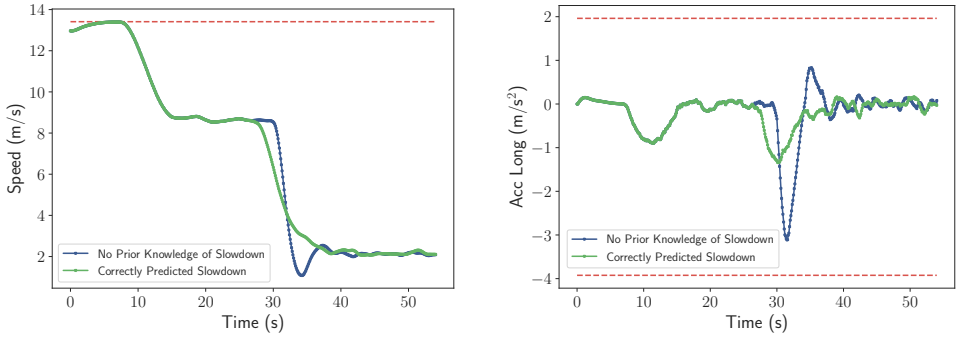


Figure 7.13: Optimisation result for vehicle following with abnormal actor braking. Recall as before that the black-dashed line indicates the actor speed. The ego vehicle only begins to slow down after the actor has, as it had no information from the prediction model that a slow-down was likely to occur. It overshoots the actor speed because of this, then returns back to a following speed similar to the actor's.

From Fig. 7.13, we see that to begin, the ego vehicle follows the actor and stabilises to some speed and distance. After some time, the actor brakes suddenly and far harder than the ego vehicle is allowed to (note that the actor instantly applies a $0.8g$ deceleration), and the ego vehicle reacts with a prolonged period of deceleration reaching around $-3\,\mathrm{m/s^2}$. It avoids both violating any hard constraints during the optimisation and hitting the actor. An obvious extension to this however is to ask how different would this behaviour look with a prediction model that somehow predicted the actor would slow down, and how would comfort be impacted in such a case? To investigate this, we run the simulation again however this time we specify the actor's distribution to reflect the true behaviour, seeing the slow-down as soon as it is within our prediction horizon rather than as it is happening. We then plot the resulting speed and acceleration profiles in Fig. 7.14.

From Fig. 7.14b we see that when we correctly predict the actor slow-down, even though it is harder than we allow the ego vehicle to brake, it is accounted for by holding a less extreme deceleration behaviour for an extended period of time. In the case where we did not predict it, the ego vehicle has to brake far harder and reach a higher deceleration value to account for the abnormal actor behaviour. Note also that from Fig. 7.14a when we correctly predict the slowdown, we do not overshoot the velocity, whereas we do in the abnormal case due to the high deceleration values required to avoid a collision and remain within our comfort constraints. These two cases in some sense represent the two reasonable extremes: a prediction that has no idea the actor will brake, and a prediction that knows the actor's behaviour perfectly. In reality, we might expect a situation between these two, with some knowledge from a prediction model informed by a combination of the actor states and environment around



(a) Speed results for abnormal actor behaviour.

(b) Acceleration results for abnormal actor behaviour.

Figure 7.14: Comparison of speed and acceleration results for a scenario with an actor slowdown, considering cases where we had no knowledge of the slowdown, and where we predicted it before it happened. The first 27 seconds of the two cases are exactly the same as the actor and prediction models are the same up to this point.

it. It also highlights that reasonable comfort can be sustained even when others perform actions the ego vehicle is unable to match within the specified comfort constraints, as long as we are able to predict them far enough ahead of their occurrence. As such, the ego vehicle is not mirroring what the vehicle ahead is doing, it is accounting for it and the specified comfort profile jointly.

7.4.7 Stability of future trajectories

In the framework of MPC, we not only solve for the optimal actions at the current time-step, we also consider some horizon τ^* ahead, yielding expected optimal actions at these steps. We find that our methods are reasonably stable in predicting future states for our toy problems, with examples of future expected states already seen in each scenario considered. The stability of this look-ahead planning directly relates to how the dynamic environment changes between time-steps, and choosing an appropriate look-ahead time is a compromise between computation power, desired comfort and safety, and the risk of entering a situation one cannot escape without a harsh manoeuvre. This look-ahead time however is coupled with the question of how far ahead can a predictive distribution actually yield informative, precise predictions in driving scenarios. Such a question is a topic of active research, however in some scenarios, there may be justification to believe that forecast horizons could be quite large for A road and motorway scenarios. Consider an example of an A road in the UK, and where we observe some actor ahead of the ego vehicle, driving near to the speed limit. Openly available data provided by various sources, one being NTIS, would allow one to determine the state of the road long distances ahead, and if we are not near to an exit or entry of

the A road, we can make sensible predictions of the future road state, and the actions an actor we observe locally might have to take. As an example, if we observe higher density of traffic some distance ahead that would not be visible by on-board sensors or human vision, but can be detected using the real-time data-streams openly available, we might then know the actor will have to slow down, and have a reasonable estimate of where and when this will occur, which in turn will allow us to make longer range predictions than only considering information locally obtained from the ego vehicle. Such an idea is referred to as 'V2X' or 'Vehicle-to-everything' in autonomous driving frameworks, and offers a clear way to make longer range predictions for actor behaviour, accounting for factors not immediately visible to the ego vehicle.

7.5 Summary & conclusions

In this work, we have explored what behaviours emerge as a consequence of abstract definitions of safety, progress and comfort in an autonomous driving setting. The novelty of this method compared to those already discussed in the literature is three fold. Firstly, we couple the problems of finding an optimal route in (x,y) with that of lateral and longitudinal control of the vehicle. In some cases, work assumes a given path and aims to best follow it with some optimal control approach. We instead only consider a local path, that is dynamically updated in a receding horizon style, and is impacted by growing uncertainty in predictions of actor states. Secondly, our cost function contains within it an abstract notion of safety based on both minimising velocity difference of potential collisions and a probabilistic interpretation of future actor states. Specifically, we have a hard constraint avoiding some region of space we deem highly likely for an actor to occupy, and the remaining space has a cost associated with occupying at various speeds which is determined by the uncertainty in the actor prediction. Doing so avoids specifying some cut off or influence distance of a potential function, and instead determines our actions based on our certainty of future actor behaviour. Since we dynamically update our path, we are able to account for predictions varying in certainty over time. Finally, the model retains interpretability, which can be seen through the concept of 'compromise', without resorting to a black box method or explicit rules. This allows practical concerns to be raised, for example 'we are following actors too closely' or 'we need to leave more lateral distance when overtaking' and for engineers to directly see how one might address them. It is perhaps a complex social and legal question to ask what the optimal compromise between safety, comfort and progress is, however from a modelling approach building this into the path planning allows for such discussions in the future. Our focus has consistently been on exploring emergent behaviours that recreate complex

driving tasks, rather than immediate applicability to a real-world vehicle.

We have shown how our approach generalises to vehicle following, vehicle passing, and overtaking both with and without oncoming traffic. Such behaviour is complex, yet we use a minimal description of fundamental quantities we value to promote it. This also avoids any sort of expert systems approach, not defining an 'overtake' algorithm and a 'following' algorithm, instead suggesting that driving behaviour is inherently a balance of compromises, and these novel behaviours and actions are implied by certain choices of compromise. In other phrasing, we have designed a system where some typical driving behaviours are emergent rather than rigidly enforced. A first specific example of emergent behaviours include a following distance emerging as a function of future actor uncertainty and how much we value safety relative to other components. A second is slowing down, moving laterally, or both when passing slow vehicles in the outside lane. Further, how early to overtake, when to return to the original lane of travel and avoiding oncoming traffic are also seen to be emergent behaviours that in turn define overtake manoeuvres. Finally, we have seen that even if actors perform actions that violate our comfort profile, we are still able to react within our constraints, and observed the importance of predicting such things ahead of time rather than being purely reactive to them.

It should be noted that in the considered cases, other drivers should be unimpeded by our actions, meaning we do not force them to change their planned actions, slow down and so forth. This is reasonable for many scenarios encountered in driving, however for more complex urban scenarios one may have to consider how our planned actions impact the predicted future actor states. In these cases, the prediction model would output a distribution, conditioned on a given input action. If we are forced to impact the actions of another driver, we should then consider how likely they are to alter their actions given our planned trajectory, which clearly relates to our prediction models and safety cost. Such cases would create another variable altering the cost function. However, for the considered scenarios and many typically encountered, we can reasonable perform our actions and plan a path that will not require any change of actions from other actors. Further, dynamically updating our decisions can account for the changing behaviour of others relative to our own.

In the future, one could consider heuristics to make this optimisation process more efficient and practical for real-time use. We do not claim the current method is real-time, however there are many refinements and adaptations that could be done to account for this from an implementation stand-point. An alternative and intriguing suggestion is to collect data from an expert driver in many scenarios, and parametrise the weights of our model on this data. Then, one could generate many orders of magnitude more driving examples and scenarios, run our already tuned model in them, and generate a large

synthetic dataset of optimal control actions and behaviours given a scene. Finally, one could then use this synthetic dataset to train a complex neural network to approximate our emergent behaviour model. Doing this will avoid the computational constraints of solving the full optimisation problem, however the fact that our model has significant amounts of driving concerns explicitly incorporated into it means it should be far easier to fit to data than a neural network. Doing this side-steps the enormous cost of collecting enough data to train a neural network for autonomous driving, as well as the computational complexity of our approach. An investigation into the most appropriate network structure to approximate such a model is needed if this avenue is pursued.

Functional safety is also paramount to consider for applications of any autonomous driving methodologies. In general, this describes an umbrella term that encapsulates a wide variety of assurances one would require for the full autonomous system: ensuring perception works in all weather an environment conditions, ensuring that for all possible outputs of perception the prediction models function, understanding all potential outputs that a system may yield and so forth. Whilst other non-rule based methods, for example reinforcement learning in the case of [205], have examined this problem, it is still very much a concern when applying highly complex and non-linear methods. If our method were to be taken closer to implementation, one would have to consider how to adhere to such standards and what the accepted standards will indeed be at the time autonomous vehicles are ready for significant commercialisation.

Chapter 8

Conclusions and future work

In this thesis, we have applied techniques from data science and mathematical modelling to tackle a number of problems in the domains of intelligent transportation systems and intelligent mobility. For the first three discussed problems, the ultimate goal was to exploit the wealth of data provided across the SRN to better monitor, understand and model existing infrastructure. For the final discussed problem, the focus turned to a pressing issue for industry in autonomous vehicles, and how one might model behaviour on a smaller, individual vehicle scale rather than across the entire network. Our results throughout this work have shown how we have successfully used the wealth of data provided on the SRN to develop new anomaly detection methods, apply novel mathematical models and derive new understanding from the data, and address key issues raised in the literature.

In chapter 4, we proposed a non-parametric methodology to define the typical behaviour of a link, and proposed a principled way of alerting operators to problems on the link in real-time by tracking fluctuations from this typical behaviour. Suggesting a natural definition of severity along with this, we showed that these periods of atypical behaviour aligned closely with incident labels in the data and spikes in travel time. Upon comparison, the method was shown to be comparable or superior to existing incident detection methods, particularly if bimodality was present in the speed data conditioned on the time of week. Clear future extensions exist for this section of work that would be relevant if it were to be applied in industry. An obvious one would be to try this approach directly on loop level data rather than on aggregated link-level data as we have done. This may provide a more granular view of what is happening on the network and may further increase the speed with which information about significant deviations can be extracted. However, to properly validate such a method, one would ideally use the true incident locations, which are currently removed before data is input into NTIS. Another avenue for future work would be to consider what performance criterion should be optimised

to combine false alarm rate, detection rate and mean time to detect. Whilst we have used the standard performance index, it requires arbitrary choices of perturbations to avoid trivial minima, and these choices of course influence the optimal parameter values. Ideas from the multi-objective optimisation literature may be appropriate for this, in conjunction with close discussion with traffic operators.

In chapter 5, we applied a self-exciting point process model to NTIS incident data. This allowed us to explore both a set of background components, representing spatial and temporal variation in the incident rate that resulted in primary incidents, and to model the fact that the occurrence of incidents may themselves alter the likelihood of another nearby in space and time, generating secondary incidents. Analysis of each component is given at length in chapter 5, with the main conclusions being two spatial hotspots and a double peak daily structure existed in the background, and 6-7% of incidents are most likely due to self-excitation under the assumptions of the model. Time and length scales of around 100 minutes and 1 kilometre were identified for self-excitation. In the future, it would be very informative to repeat the analysis for a major road without the MIDAS system or other smart motorway features since we expect this infrastructure to reduce the risk of secondary incidents. General comments about clustering of incidents on urban roads, made in [39], suggest that this behaviour is not exclusive to smart motorways, but also occurs on inner-city roads. Additionally, determining the amount of self-excitation observed on the other side of a carriageway to an incident would be informative, to quantify the 'rubbernecking' effect. More relevant variables could also be incorporated into the model, for example weather conditions. One could also extend our out-of-sample validation to include a trend component, however a parametric form would be required, as we cannot extrapolate the non-parametric estimate. Finally, our method of limiting the model freedom and improving interpretability by enforcing monotonic triggering functions, along with unidirectional triggering, are not specific to the discussed application, and can be used when models of this form are applied to different domains with similar practical considerations.

Chapter 6 considered dynamic prediction of incident durations, incorporating information from the sensor network to update predictions as new features became evident. We addressed a number of problems raised in the literature, aiming to create a system of practical use to operators, and compared multiple models across different performance criteria. In the static setting, there was little to differentiate the models, however a random survival forest model generally showed optimal performance. In the dynamic setting, non-parametric neural network models and random survival forests were competitive, and temporal convolutions improved upon manually engineered features from the

time series. One avenue for future work here is to incorporate more features in the dynamic models. These include when recovery vehicles arrive, police involvement and details from on-site reports, along with social media and weather data that can be captured, and can develop over time. These will likely have predictive power for an incidents duration, however attaining and incorporating the data types they present remains a challenge in dynamic prediction. Incorporation of more spatial information also remains an avenue for future work. This can be provided by the loop sensors, and may offer some additional insight into an incident's duration that is not captured by the link averaged values we have used. Finally, from our analysis in chapter 6 we suspect that if one was able to derive robust, complex features from the time series and feed them into a random survival forest model, we may see improved performance. One way to accomplish this is to consider an auto-encoder framework in which we pre-train a model to determine hidden representations of the series, however it is unclear if these will offer the same predictive power as we have observed from the CNN.

Chapter 7 considered the path planning problem for autonomous vehicles. We formulated an optimisation problem with measures of safety, comfort and progress, showing that dynamically solving this resulted in the emergence of a number of complex driving behaviours. Uncertainty in the prediction of other actors influenced these driving behaviours. Considering future work, one could adapt our model to better mimic social norms, including eco-friendly driving and avoiding undertaking. Another avenue would be to consider cases where no optimisation solution exists such that the probability of occupying the same space as another actor is above some minimum threshold. In such a case, it may be irresponsible to proceed and some 'emergency case' would likely need to be implemented. As discussed within the chapter, first calibrating the emergent behaviour model to data, then generating simulated data from the model to train a neural network could side-step both the cost of collecting enough data to train a deep neural network and the computational demand of the emergent behaviour model. An investigation into the most appropriate network structure to approximate such a model is needed if this avenue is pursued. Additionally, it would be interesting to see what emergent behaviours are observed in extreme scenarios based on real-world experience, for example situations that resulted in an accident. Open questions in this context include could the optimisation problem have avoided the accident within the current constraints? Would slack variables have to be introduced to allow the model to exceed certain constraints, and where so? Further, do the prediction models offer enough warning to avoid the incident entirely? Finally, refinement and improvement of the prediction models determining the future probability distribution of actors is fundamental to this method, so any work to validate and improve these

would greatly improve the usefulness of our methodology.

The methodologies we have developed offer practically useful ways for operators to better manage road infrastructure, and the conclusions drawn throughout this thesis are useful in understanding different aspects of the data, infrastructure and traffic patterns on it. Further, these methodologies and analyses are novel in the considered domain. There are of course still significant amounts of research that could be done to improve upon each aspect of our work, or tackle different problems entirely in intelligent mobility and intelligent transportation systems. Such research will be required to truly realise the full potential of existing infrastructure and optimally incorporate new technologies as they become available. Ultimately, this could continue to improve the reliability and safety of journeys for road users and on a wider scale, the efficiency of all transportation avenues through intelligent usage of data science and mathematical modelling.

Appendix A

Appendix to Chapter 2

A.1 Further discussion on kernel density estimation

We follow the discussion in [111, chapter 3] and [110] to detail where formulae discussed in chapter 2 arise.

A.1.1 Some properties of the estimator

Throughout the following we consider a Gaussian kernel with mean 0 and variance 1, written as

$$k(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{A.1}$$

and denote the kernel estimator with bandwidth ω as

$$\hat{p}_\omega(x) = \frac{1}{N\omega} \sum_{i=1}^{N} k\left(\frac{x - X_i}{\omega}\right). \tag{A.2}$$

We note that the kernel is symmetric, and assume the data-points we have are independent, identically distributed random variables.

The bias

The bias of the estimator $\hat{p}_\omega$ of the true density $p(x)$ at some point x is defined as

$$Bias(\hat{p}_\omega(x), p(x)) = \mathbb{E}\left[\hat{p}_\omega(x)\right] - p(x) \tag{A.3}$$

One can compute $\mathbb{E}\left[\hat{p}_\omega(x)\right]$ as follows

$$\begin{aligned}\mathbb{E}\left[\hat{p}_\omega(x)\right] &= \frac{1}{N\omega}\sum_{i=1}^{N}\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)\right] \\ &= \frac{1}{\omega}\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)\right] \\ &= \int_{-\infty}^{\infty}\frac{1}{\omega}k\left(\frac{x-u}{\omega}\right)p(u)du.\end{aligned} \tag{A.4}$$

We now use the change of variables $y = \frac{x-u}{\omega}$ and attain

$$\mathbb{E}\left[\hat{p}_\omega(x)\right] = \int_{-\infty}^{\infty}k\left(y\right)p(x-\omega y)dy. \tag{A.5}$$

Taylor expansion of $p(x-\omega y)$ yields

$$\begin{aligned}\mathbb{E}\left[\hat{p}_\omega(x)\right] &= \int_{-\infty}^{\infty}k\left(y\right)\left[p(x)-\omega y p'(x)+\frac{\omega^2y^2}{2}p''(x)+\dots\right]dy \\ &= p(x)\int_{-\infty}^{\infty}k\left(y\right)dy-\omega p'(x)\int_{-\infty}^{\infty}k\left(y\right)ydy+\frac{\omega^2p''(x)}{2}\int_{-\infty}^{\infty}k\left(y\right)y^2dy+\mathcal{O}\left(\omega^3\right) \\ &= p(x)+\frac{\omega^2p''(x)}{2}m_2(k)+\mathcal{O}\left(\omega^3\right)\end{aligned} \tag{A.6}$$

where $m_2(k) = \int_{-\infty}^{\infty}k\left(y\right)y^2dy$. The bias of the estimator is therefore

$$Bias(\hat{p}_\omega(x), p(x)) = \frac{\omega^2p''(x)}{2}m_2(k)+\mathcal{O}\left(\omega^3\right) \tag{A.7}$$

Finally, for small ω, we have

$$Bias(\hat{p}_\omega(x), p(x)) \approx \frac{\omega^2p''(x)}{2}m_2(k) \tag{A.8}$$

where high order terms in ω are discarded.

The variance

The variance of the estimator at a point x can be derived as follows

$$\begin{aligned}Var(\hat{p}_\omega(x)) &= Var\left(\frac{1}{N\omega}\sum_{i=1}^{N}k\left(\frac{x-X_i}{\omega}\right)\right) \\ &= \frac{1}{N^2\omega^2}Var\left(\sum_{i=1}^{N}k\left(\frac{x-X_i}{\omega}\right)\right) \\ &= \frac{1}{N\omega^2}Var\left(k\left(\frac{x-X_i}{\omega}\right)\right) \\ &= \frac{1}{N\omega^2}\left(\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)^2\right]-\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)\right]^2\right).\end{aligned} \tag{A.9}$$

We have seen from Eqs. (A.3), (A.5) and (A.6) that we can replace $\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)\right]$ with $\omega\left[p(x)+\frac{\omega^2 p''(x)}{2}m_2(k)+\mathcal{O}\left(\omega^3\right)\right]$. For the remaining term, we see

$$\frac{1}{N\omega^2}\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)^2\right]=\frac{1}{N\omega}\int_{-\infty}^{\infty}\frac{1}{\omega}k\left(\frac{x-u}{\omega}\right)^2 p(u)du \tag{A.10}$$

which we can again apply the change of variables $y=\frac{x-u}{\omega}$ to and attain

$$\frac{1}{N\omega^2}\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)^2\right]=\frac{1}{N\omega}\int_{-\infty}^{\infty}k\left(y\right)^2 p(x-\omega y)dy. \tag{A.11}$$

Taylor expanding yields

$$\begin{aligned}\frac{1}{N\omega^2}\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)^2\right]&=\frac{1}{N\omega}\int_{-\infty}^{\infty}k\left(y\right)^2\left[p(x)-\omega y p'(x)+\frac{\omega^2 y^2}{2}p''(x)+\dots\right]dy\\&=\frac{1}{N\omega}\left[p(x)\int_{-\infty}^{\infty}k\left(y\right)^2 dy-\omega p'(x)\int_{-\infty}^{\infty}k\left(y\right)^2 y dy+\frac{\omega^2 p''(x)}{2}\int_{-\infty}^{\infty}k\left(y\right)^2 y^2 dy+\mathcal{O}\left(\omega^3\right)\right].\end{aligned} \tag{A.12}$$

Hence, we see

$$\begin{aligned}Var(\hat{p}_\omega(x))&=\frac{1}{N\omega^2}\left(\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)^2\right]-\mathbb{E}\left[k\left(\frac{x-X_i}{\omega}\right)\right]^2\right)\\&=\frac{p(x)}{N\omega}\int_{-\infty}^{\infty}k\left(y\right)^2 dy-\frac{p'(x)}{N}\int_{-\infty}^{\infty}k\left(y\right)^2 y dy+\frac{\omega p''(x)}{2N}\int_{-\infty}^{\infty}k\left(y\right)^2 y^2 dy-\frac{1}{N}\left[p(x)+\frac{\omega^2 p''(x)}{2}m_2(k)+\mathcal{O}\left(\omega^3\right)\right]^2+\mathcal{O}\left(\frac{\omega^2}{N}\right).\end{aligned} \tag{A.13}$$

If we now consider the amount of data to be large, $(N\to\infty)$ and a small ω, we see most of the terms here can be omitted, leaving

$$Var(\hat{p}_\omega(x))\approx\frac{p(x)}{N\omega}\int_{-\infty}^{\infty}k\left(y\right)^2 dy. \tag{A.14}$$

A.1.2 The bias-variance trade-off

One can view the MSE for KDE as a trade-off between the bias and variance of the estimator. First consider expected MSE, given by

$$MSE\left(\hat{p}_\omega\right)=\mathbb{E}\left[\left(\hat{p}_\omega(x)-p(x)\right)^2\right] \tag{A.15}$$

where as before $\hat{p}_\omega$ is the smoothed estimate and p is the true density function we wish to approximate. We can re-write Eq. (A.15) in terms of bias and

variance as follows

$$\begin{aligned}
MSE\left(\hat{p}_\omega\right) &= \mathbb{E}\left[\left(\hat{p}_\omega(x) - \mathbb{E}[\hat{p}_\omega(x)] + \mathbb{E}[\hat{p}_\omega(x)] - p(x)\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{p}_\omega(x) - \mathbb{E}[\hat{p}_\omega(x)]\right)^2 + 2(\hat{p}_\omega(x) - \mathbb{E}[\hat{p}_\omega(x)])(\mathbb{E}[\hat{p}_\omega(x)] - p(x)) + \left(\mathbb{E}[\hat{p}_\omega(x)] - p(x)\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{p}_\omega(x) - \mathbb{E}[\hat{p}_\omega(x)]\right)^2\right] + 2\mathbb{E}\left[(\hat{p}_\omega(x) - \mathbb{E}[\hat{p}_\omega(x)])(\mathbb{E}[\hat{p}_\omega(x)] - p(x))\right] + \mathbb{E}\left[\left(\mathbb{E}[\hat{p}_\omega(x)] - p(x)\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{p}_\omega(x) - \mathbb{E}[\hat{p}_\omega(x)]\right)^2\right] + 2\mathbb{E}\left[\hat{p}_\omega(x) - \mathbb{E}[\hat{p}_\omega(x)]\right]\left(\mathbb{E}[\hat{p}_\omega(x)] - p(x)\right) + \left(\mathbb{E}[\hat{p}_\omega(x)] - p(x)\right)^2 \\
&= \mathbb{E}\left[\left(\hat{p}_\omega(x) - \mathbb{E}[\hat{p}_\omega(x)]\right)^2\right] + 2\left(\mathbb{E}\left[\hat{p}_\omega(x)\right] - \mathbb{E}[\hat{p}_\omega(x)]\right)\left(\mathbb{E}[\hat{p}_\omega(x)] - p(x)\right) + \left(\mathbb{E}[\hat{p}_\omega(x)] - p(x)\right)^2 \\
&= \mathbb{E}\left[\left(\hat{p}_\omega(x) - \mathbb{E}[\hat{p}_\omega(x)]\right)^2\right] + \left(\mathbb{E}[\hat{p}_\omega(x)] - p(x)\right)^2 \\
&= Var(\hat{p}_\omega(x)) + Bias(\hat{p}_\omega(x), p(x))^2.
\end{aligned} \tag{A.16}$$

Clearly, choosing the smoothing parameter is a trade-off between the bias of the estimator and its variance. However, MSE is point-wise error, and one is interested in the global accuracy of $\hat{p}_\omega$, resulting in the consideration of MISE, written as

$$\begin{aligned}
MISE\left(\hat{p}_\omega\right) &= \mathbb{E}\left[\int_{-\infty}^{\infty}\left(\hat{p}_\omega(x) - p(x)\right)^2 dx\right] \\
&= \int_{-\infty}^{\infty} Bias(\hat{p}_\omega(x), p(x))^2 dx + \int_{-\infty}^{\infty} Var(\hat{p}_\omega(x)) dx.
\end{aligned} \tag{A.17}$$

A.1.3 Determining AMISE

If we substitute the asymptotic estimates of bias and variance given by Eqs. (A.8) and (A.14) into Eq. (A.17), we attain the asymptotic mean integrated square error (AMISE), which has the form

$$AMISE\left(\hat{p}_\omega\right) = \frac{\omega^4 m_2^2(k)}{4}\left(\int_{-\infty}^{\infty} p''(x)^2 dx\right) + \frac{\int_{-\infty}^{\infty} k\left(y\right)^2 dy}{N\omega}\left[\int_{-\infty}^{\infty} p(x) dx\right]. \tag{A.18}$$

If we denote $R(\nu) = \int_{-\infty}^{\infty} \nu^2(z) dz$ then we see that this can be written as

$$AMISE\left(\hat{p}_\omega\right) = \frac{R(k)}{N\omega} + \frac{\omega^4}{4}\left(m_2(k)\right)^2 R(p'') \tag{A.19}$$

which matches Eq. (2.4) exactly.

A.1.4 The minimiser of AMISE

Eq. (A.19) clearly has a minimiser, and it can be derived using standard methods. If we follow the methods in [206], we consider a general function of the form

$$f(x) = Ax^\alpha + Bx^{-\beta} \tag{A.20}$$

with $\alpha, \beta \neq 0$, we see that

$$f'(x) = A\alpha x^{\alpha-1} - B\beta x^{-\beta-1}, \tag{A.21}$$

and the minimum of the function is found at

$$x = \left(\frac{\beta B}{\alpha A}\right)^{\frac{1}{\alpha+\beta}}. \tag{A.22}$$

Applying this to Eq. (A.19), we see that $A = \frac{(m_2(k))^2 R(p'')}{4}$, $B = \frac{R(k)}{N}$, $\alpha = 4$ and $\beta = 1$. It is therefore clear that the minimiser of Eq. (A.19), called ω_{optimal} is

$$\omega_{\text{optimal}} = \left(\frac{R(k)}{(m_2(k))^2 R(p'')}\right)^{\frac{1}{5}} N^{-\frac{1}{5}}. \tag{A.23}$$

Taking a step back, we realise that our data collection tells us of the value of N, and our choice of kernel allows us to compute $R(k)$ and $(m_2(k))^2$. However, we still do not know p, and therefore cannot directly compute $R(p'')$.

The idea behind plug-in methods is to choose some reference bandwidth ω_0 to estimate $R(p'')$, and plug this estimate into Eq. (A.23). As discussed in [111], the final density estimate $\hat{p}_\omega(x)$ is less sensitive to the choice of ω_0 than ω, justifying the idea. One can even go further, approximating integrals of higher order derivatives and see the influence of the initial choice decay further. Fundamentally, these methods require a choice of bandwidth at some stage, however they exploit the calculations throughout this section to make a choice that the final approximation is less sensitive to compared to directly choosing ω. These methods can be extended to the multivariate setting, where a matrix of smoothing values is determined. This is discussed at length in [110].

A.2 Further discussion on model interpretability & SHAP values

To offer more details on SHAP values, we follow discussions in [207] throughout this section.

A.2.1 Coalition games

To begin, we consider a coalition game, with a pair (N, v) where:

- N is a finite set of players
- v is a function that yields a pay-off for each coalition $S \subseteq N$
- Assume an empty set gives no pay-off

Shapley values arise from a theory from coalition games, stating there must be a unique way to divide the pay-off between a coalition that satisfies three axioms: 'symmetry', 'dummy player' and 'additivity'.

A.2.2 Symmetry

Two players i and j are interchangeable if they always contribute the same amount to every coalition of other players. Formally, for all S that contain neither i or j, they are interchangeable if

$$v(S \cup \{i\}) = v(S \cup \{j\}). \tag{A.24}$$

The symmetry axiom states that if i and j are interchangeable, then $\psi_i(N, v) = \psi_j(N, v)$, that is their pay off should be the same.

A.2.3 Dumb players

A dumb player gains nothing personally from joining a coalition, resulting in

$$\forall\ S \text{ s.t. } i \notin S, v(S \cup \{i\}) - v(S) = v(\{i\}). \tag{A.25}$$

Their pay-off must be $\psi_i(N, v) = v(\{i\})$.

A.2.4 Additivity

For any two v_1, v_2, we have for any player i that

$$\psi_i(N, v_1 + v_2) = \psi_i(N, v_1) + \psi_i(N, v_2) \tag{A.26}$$

where the game $(N, v_1 + v_2)$ is defined by $(v_1 + v_2)(S) = v_1(S) + v_2(S)$ for every coalition S.

A.2.5 Shapley values

Shapley values tell us how to divide the pay-off in a coalition game in a unique way such that the 3 discussed axioms are met. The value for player i is computed as

$$\phi_i(N, v) = \frac{1}{|N|!} \sum_{S \subseteq N \setminus i} |S|!(|N| - |S| - 1)! \left[v(S \cup \{i\}) - v(S)\right]. \tag{A.27}$$

We can intuitively consider how Eq. (A.27) arises as follows. Imagine at first we have an empty coalition of players. Then, at random, we start adding players to it. Whenever we happen to add player i, the change in value will be

$$v(S \cup \{i\}) - v(S). \tag{A.28}$$

Now consider, at the point we add i into S, how many ways could S have formed into its current state? There are $|S|!$ ways this could have happened.

Further, there are $(|N| - |S| - 1)!$ ways that the remaining players could be added after i has been. This suggests for a single instance we have

$$|S|!(|N| - |S| - 1)! \left[v(S \cup \{i\}) - v(S)\right]. \tag{A.29}$$

However, this is just for a single S, so we then sum over all possible sets S and then average the value by dividing by the the total number of possible orderings of players. Combining all of this, we attain Eq. (A.27).

A.2.6 Interpretability in machine learning

Shapley values should be interpreted in the machine learning context similar to how they are game theory, where the model features represent players and the game result is the model output. Since computing these values involves summing over all feature subsets, the raw implementation is quite computationally expensive. However, a number of computational improvements are discussed in [132]. Two further complications exist. Firstly, neural networks cannot take arbitrary missing values. To avoid this problem, one instead considers some 'reference' or 'background' dataset, from which values are taken as replacements when considering alterations of feature vectors [131]. The appropriate choice of background is an open problem in applying these methods. For image tasks, it might be clear than one can use a blank image as a reference, however in many domains an intuitive reference value does not exist. Instead, it is common to provide a dataset with many records in as the background, and average over it, which is the approach we take.

Secondly, there is often structure to feature vectors in problems, and elements are not simply a random collection of possible values for each entry. One can consider this structure most clearly when considering a one-hot encoded example. Suppose our feature vector contained three binary values, which indicate morning, afternoon or evening, which of course correspond to a time of day being discretised and then one-hot encoded. In the raw methodology for SHAP in [131], one would go through each feature, and consider setting it to a reference value and use the change in model output as a measure of importance. However, suppose we had our time of day encoding, and an entry occurred during the morning, with a feature vector of $[1, 0, 0]$. We may then consider altering the entry corresponding to the binary label for afternoon, which might lead us to considering a feature vector of $[1, 1, 0]$. Such an input is not practically possible, and so the model output for it is irrelevant to our problem. To avoid this, recent work in [132] incorporates the structure in the data before performing any perturbations. A partition of the data into highly correlated components is performed, and then when considering feature importance, instead of altering a single element, a set are altered if they are

grouped together. We do however note that we compute the SHAP values both incorporating this structure and neglecting it, and see similar results in both cases.

A.3 Further discussion on optimisation

A.3.1 Gradient decent and Newton's method from Taylor series

The goal of both gradient descent and Newton's method is, given a current point $\boldsymbol{x}$, choose an optimal step $\Delta\boldsymbol{x}$ to move towards the minima of a function f. If we consider a Taylor expansion of f, with some $\boldsymbol{v}$ up to second order terms, we have

$$f(\boldsymbol{x}+\boldsymbol{v}) \approx f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T\boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{v}. \tag{A.30}$$

To choose the optimal $\boldsymbol{v}$ that minimises this approximation, we differentiate and set to 0, giving

$$0 = \nabla f(\boldsymbol{x}) + \nabla^2 f(\boldsymbol{x})\boldsymbol{v}, \tag{A.31}$$

and hence the optimal step $\Delta\boldsymbol{x}$ is

$$\Delta\boldsymbol{x} = -\left(\nabla^2 f(\boldsymbol{x})\right)^{-1}\nabla f(\boldsymbol{x}). \tag{A.32}$$

This suggests an iterative procedure of

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \left(\nabla^2 f(\boldsymbol{x})\right)^{-1}\nabla f(\boldsymbol{x}) \tag{A.33}$$

which is exactly Newton's method. Gradient descent on the other hand does not use any information on $\nabla^2 f(\boldsymbol{x})$. For direct comparison, the methods are

$$\begin{aligned} \boldsymbol{x}^{(n+1)} &= \boldsymbol{x}^{(n)} - \eta\nabla f(\boldsymbol{x}) && \text{(Gradient decent)} \\ \boldsymbol{x}^{(n+1)} &= \boldsymbol{x}^{(n)} - \eta\left(\nabla^2 f(\boldsymbol{x})\right)^{-1}\nabla f(\boldsymbol{x}) && \text{(Newton's method).} \end{aligned} \tag{A.34}$$

Here, we have included some step size η that influences how far we move at each iteration, and can be chosen by backtracking line search.

A.3.2 Newton's method with equality constraints

Equality constraints can be incorporated into Newton's method naturally by doing as follows. If we have the general problem

$$\min\left(f(\boldsymbol{x})\right) \text{ s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{A.35}$$

we recall that Newton's method is attempting to choose the optimal value of $\boldsymbol{v}$ to minimise and satisfy

$$\begin{aligned} \hat{f}(\boldsymbol{x}+\boldsymbol{v}) &= f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^T \nabla^2 f(\boldsymbol{x})\boldsymbol{v}, \\ \boldsymbol{A}(\boldsymbol{x}+\boldsymbol{v}) &= \boldsymbol{b}. \end{aligned} \tag{A.36}$$

If we assume that we start at a feasible point, then the later equation becomes $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}$. We can write the Lagrangian and its derivative for this problem as

$$\begin{aligned} \mathcal{L} &= f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^T \nabla^2 f(\boldsymbol{x})\boldsymbol{v} + \boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{v} \\ \nabla\mathcal{L} &= \nabla f(\boldsymbol{x}) + \nabla^2 f(\boldsymbol{x})\boldsymbol{v} + \boldsymbol{A}^T\boldsymbol{\lambda}. \end{aligned} \tag{A.37}$$

The Karush-Kuhn-Tucker conditions yield

$$\begin{aligned} \nabla f(\boldsymbol{x}) + \nabla^2 f(\boldsymbol{x})\boldsymbol{v} + \boldsymbol{A}^T\boldsymbol{\lambda} = \boldsymbol{0} \ \text{(stationary)} \\ \boldsymbol{A}\boldsymbol{v} = \boldsymbol{0} \ \text{(feasibility)}. \end{aligned} \tag{A.38}$$

This yields the final system, giving the optimal step $\Delta\boldsymbol{x}$ as

$$\begin{pmatrix} \nabla^2 f(\boldsymbol{x}) & \boldsymbol{A}^T \\ \boldsymbol{A} & 0 \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} -\nabla f(\boldsymbol{x}) \\ 0 \end{pmatrix}. \tag{A.39}$$

If the equality constraint is non-linear, written as $h^c(\boldsymbol{x}) = 0$, then this system becomes

$$\begin{pmatrix} \nabla^2 f(\boldsymbol{x}) & \nabla h^c(\boldsymbol{x})^T \\ \nabla h^c(\boldsymbol{x}) & 0 \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} -\nabla f(\boldsymbol{x}) \\ -h^c(\boldsymbol{x}) \end{pmatrix} \tag{A.40}$$

where the second equation is derived from considering the Newton step for solving $h^c(\boldsymbol{x}) = 0$.

A.3.3 The barrier method

The barrier method is an optimisation technique that incorporates inequality constraints into an optimisation, not just equality ones. Suppose we have an optimisation problem

$$\min\left(f(\boldsymbol{x})\right) \text{ s.t. } h_i^c(\boldsymbol{x}) \leq 0 \tag{A.41}$$

$\forall\, i \in \{1, \ldots M_c\}$, meaning we have a function to minimise and have M_c inequality constraints captured by h_i^c. As discussed in section 2.5.4, the barrier method replaces the inequality constraints with a penalty term directly

in the optimisation function, giving

$$\min\left(f(\boldsymbol{x}) - \frac{1}{\kappa}\sum_{i=1}^{M_c}\log(-h_i^c(\boldsymbol{x}))\right) \tag{A.42}$$

if the log-barrier function is used, and recall that $\kappa > 0$ is a parameter that can vary to ensure the constraints are enforced. We can now solve Eq. (A.42) subject to any equality constraints using Newton's method, and in practice the barrier method performs the steps listed in algorithm 3.

Algorithm 3: Barrier method

Input: Starting point $\boldsymbol{x}^{(0)}$, Stopping threshold $\epsilon > 0$, Inequality constraints $h_i^c(\boldsymbol{x})$, Update factor $\mu > 1$

Let M_c be the number of constraints, initialise κ to some positive value.
while $\frac{M_c}{\kappa} > \epsilon$ **do**
    Solve the optimisation problem defined in Eq. (A.42) using κ and starting point $\boldsymbol{x}^{(0)}$
    Update starting point $\boldsymbol{x}^{(0)}$ as solution to current problem
    Update κ with $\kappa \to \mu\kappa$
end

Output: Final optimization result $\boldsymbol{x}$

A.3.4 Backtracking line search

We have already seen that Newton's method applies a step of $\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \eta\left(\nabla^2 f(\boldsymbol{x})\right)^{-1}\nabla f(\boldsymbol{x})$, however there are methods to choose η such that one takes optimal step lengths, rather than fixed ones. Such methods are called backtracking line search, and pseudocode describing them is given in algorithm 4. One can apply algorithm 4 at each step of the optimisation problem to

Algorithm 4: Backtracking line search

Input: Parameters $\alpha \in (0, \frac{1}{2}]$ and $\beta \in (0, 1)$, function to minimise $f(\boldsymbol{x})$

Set $\eta = 1$
Let $\boldsymbol{v} = -\left(\nabla^2 f(\boldsymbol{x})\right)^{-1}\nabla f(\boldsymbol{x})$
while $f(\boldsymbol{x} + \eta\boldsymbol{v}) > f(\boldsymbol{x}) + \alpha\eta\nabla f(\boldsymbol{x})^T\boldsymbol{v}$ **do**
    Update η with $\eta \to \beta\eta$
end

Output: Optimal step size η

take optimal sized steps, yielding faster and more stable convergence compared to using a fixed η.

Appendix B

Appendix to Chapter 3

B.1 Data pipeline

An outline of the data pipeline developed for this work is given in Fig. B.1.

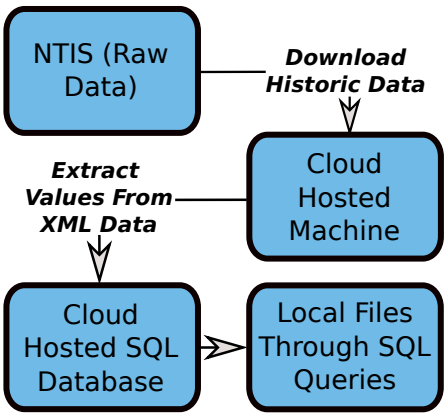


Figure B.1: Outline of the flow of data from the NTIS system to a local computer.

1 - Acquire raw NTIS data in XML format from Highways England.

2 - Extract required information by parsing XML files.

3 - Insert time series, incident, sign and speed limit data into an SQL database.

4 - Query the database to get local files.

B.2 Conversion of TMS to SMS

Denoting SMS $= \bar{v}_{\mathrm{SM}}$ and TMS $= \bar{v}_{\mathrm{TM}}$, it was shown by Wardrop in [142] that, assuming constant road topology, one can write

$$\bar{v}_{\mathrm{TM}} = \bar{v}_{\mathrm{SM}} + \frac{\sigma^2_{\mathrm{SM}}}{\bar{v}_{\mathrm{SM}}} \tag{B.1}$$

where σ_{SM} is the standard deviation of the instantaneous vehicle speeds. For our purposes, Eq. (B.1) is not directly applicable as we do not have the resolution of data to measure the instantaneous speed distribution. Instead, we turn to methods developed in [141], where the authors derive a heuristic approximation

to calculate $\bar{v}_{\mathrm{SM}}$ given $\bar{v}_{\mathrm{TM}}$. In short, the authors did as follows: first, suppose we use the definition of variance and linearity of expectation to write

$$\begin{aligned}\sigma_{\mathrm{SM}}^2 &= \mathbb{E}\left[(v_i - \bar{v}_{\mathrm{SM}})^2\right] \\ &= \mathbb{E}\left[v_i^2 + \bar{v}_{\mathrm{SM}}^2 - 2v_i\bar{v}_{\mathrm{SM}}\right] \\ &= \mathbb{E}\left[v_i^2\right] + \bar{v}_{\mathrm{SM}}^2 - 2\bar{v}_{\mathrm{SM}}\mathbb{E}\left[v_i\right].\end{aligned} \tag{B.2}$$

From Eq. (B.2), we do not have data measuring $\mathbb{E}\left[v_i\right]$ or $\mathbb{E}\left[v_i^2\right]$. If we assume the traffic stream maintains the same levels of flow and speed along a specified stretch of road in a single measurement interval, we can approximate $\mathbb{E}\left[v_i\right] \approx \bar{v}_{\mathrm{TM}}$, meaning we expect the instantaneous speed of any vehicle i to be equal to the average speed measured at the detector location under these assumptions. Substituting this into Eq. (B.2), we get

$$\sigma_{\mathrm{SM}}^2 \approx \mathbb{E}\left[v_i^2\right] + \bar{v}_{\mathrm{SM}}^2 - 2\bar{v}_{\mathrm{SM}}\bar{v}_{\mathrm{TM}}. \tag{B.3}$$

Replacing σ_{SM}^2 using Eq. (B.1), we attain

$$\bar{v}_{\mathrm{SM}}\bar{v}_{\mathrm{TM}} - \bar{v}_{\mathrm{SM}}^2 \approx \mathbb{E}\left[v_i^2\right] + \bar{v}_{\mathrm{SM}}^2 - 2\bar{v}_{\mathrm{SM}}\bar{v}_{\mathrm{TM}} \tag{B.4}$$

from which we can attain the final quadratic in $\bar{v}_{\mathrm{SM}}$ as

$$2\bar{v}_{\mathrm{SM}}^2 - 3\bar{v}_{\mathrm{TM}}\bar{v}_{\mathrm{SM}} + \mathbb{E}\left[v_i^2\right] \approx 0. \tag{B.5}$$

Whilst there are actually two unknown quantities in Eq. (B.5), $\bar{v}_{\mathrm{SM}}$ and $\mathbb{E}\left[v_i^2\right]$, [141] derived an empirical relationship between the observed $\bar{v}_{\mathrm{TM}}$ and unknown $\mathbb{E}\left[v_i^2\right]$, fitting such a relationship across multiple links to ensure generality.

To fit this relationship, the authors took Automatic Number Plate Recognition (ANPR) and sensor data from NTIS. Using these two data sources, and inspecting 2 weeks of data across 6 different links, the authors proposed a quadratic relationship between $\bar{v}_{\mathrm{TM}}$ and $\mathbb{E}\left[v_i^2\right]$, of the form

$$\mathbb{E}\left[v_i^2\right] = a\bar{v}_{\mathrm{TM}}^2 + b\bar{v}_{\mathrm{TM}} + c. \tag{B.6}$$

for some constants a, b, c.

Fitting the relationship in Eq. (B.6) to their datasets, the authors found such a fit had an R^2 value of 0.9718 with each coefficient being significantly different to 0 when tested. The final coefficients derived for the relationships were: $a = 1.22, b = -15.21, c = 207.95$.

Given this approximation for $\mathbb{E}\left[v_i^2\right]$, we then simply solve Eq. (B.5) using the quadratic formula, knowing to take the positive square root through realistic restrictions on the values of space mean speeds, yielding a final approximation

of

$$\bar{v}_{\mathrm{SM}} \approx \frac{3\bar{v}_{\mathrm{TM}} + \sqrt{9\bar{v}_{\mathrm{TM}}^2 - 8E\left[v_i^2\right]}}{4}. \tag{B.7}$$

The proposed method generated MAPE between 3.6% and 17.9% depending on link, roughly a 10% improvement compared to simply assuming $\bar{v}_{\mathrm{TM}}$ and $\bar{v}_{\mathrm{SM}}$ were equal.

B.3 Seasonality Plots

Seasonality in the data for each link is shown for flow, speed and density in Figs. B.2, B.3 and B.4 respectively. These seasonal visualisations are constructed by grouping the data-points by time of week, and computing the median value for each quantity.

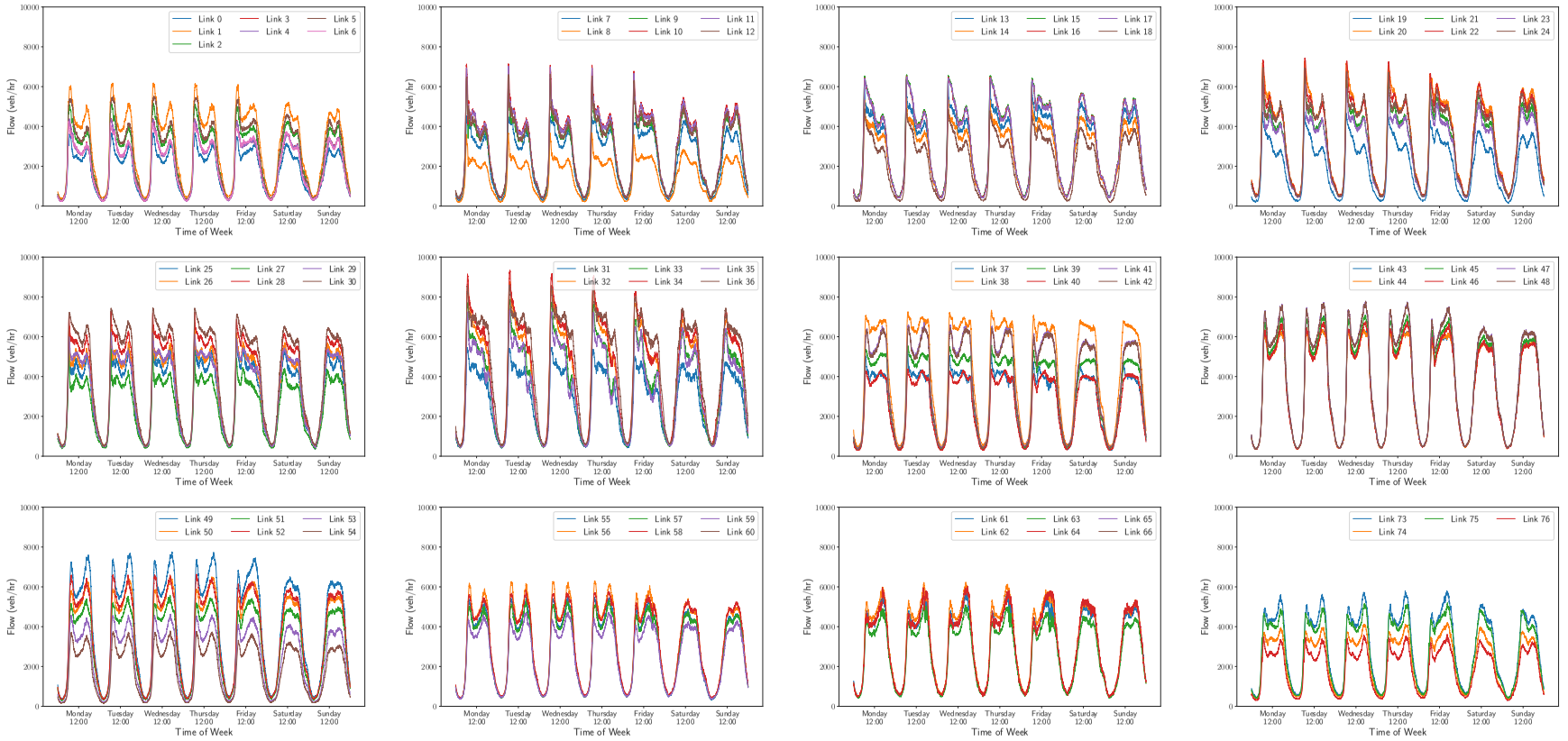


Figure B.2: Seasonality visualisations for flow on each link in the dataset, plotted in groups of 6 for brevity. Links are numbered starting at 0. Spatially, link 0 is located just below the Dartford crossing, and links follow spatially in a clockwise direction as link number increases.

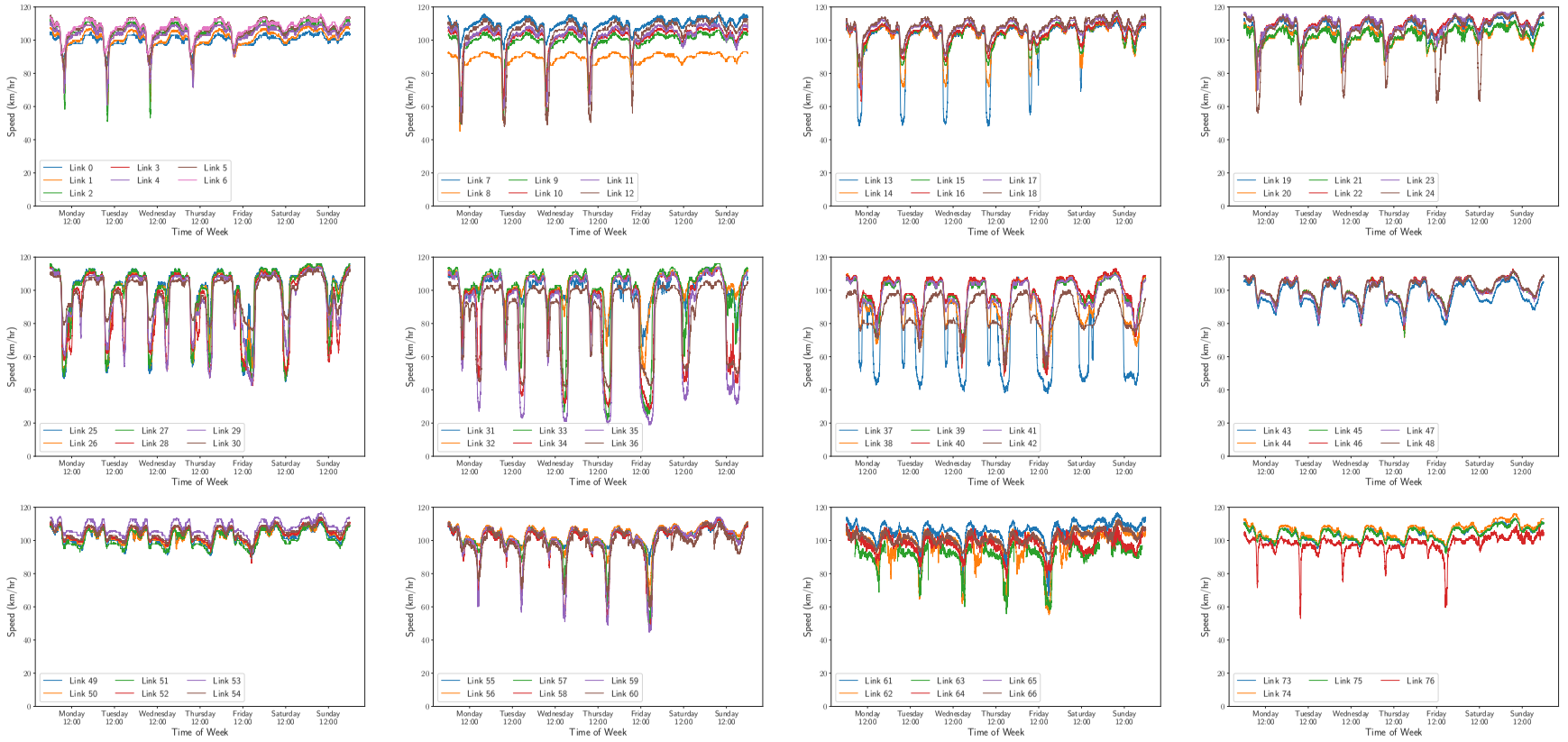


Figure B.3: Seasonality visualisations for speed on each link in the dataset, plotted in groups of 6 for brevity. Links are numbered starting at 0. Spatially, link 0 is located just below the Dartford crossing, and links follow spatially in a clockwise direction as link number increases.

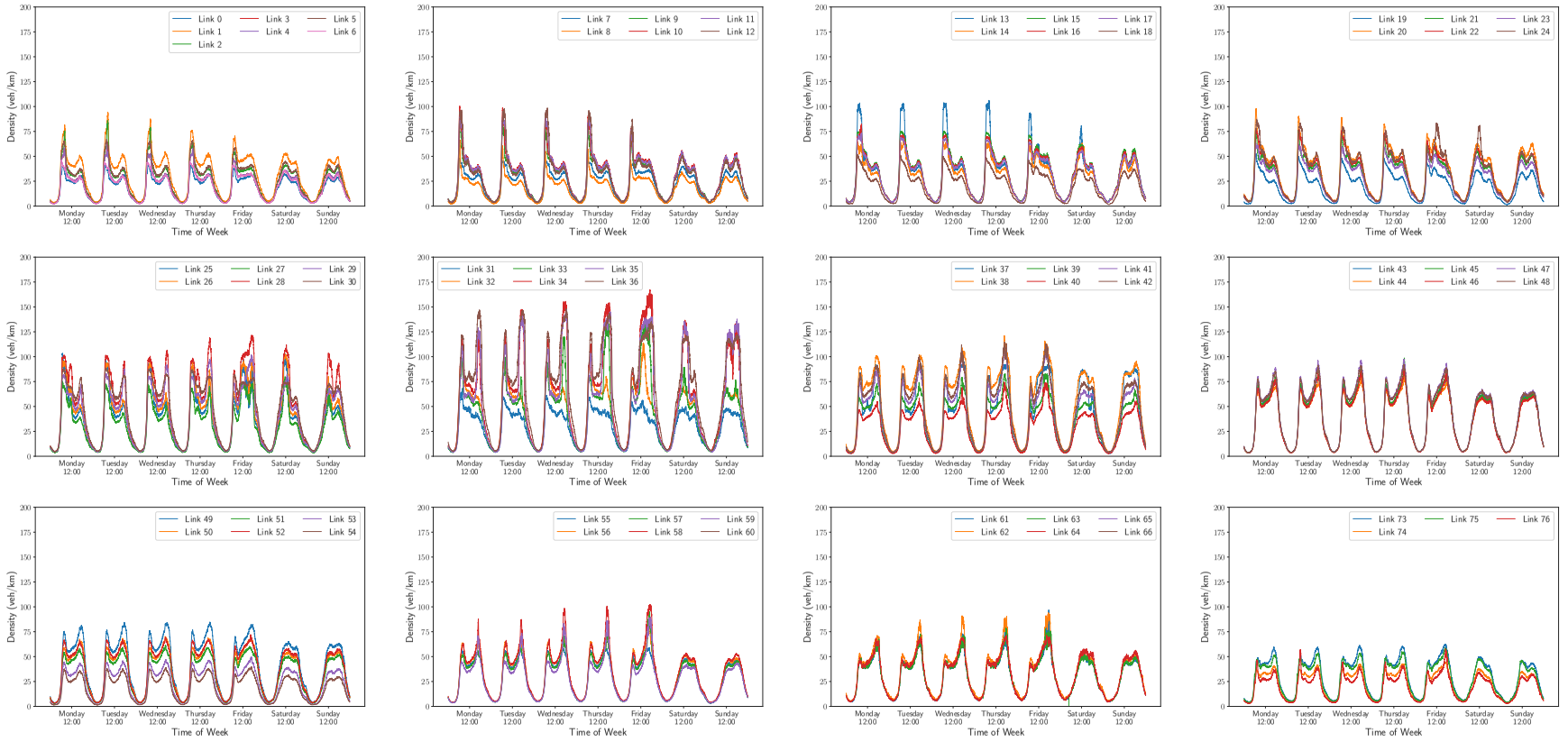


Figure B.4: Seasonality visualisations for density on each link in the dataset, plotted in groups of 6 for brevity. Links are numbered starting at 0. Spatially, link 0 is located just below the Dartford crossing, and links follow spatially in a clockwise direction as link number increases.

B.4 Summary statistics of data subsets

B.4.1 Data between April 7th 2017 and June 16th 2017

In table B.1, we show summary statistics for the 10 week subset of data between April 7th 2017 and June 16th 2017. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. Firstly, links 11, 12, 13 and 14 report only missing values in this subset, but had non-missing values for parts of the entire dataset. The same is true for links 61, 62, 63 and 64. Comparing central tendencies of the data, we see that generally links have similar median speeds, with the largest differences being on links 8, 42 and 62. Link 62 has the largest change, with a median speed of 101 km/hr in the full dataset but 85 km/hr in table B.1, however a very large amount of the data for link 62 is missing in this subset, so we do not expect the small amount we have to be as representative of the wider behaviour as we observe in the full dataset. The same reasoning applies to why we observe the largest difference in median travel time on this link. Median flow and density values appear consistent across both tables B.1 and 3.1.

Comparing the spread of data, we see the largest changes in inter-quartile range of speed values are observed on links 35 and 62. The changes on link 62 can again be attributed to missing data, however for link 35 we see that the subset of data given in table B.1 has a reduction in inter-quartile range of 11 km/hr. Inter-quartile ranges of flows are consistent between the full dataset and subset, and for densities generally vary by less than 4 vehicles per kilometre.

Since the data in table 3.1 spans over a full year of collection, we will observe data generated in each season. The subset in table B.1 is collected between April and June however, so we expect the average temperature to be higher during this subset compared to the entire dataset, and there to be less rain, fog and similar conditions that are common in winter in the UK. A final point of note is that this dataset is used for exploratory analysis in chapter 4, and there are clearly a significant amount of links appropriate for analysis.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	11	100	5	13	112	268	0	2047	1952	60	4056	553	1	26	1	24	202	268	0	20	20	1	123	553	1
1	3.6	13	26	103	7	21	114	273	0	3526	3348	194	6863	336	0	127	9	115	623	273	0	34	35	2	158	336	0
2	1.0	3	7	108	7	29	122	178	0	2819	2643	60	5766	367	0	33	2	29	333	178	0	26	26	1	113	367	0
3	0.5	2	8	106	7	18	124	423	0	2329	2186	60	4980	555	1	17	1	15	101	423	0	22	22	1	117	651	1
4	0.4	1	6	106	7	18	124	329	0	2340	2201	60	4949	355	0	14	1	12	79	329	0	22	22	1	114	376	0
5	4.5	2	17	109	8	23	128	175	0	2978	2890	180	6360	394	0	149	11	127	1800	175	0	27	28	2	135	399	0
6	0.9	2	6	110	7	11	129	641	1	2362	2215	64	4786	490	0	28	1	24	289	641	1	21	21	1	118	705	1
7	5.1	2	20	111	7	11	126	16	0	2678	2560	168	5374	441	0	166	11	146	1683	16	0	24	24	2	166	441	0
8	1.0	1	22	96	7	5	130	28757	29	1740	1673	60	3476	269	0	40	3	29	753	28757	29	21	10	1	381	28857	29
9	0.7	1	22	101	6	5	128	409	0	3098	2853	60	7097	540	1	23	2	18	469	409	0	30	29	1	279	540	1
10	0.9	2	35	102	7	4	130	12166	12	3588	3336	180	7980	266	0	31	2	25	798	12166	12	37	29	2	390	12262	12
11	5.9	13	123	102	10	3	130	30568	30	-	-	-	-	100800	100	209	20	163	7081	30568	30	-	-	-	-	100800	100
12	1.0	2	33	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100
13	6.8	17	99	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100
14	1.0	2	9	106	8	27	122	23270	23	-	-	-	-	100800	100	34	3	29	131	23270	23	-	-	-	-	100800	100
15	1.2	4	16	107	7	18	121	283	0	3810	3627	144	7516	441	0	40	2	35	232	283	0	35	35	1	162	441	0
16	1.5	6	16	108	7	24	121	165	0	3735	3566	144	7375	399	0	51	3	45	227	165	0	34	34	2	149	399	0
17	0.5	0	7	110	7	20	126	173	0	3744	3582	60	7317	264	0	15	1	14	135	173	0	33	33	1	169	277	0
18	0.2	1	5	110	7	14	127	183	0	2542	2534	60	5687	410	0	6	0	5	61	183	0	23	24	1	117	410	0
19	1.2	4	8	110	7	13	127	183	0	2394	2417	60	5388	399	0	41	3	35	411	183	0	21	23	1	116	399	0
20	2.0	4	16	106	8	14	119	403	0	4251	3929	195	8251	359	0	69	5	61	537	403	0	40	40	2	200	488	0
21	3.0	6	13	105	9	7	123	402	0	3526	3309	180	7143	884	1	102	8	87	1465	402	0	33	34	2	165	884	1
22	10.0	22	54	112	7	23	126	273	0	4022	3740	275	8335	330	0	322	22	286	1580	273	0	36	36	2	123	330	0
23	0.5	1	23	113	8	18	130	321	0	3385	3179	60	7382	391	0	16	1	14	103	321	0	31	31	1	148	391	0
24	6.4	13	91	111	10	16	125	276	0	3891	3692	240	7854	338	0	208	19	185	1446	276	0	36	38	2	154	338	0
25	0.8	2	47	108	14	5	125	284	0	3672	3297	300	7411	69739	69	26	3	22	517	284	0	36	43	3	152	69739	69
26	2.6	9	48	107	15	4	122	283	0	3969	3744	69	7980	350	0	86	13	75	2166	283	0	37	43	1	161	350	0
27	0.9	2	37	108	15	5	125	710	1	2909	2847	60	6120	865	1	29	4	25	723	710	1	27	31	1	144	985	1
28	6.9	13	89	104	22	5	121	1264	1	4467	4379	60	8247	769	1	240	56	206	4929	1264	1	43	54	2	162	1321	1
29	0.9	3	22	105	23	13	130	925	1	3793	3851	60	7260	976	1	32	8	26	258	925	1	38	50	0	151	1121	1
30	2.5	4	15	99	17	18	122	280	0	4757	4696	170	8335	402	0	91	16	74	502	280	0	48	58	2	152	402	0
31	1.2	1	11	103	11	18	120	297	0	3179	3180	146	6240	373	0	42	5	36	235	297	0	33	35	1	117	373	0
32	3.9	6	72	106	10	18	120	277	0	4499	4679	300	8854	339	0	131	13	116	770	277	0	45	52	3	214	339	0
33	1.0	3	104	108	12	5	122	291	0	3592	4200	180	8302	354	0	33	4	29	683	291	0	41	50	2	275	354	0
34	1.9	3	98	105	14	15	119	291	0	4827	4847	180	10020	365	0	67	10	59	497	291	0	54	64	2	238	365	0
35	1.7	5	95	102	26	9	122	128	0	3584	4073	264	7806	175	0	60	18	50	659	128	0	48	61	3	243	175	0
36	1.6	3	42	99	35	24	114	123	0	5730	4920	199	9269	216	0	57	27	50	233	123	0	63	81	4	207	216	0
37	1.2	3	39	101	50	14	127	116	0	3356	3196	180	5664	185	0	44	35	35	317	116	0	35	67	2	139	185	0
38	7.1	17	39	98	29	20	118	105	0	5615	5033	123	7960	153	0	263	91	219	1264	105	0	59	72	2	164	153	0
39	0.7	2	13	100	13	20	117	423	0	4100	3685	87	6808	171	0	26	4	22	130	423	0	43	42	1	108	470	0
40	0.6	2	25	102	13	6	123	543	1	3286	3021	60	4934	590	1	23	3	19	389	543	1	33	33	1	146	733	1
41	1.9	5	19	101	14	8	120	734	1	4640	4372	60	7387	509	1	67	9	56	844	734	1	47	48	2	192	827	1

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	17	96	22	6	121	751	1	4547	4341	60	7560	550	1	49	12	39	823	751	1	48	53	1	262	930	1
43	4.6	11	18	99	12	6	120	635	1	4528	4296	60	7061	465	0	166	20	137	2656	635	1	47	48	1	201	790	1
44	0.8	1	12	103	10	8	121	312	0	4590	4336	60	7215	536	1	28	2	24	399	312	0	45	45	1	344	655	1
45	1.6	5	9	103	10	8	121	0	0	4705	4526	187	7814	220	0	57	6	48	811	0	0	46	47	2	184	220	0
46	0.2	1	8	103	10	5	121	121	0	4478	4326	60	7597	3496	3	9	1	7	176	121	0	44	45	1	198	3496	3
47	0.9	3	7	100	11	15	116	112	0	4963	4864	107	8551	177	0	33	3	28	224	112	0	50	52	2	177	177	0
48	1.4	4	8	102	10	22	116	107	0	4941	4826	114	8520	239	0	48	5	42	226	107	0	49	51	1	157	239	0
49	1.8	4	8	104	9	31	121	107	0	4959	4835	109	8726	242	0	64	5	55	213	107	0	48	50	1	141	242	0
50	2.3	7	5	105	8	19	126	118	0	4296	4169	103	7371	183	0	79	6	66	437	118	0	41	42	1	170	188	0
51	0.9	3	4	105	10	9	124	136	0	3771	3679	180	6235	221	0	31	3	27	360	136	0	36	38	2	141	221	0
52	3.8	12	11	106	8	14	122	105	0	4436	4351	198	8256	167	0	130	11	113	983	105	0	42	44	2	146	167	0
53	1.0	3	7	109	8	10	127	133	0	2877	3016	84	6105	192	0	32	3	28	347	133	0	26	29	1	138	192	0
54	0.4	2	7	105	7	3	124	189	0	2160	2352	62	5520	307	0	14	1	12	478	189	0	20	23	1	142	307	0
55	0.5	1	10	104	6	5	119	128	0	3413	3351	90	7560	282	0	18	1	16	379	128	0	33	34	1	156	282	0
56	6.1	16	28	107	8	9	120	105	0	3723	3687	87	7210	159	0	205	16	182	2422	105	0	35	37	2	151	159	0
57	0.8	3	33	106	8	5	122	106	0	3289	3232	108	6397	174	0	26	2	22	542	106	0	32	33	1	156	174	0
58	4.0	12	69	104	8	7	118	106	0	3684	3631	60	7054	173	0	139	10	123	2072	106	0	36	38	3	179	173	0
59	0.9	3	40	106	9	5	130	204	0	2862	2772	60	5778	310	0	32	3	26	682	204	0	28	28	0	209	310	0
60	3.4	11	60	101	12	5	129	204	0	-	-	-	-	100800	100	120	14	95	2436	204	0	-	-	-	-	100800	100
61	0.8	2	50	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100
62	7.8	20	139	85	35	14	111	99276	98	-	-	-	-	100800	100	329	160	253	2080	99276	98	-	-	-	-	100800	100
63	1.0	3	17	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100
64	4.7	11	47	91	7	11	110	86362	86	-	-	-	-	100800	100	187	14	155	1618	86362	86	-	-	-	-	100800	100
65	0.6	2	20	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100	-	-	-	-	100800	100
66	1.8	4	34	104	8	6	124	308	0	-	-	-	-	100800	100	62	4	52	1074	308	0	-	-	-	-	100800	100
67	4.5	11	46	102	7	12	118	110	0	3387	3177	60	6694	209	0	158	11	136	1328	110	0	33	32	1	160	261	0
68	0.9	2	12	105	8	8	125	30617	30	-	-	-	-	100800	100	31	2	26	406	30617	30	-	-	-	-	100800	100
69	0.5	1	14	102	7	16	119	123	0	-	-	-	-	100800	100	16	1	14	104	123	0	-	-	-	-	100800	100
70	1.9	6	9	105	8	11	121	111	0	3677	3631	93	7782	171	0	66	5	57	628	111	0	35	37	1	151	171	0
71	9.5	27	30	106	9	16	122	119	0	3679	3627	240	7655	169	0	324	27	281	2173	119	0	34	36	2	129	169	0
72	0.9	3	33	102	8	7	118	791	1	2518	2427	60	5315	1676	2	31	3	26	423	791	1	24	25	1	135	1916	2
73	3.4	10	15	103	8	18	119	2858	3	3523	3464	60	6797	1112	1	121	9	104	685	2858	3	35	35	1	144	2944	3
74	1.1	4	12	106	9	10	124	4015	4	2669	2435	60	5281	4098	4	36	3	31	399	4015	4	25	24	1	109	5055	5
75	7.8	24	29	105	7	11	118	4244	4	3199	3068	60	6113	1502	1	267	19	236	2541	4244	4	32	30	1	151	4305	4
76	1.2	2	41	102	7	3	122	3265	3	2176	1980	60	4410	3255	3	42	3	34	1376	3265	3	22	20	1	234	3577	4

Table B.1: Summary statistics of data, computed using data between April 7th 2017 and June 16th 2017. This subset is discussed in section 4.2 for an initial exploratory analysis of our incident detection methodology. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.2 Data between April 7th 2017 and April 27th 2017

In table B.2, we show summary statistics for the subset of data recorded between April 7th 2017 and April 27th 2017. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. Firstly, as this data is a subset of the data discussed in appendix B.4.1, the same comments regarding missing data hold here. Similarly, the same comments regarding expected difference in weather conditions hold. If we compare median speed values, we see that the largest differences between tables 3.1 and B.2 are seen for links 8, 41, 42 and 64. There may have been a systematic reason for the reduction in median speeds we observe considering two of the most significant changes are seen at neighbouring spatial locations. Median flows are generally consistent across the full data and this subset, with minor decreases on parts of the network. Median density values are also consistent across both datasets, with the largest difference being 4 veh/km on link 42.

Considering the spread of data, we see the largest differences in inter-quartile range are found on links 28, 37, 38, 42 and 43 when comparing tables 3.1 and B.2. It may be that during this shorter collection period, significant incidents can impact traffic speed, and their impact is more noticeable through changes in inter-quartile range when looking at small subsets of data. Such events would also likely effect neighbouring links, which may be why we see links 37 and 38, along with links 42 and 43 showing the largest changes in inter-quartile range. Links 37 and 38 similarly show some of the largest changes in flow and density inter-quartile range.

This dataset, along with those discussed in appendix B.4.3 and B.4.4 is used to verify stationarity of the typical density-flow relationship in time, and exploratory analysis of our incident detection methodology in chapter 4.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	6	101	5	13	112	0	0	2005	1978	145	3988	0	0	26	1	24	202	0	0	20	20	1	123	0	0
1	3.6	13	8	103	7	21	113	0	0	3420	3328	240	6788	0	0	127	9	116	623	0	0	33	34	2	158	0	0
2	1.0	3	4	108	6	29	120	0	0	2736	2645	145	5653	0	0	33	2	30	333	0	0	25	26	1	98	0	0
3	0.5	2	4	107	6	35	119	185	1	2276	2194	60	4796	89	0	17	1	16	97	185	1	21	21	1	87	185	1
4	0.4	1	4	107	6	35	119	91	0	2280	2210	60	4949	94	0	14	1	12	76	91	0	21	21	1	99	94	0
5	4.5	2	8	109	7	46	128	0	0	2936	2911	180	6164	0	0	149	10	127	1800	0	0	26	28	2	99	0	0
6	0.9	2	2	111	6	31	126	0	0	2319	2235	136	4705	0	0	28	2	25	103	0	0	21	21	1	90	0	0
7	5.1	2	7	112	7	22	124	0	0	2629	2571	168	5374	0	0	164	10	148	826	0	0	23	23	2	105	0	0
8	1.0	1	7	97	7	5	119	8581	28	1709	1680	60	3476	0	0	39	2	32	753	8581	28	21	10	2	380	8581	28
9	0.7	1	7	101	6	11	111	0	0	3087	2894	60	7001	0	0	23	1	21	213	0	0	30	29	1	156	0	0
10	0.9	2	11	101	6	4	120	0	0	3560	3397	180	7833	0	0	32	2	27	798	0	0	34	34	2	385	0	0
11	5.9	13	37	102	9	3	129	9809	32	-	-	-	-	30240	100	208	18	165	7081	9809	32	-	-	-	-	30240	100
12	1.0	2	8	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
13	6.8	17	31	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
14	1.0	2	4	105	9	27	122	12709	42	-	-	-	-	30240	100	34	3	29	131	12709	42	-	-	-	-	30240	100
15	1.2	4	6	106	7	18	119	0	0	3784	3618	144	7246	0	0	40	3	36	232	0	0	35	36	2	151	0	0
16	1.5	6	6	108	7	24	120	0	0	3681	3525	144	7045	0	0	51	3	45	227	0	0	33	34	2	149	0	0
17	0.5	0	2	109	7	20	125	0	0	3680	3540	155	7140	0	0	16	1	14	135	0	0	33	33	1	156	0	0
18	0.2	1	2	110	8	14	127	0	0	2564	2532	60	5496	0	0	6	0	5	61	0	0	23	24	1	111	0	0
19	1.2	4	3	110	8	13	127	0	0	2426	2418	60	5380	0	0	40	3	35	411	0	0	22	23	1	115	0	0
20	2.0	4	3	106	7	14	119	0	0	4211	4075	240	7885	0	0	69	5	61	537	0	0	40	41	2	173	0	0
21	3.0	6	3	106	8	7	118	0	0	3466	3394	240	7129	0	0	101	7	91	1465	0	0	33	35	2	142	0	0
22	10.0	22	18	112	7	23	126	0	0	3936	3775	275	8026	0	0	321	20	286	1580	0	0	35	36	2	114	0	0
23	0.5	1	8	113	8	27	127	0	0	3408	3257	240	7382	0	0	16	1	15	69	0	0	31	32	2	112	0	0
24	6.4	13	29	111	9	16	125	0	0	3780	3726	240	7854	0	0	208	17	185	1446	0	0	35	38	2	139	0	0
25	0.8	2	13	109	13	5	125	0	0	-	-	-	-	30240	100	26	3	22	517	0	0	-	-	-	-	30240	100
26	2.6	9	12	108	16	4	122	0	0	3868	3774	69	7620	0	0	85	13	75	2166	0	0	36	42	2	146	0	0
27	0.9	2	12	109	14	5	124	394	1	2905	2850	60	6004	283	1	29	4	25	723	394	1	27	31	1	115	403	1
28	6.9	13	25	103	25	5	120	414	1	4465	4412	60	8174	222	1	241	66	207	4929	414	1	43	57	3	162	414	1
29	0.9	3	11	106	23	13	130	394	1	3801	3897	60	7106	258	1	32	9	26	258	394	1	38	52	0	140	403	1
30	2.5	4	6	99	19	18	116	0	0	4677	4775	170	8277	0	0	91	19	78	502	0	0	47	62	3	152	0	0
31	1.2	1	4	103	12	18	120	0	0	3149	3212	146	6095	0	0	42	5	36	235	0	0	32	36	1	117	0	0
32	3.9	6	21	104	10	20	119	0	0	4531	4777	300	8854	0	0	134	13	117	690	0	0	47	55	3	214	0	0
33	1.0	3	33	108	12	5	122	0	0	3548	4321	275	8302	0	0	33	3	29	683	0	0	43	52	3	275	0	0
34	1.9	3	26	106	17	18	119	0	0	4818	4961	360	9811	0	0	66	12	59	381	0	0	53	68	3	224	0	0
35	1.7	5	30	102	34	13	119	0	0	3587	4191	305	7806	0	0	60	27	52	472	0	0	48	68	3	243	0	0
36	1.6	3	12	98	40	37	114	0	0	5768	5070	360	9269	0	0	58	33	50	153	0	0	62	92	4	172	0	0
37	1.2	3	15	100	57	16	120	0	0	3442	3271	180	5626	0	0	44	45	37	277	0	0	36	75	2	125	0	0
38	7.1	17	15	98	30	27	118	0	0	5649	5166	339	7620	0	0	264	94	219	941	0	0	58	73	3	164	0	0
39	0.7	2	4	100	13	26	116	299	1	4141	3761	258	5847	0	0	26	4	22	100	299	1	43	42	3	108	299	1
40	0.6	2	10	103	12	14	123	420	1	3362	3070	60	4897	358	1	23	3	19	162	420	1	34	34	1	146	501	2
41	1.9	5	7	103	13	23	120	629	2	4667	4422	60	7335	311	1	66	9	56	300	629	2	47	47	3	169	629	2

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	8	103	15	15	121	629	2	4542	4393	60	7227	249	1	46	6	39	314	629	2	46	48	3	200	629	2
43	4.6	11	5	101	17	16	120	629	2	4539	4336	60	7052	304	1	162	28	137	1048	629	2	46	47	3	150	629	2
44	0.8	1	6	104	10	13	121	299	1	4636	4376	60	7140	432	1	27	3	24	220	299	1	45	45	1	340	538	2
45	1.6	5	5	104	10	13	120	0	0	4652	4589	240	7673	0	0	56	6	49	450	0	0	45	48	2	184	0	0
46	0.2	1	3	104	10	10	119	0	0	4427	4395	208	7597	0	0	9	1	7	89	0	0	43	46	2	198	0	0
47	0.9	3	3	100	11	15	115	0	0	4884	4924	218	8457	0	0	33	3	29	224	0	0	49	54	2	177	0	0
48	1.4	4	4	102	11	22	116	0	0	4863	4888	240	8445	0	0	48	5	42	226	0	0	48	52	2	157	0	0
49	1.8	4	4	104	10	31	121	0	0	4875	4890	264	8534	0	0	64	6	55	213	0	0	47	51	2	141	0	0
50	2.3	7	1	105	9	19	120	0	0	4227	4235	240	7294	0	0	79	7	69	437	0	0	40	44	2	170	0	0
51	0.9	3	2	106	11	9	122	0	0	3718	3739	194	6197	0	0	31	3	27	360	0	0	35	39	2	141	0	0
52	3.8	12	6	107	10	14	122	0	0	4353	4424	240	7237	0	0	129	12	113	983	0	0	41	45	2	146	0	0
53	1.0	3	2	109	8	10	127	0	0	2825	3076	84	5462	0	0	32	3	28	347	0	0	26	30	1	138	0	0
54	0.4	2	4	105	7	3	117	0	0	2129	2402	63	4552	0	0	14	1	13	478	0	0	20	24	1	142	0	0
55	0.5	1	4	105	7	5	119	0	0	3387	3371	90	6526	0	0	18	1	16	379	0	0	33	34	2	156	0	0
56	6.1	16	7	108	8	13	120	0	0	3673	3727	87	7210	0	0	203	15	182	1663	0	0	35	37	3	137	0	0
57	0.8	3	5	106	7	5	122	0	0	3301	3276	147	6397	0	0	26	2	22	542	0	0	32	33	1	156	0	0
58	4.0	12	13	105	8	7	118	0	0	3700	3697	60	7054	0	0	138	11	123	1983	0	0	36	38	3	143	0	0
59	0.9	3	8	107	9	5	130	0	0	2880	2820	60	5778	0	0	32	3	26	682	0	0	28	28	0	154	0	0
60	3.4	11	11	102	12	5	129	0	0	-	-	-	-	30240	100	119	14	95	2436	0	0	-	-	-	-	30240	100
61	0.8	2	16	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
62	7.8	20	46	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
63	1.0	3	4	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
64	4.7	11	14	90	11	11	110	26330	87	-	-	-	-	30240	100	190	23	155	1618	26330	87	-	-	-	-	30240	100
65	0.6	2	5	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
66	1.8	4	4	105	7	18	120	0	0	-	-	-	-	30240	100	61	4	54	365	0	0	-	-	-	-	30240	100
67	4.5	11	12	103	7	29	115	0	0	3351	3223	300	6694	0	0	156	10	140	558	0	0	32	33	3	122	0	0
68	0.9	2	2	105	8	12	124	9809	32	-	-	-	-	30240	100	31	2	26	275	9809	32	-	-	-	-	30240	100
69	0.5	1	3	102	7	26	119	0	0	-	-	-	-	30240	100	16	1	14	62	0	0	-	-	-	-	30240	100
70	1.9	6	2	106	8	11	121	0	0	3641	3696	230	7782	0	0	65	5	57	628	0	0	34	37	2	151	0	0
71	9.5	27	8	106	8	39	122	0	0	3629	3678	240	7655	0	0	324	25	281	886	0	0	33	36	2	113	0	0
72	0.9	3	1	104	8	28	118	0	0	2397	2482	120	5220	0	0	29	2	26	109	0	0	22	25	1	115	0	0
73	3.4	10	6	102	8	32	116	0	0	3364	3520	64	6727	0	0	121	9	107	388	0	0	32	36	1	127	0	0
74	1.1	4	7	106	8	10	124	1066	4	2769	2479	60	4920	1809	6	36	2	31	399	1066	4	25	25	1	106	1873	6
75	7.8	24	11	104	8	17	118	2760	9	3191	3055	60	6008	933	3	268	19	237	1642	2760	9	32	29	1	139	2760	9
76	1.2	2	10	102	7	3	122	2241	7	2254	1909	60	4254	2231	7	42	3	34	1376	2241	7	22	19	1	234	2372	8

Table B.2: Summary statistics of data, computed using data between 7th April 2017 and 27th April 2017. This subset is discussed in sections 4.3.3 and 4.5.2. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.3 Data between April 28th 2017 and May 18th 2017

In table B.3, we show summary statistics for data collected between April 28th 2017 and May 18th 2017. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. As this data is again a subset of the data discussed in appendix B.4.1, the same comments regarding missing data and expected difference in weather conditions hold here. If we compare median speed values, we see that the largest differences between tables 3.1 and B.2 are seen for links 8, 29, 41, 42 and 64, which is generally in-line with appendix B.4.2. Median flows are generally consistent across the full data and this subset, with peak increases of 100 veh/hr around links 29 and 31. The most significant decreases in median flows are seen on links 36 and 38, with differences being around 200 veh/hr. As in appendix B.4.2, link 42 shows the most significant difference in median density values.

Considering the spread of data, the largest differences in inter-quartile range of speeds are found on links 35, 36 and 37 when comparing tables 3.1 and B.2. The spatial structure here clearly suggests that this is either due to major incidents impacting this part of the network in the collection window, or a similar physical aspect altering the traffic flow, rather than random variation. Similarly, the largest differences in inter-quartile range for flow and density are also observed on these links.

This dataset, along with those discussed in appendix B.4.2 and B.4.4 is used to verify stationarity of the typical density-flow relationship in time, and exploratory analysis of our incident detection methodology in chapter 4.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	4	100	5	20	110	165	1	2017	1879	60	4056	224	1	26	1	24	134	165	1	20	19	1	116	224	1
1	3.6	13	9	102	7	25	113	170	1	3472	3314	194	6863	217	1	128	9	116	516	170	1	34	34	2	138	217	1
2	1.0	3	1	107	7	34	120	76	0	2781	2647	60	5766	242	1	33	2	30	105	76	0	26	26	1	113	242	1
3	0.5	2	2	106	7	18	123	136	0	2286	2189	120	4980	341	1	17	1	15	101	136	0	22	22	1	117	341	1
4	0.4	1	2	106	7	18	123	136	0	2300	2209	120	4860	143	0	14	1	12	79	136	0	22	22	1	114	161	1
5	4.5	2	7	109	7	23	123	78	0	2922	2870	180	6360	269	1	150	10	133	695	78	0	27	28	2	135	274	1
6	0.9	2	2	110	7	35	129	260	1	2314	2211	125	4757	310	1	29	1	24	92	260	1	21	21	1	72	310	1
7	5.1	2	8	112	7	39	125	0	0	2640	2550	169	5166	316	1	165	10	147	662	0	0	23	23	2	94	316	1
8	1.0	1	8	96	7	14	129	8646	29	1725	1679	60	3412	143	0	39	3	29	274	8646	29	21	9	1	183	8725	29
9	0.7	1	5	101	6	15	128	312	1	3073	2893	240	7071	415	1	23	1	18	156	312	1	30	29	2	153	415	1
10	0.9	2	7	101	6	11	130	633	2	3558	3381	240	7891	148	0	32	2	25	289	633	2	35	34	2	326	718	2
11	5.9	13	28	102	8	8	130	9226	31	-	-	-	-	30240	100	207	17	163	2668	9226	31	-	-	-	-	30240	100
12	1.0	2	5	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
13	6.8	17	19	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
14	1.0	2	1	105	8	33	122	10445	35	-	-	-	-	30240	100	34	3	29	109	10445	35	-	-	-	-	30240	100
15	1.2	4	3	107	6	34	121	181	1	3727	3630	156	7516	316	1	40	2	35	126	181	1	34	35	1	154	316	1
16	1.5	6	2	108	6	37	120	63	0	3660	3579	236	7375	275	1	51	2	46	149	63	0	33	34	2	138	275	1
17	0.5	0	2	110	7	41	125	76	0	3680	3625	80	7317	145	0	15	1	14	41	76	0	33	34	1	135	158	1
18	0.2	1	2	110	7	44	124	73	0	2473	2535	110	5687	283	1	6	0	5	15	73	0	22	23	1	94	283	1
19	1.2	4	2	110	8	40	124	73	0	2327	2399	99	5388	275	1	40	3	36	110	73	0	21	22	1	96	275	1
20	2.0	4	6	106	7	51	117	292	1	4151	3910	195	8251	234	1	69	4	62	142	292	1	39	39	2	133	363	1
21	3.0	6	4	106	8	34	118	292	1	3474	3312	180	7143	759	3	101	7	91	379	292	1	32	34	2	122	759	3
22	10.0	22	18	112	7	28	123	163	1	3950	3738	300	8335	206	1	322	21	294	1287	163	1	35	36	3	123	206	1
23	0.5	1	6	113	8	18	130	210	1	3401	3238	60	7149	264	1	16	1	14	103	210	1	30	31	1	132	264	1
24	6.4	13	24	111	9	30	123	166	1	3837	3740	300	7642	214	1	208	16	187	778	166	1	35	37	3	144	214	1
25	0.8	2	14	109	11	17	123	173	1	-	-	-	-	30240	100	26	3	23	161	173	1	-	-	-	-	30240	100
26	2.6	9	15	108	12	22	122	173	1	3901	3783	120	7980	225	1	85	10	75	416	173	1	36	40	1	161	225	1
27	0.9	2	9	108	13	12	125	206	1	2834	2865	60	6120	457	2	29	3	25	260	206	1	26	30	1	144	457	2
28	6.9	13	26	105	16	17	121	740	2	4385	4400	60	8247	423	1	237	38	206	1521	740	2	43	48	2	161	783	3
29	0.9	3	7	106	19	18	130	421	1	3720	3845	60	7260	593	2	32	6	26	186	421	1	37	45	1	151	593	2
30	2.5	4	4	100	15	28	122	170	1	4686	4716	239	8335	278	1	90	14	74	327	170	1	47	56	2	137	278	1
31	1.2	1	4	103	10	28	119	181	1	3121	3172	256	6240	248	1	42	4	36	154	181	1	32	34	3	89	248	1
32	3.9	6	21	107	9	18	119	166	1	4471	4719	303	8747	214	1	130	11	117	754	166	1	44	50	3	197	214	1
33	1.0	3	27	108	10	8	122	181	1	3616	4243	187	8100	229	1	33	3	29	518	181	1	40	47	2	201	229	1
34	1.9	3	26	106	11	21	119	181	1	4835	4910	180	10020	240	1	66	8	59	334	181	1	51	58	2	238	240	1
35	1.7	5	26	103	18	13	120	115	0	3617	4140	315	7772	157	1	60	12	52	471	115	0	46	54	3	170	157	1
36	1.6	3	9	100	23	31	112	110	0	5686	5006	390	9209	198	1	57	15	50	181	110	0	61	69	4	207	198	1
37	1.2	3	9	103	19	15	119	103	0	3288	3220	180	5664	167	1	43	9	37	307	103	0	33	43	2	124	167	1
38	7.1	17	10	99	23	22	117	92	0	5510	5036	360	7960	135	0	259	65	220	1178	92	0	58	65	3	152	135	0
39	0.7	2	5	100	12	20	117	111	0	3995	3699	180	6808	153	1	26	3	22	130	111	0	41	41	2	103	153	1
40	0.6	2	9	103	11	6	119	110	0	3145	2999	82	4874	214	1	23	2	20	389	110	0	31	33	1	141	214	1
41	1.9	5	7	103	11	8	118	92	0	4487	4355	240	7200	180	1	66	7	57	844	92	0	45	46	2	192	180	1

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	3	101	15	6	118	109	0	4418	4321	68	7560	279	1	47	7	41	823	109	0	44	49	1	167	279	1
43	4.6	11	6	96	11	10	118	6	0	4408	4266	94	7061	143	0	171	20	140	1721	6	0	46	49	2	201	143	0
44	0.8	1	3	103	10	8	119	13	0	4475	4316	60	7174	91	0	28	2	24	399	13	0	44	45	1	344	104	0
45	1.6	5	2	103	10	8	121	0	0	4633	4510	187	7814	202	1	57	6	48	811	0	0	45	47	2	167	202	1
46	0.2	1	3	103	9	7	121	108	0	4405	4284	120	7451	198	1	9	1	7	138	108	0	43	44	1	190	198	1
47	0.9	3	1	100	10	41	114	99	0	4865	4828	251	8551	158	1	33	3	28	80	99	0	49	52	2	142	158	1
48	1.4	4	1	101	10	49	115	99	0	4842	4785	139	8520	221	1	49	4	43	102	99	0	48	51	1	135	221	1
49	1.8	4	1	103	10	44	118	99	0	4868	4803	240	8726	189	1	65	6	56	185	99	0	47	50	2	128	189	1
50	2.3	7	3	105	8	26	120	110	0	4223	4123	235	7371	165	1	79	6	69	382	110	0	40	42	2	145	170	1
51	0.9	3	1	105	9	30	124	123	0	3682	3623	180	6235	203	1	31	3	27	110	123	0	35	37	2	119	203	1
52	3.8	12	4	105	8	61	120	92	0	4373	4287	240	8256	149	0	131	10	115	306	92	0	41	43	2	88	149	0
53	1.0	3	1	108	8	47	123	120	0	2836	2970	120	6105	174	1	33	3	28	74	120	0	26	29	1	87	174	1
54	0.4	2	1	105	6	37	115	176	1	2098	2331	64	5520	289	1	14	1	13	41	176	1	20	23	1	72	289	1
55	0.5	1	2	104	6	14	119	115	0	3275	3282	120	7560	264	1	18	1	16	134	115	0	31	33	1	101	264	1
56	6.1	16	7	107	7	9	117	92	0	3571	3589	197	7204	141	0	205	15	187	2422	92	0	34	35	2	151	141	0
57	0.8	3	10	106	7	6	121	93	0	3168	3175	108	6337	156	1	26	2	22	452	93	0	30	32	2	155	156	1
58	4.0	12	20	104	8	7	117	93	0	3578	3578	236	7017	155	1	139	10	124	2072	93	0	35	36	3	140	155	1
59	0.9	3	13	105	8	6	123	191	1	2735	2723	76	5639	292	1	32	3	28	569	191	1	27	27	2	209	292	1
60	3.4	11	23	100	11	5	123	191	1	-	-	-	-	30240	100	122	13	99	2436	191	1	-	-	-	-	30240	100
61	0.8	2	16	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
62	7.8	20	46	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
63	1.0	3	6	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
64	4.7	11	13	90	10	19	108	26068	86	-	-	-	-	30240	100	191	20	158	880	26068	86	-	-	-	-	30240	100
65	0.6	2	5	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
66	1.8	4	15	104	8	22	117	274	1	-	-	-	-	30240	100	62	4	55	294	274	1	-	-	-	-	30240	100
67	4.5	11	15	102	7	30	114	89	0	3255	3110	236	6469	188	1	158	10	141	547	89	0	32	31	2	124	219	1
68	0.9	2	3	105	8	25	122	9254	31	-	-	-	-	30240	100	31	2	27	149	9254	31	-	-	-	-	30240	100
69	0.5	1	5	102	7	37	117	109	0	-	-	-	-	30240	100	16	1	14	44	109	0	-	-	-	-	30240	100
70	1.9	6	2	105	8	36	118	98	0	3589	3532	310	7581	152	1	66	5	59	193	98	0	34	35	3	105	152	1
71	9.5	27	8	106	9	16	118	106	0	3592	3536	335	7547	151	0	324	28	291	2173	106	0	34	35	3	127	151	0
72	0.9	3	12	103	9	10	117	538	2	2454	2341	60	5315	697	2	30	3	26	309	538	2	24	24	1	116	937	3
73	3.4	10	6	102	8	43	114	1422	5	3479	3424	60	6797	490	2	122	10	109	288	1422	5	35	34	2	112	1503	5
74	1.1	4	3	105	9	14	124	1803	6	2623	2368	60	4842	1655	5	37	4	31	276	1803	6	25	23	1	58	2031	7
75	7.8	24	6	105	7	25	117	1484	5	3185	3045	60	6113	551	2	266	17	238	1111	1484	5	31	30	1	121	1527	5
76	1.2	2	10	102	6	6	122	1024	3	2105	1976	60	4410	1006	3	42	2	34	695	1024	3	21	20	1	148	1187	4

Table B.3: Summary statistics of data, computed using data between 28th April 2017 and 18th May 2017. This subset is discussed in sections 4.3.3 and 4.5.2. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.4 Data between May 19th 2017 and June 8th 2017

In table B.4, we show summary statistics for data collected between May 19th 2017 and June 8th 2017. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. As this data is again a subset of the data discussed in appendix B.4.1, the same comments regarding missing data and expected difference in weather conditions hold here. However, unlike the subsets discussed in appendix B.4.2 and B.4.3, link 25 reports mostly non-missing flow values. Comparing median speed values, the largest differences between tables 3.1 and B.4 are observed on links 8, 42, 49, 62 and 64. Median flow values are lower on almost all links during this subset, suggesting there was reduced demand on a network wide scale in this short time-window compared to the entire collection period. The same is true for density, as it is derived using speed and flow measurements.

Comparing variation, we see that the largest changes in inter-quartile range of speed values are found on links 34, 35, 37, 62 and 64. There is an increase in the speed inter-quartile range across links 30 to 40, again showing spatial structure in the changes on the network. This could point to significant events causing larger speed drops than typical, and these being more noticeable in short windows of data. There is no systematic pattern in the change in the inter-quartile range of flow values across links, with some being higher than in the full dataset and some being lower. As expected, links 30 to 40 generally show an increase in the inter-quartile range of their density values compared to the full dataset.

Note that the FA cup final took place at Wembley Stadium on May 27th 2017, between two clubs located in London, Chelsea and Arsenal. This may have impacted traffic on the M25, particularly on the western side, meaning we would observe what appeared to be a particularly busy Saturday morning and afternoon. This is the only event we are aware of occurring in the collection window that would significantly impact the dataset, in-particular when compared to those discussed in appendix B.4.2 and B.4.3.

This dataset, along with those discussed in appendix B.4.2 and B.4.4 is used to verify stationarity of the typical density-flow relationship in time, and exploratory analysis of our incident detection methodology in chapter 4.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	1	100	6	31	111	103	0	2095	1997	60	3929	329	1	26	1	24	84	103	0	21	21	1	88	329	1
1	3.6	13	7	102	7	37	114	103	0	3623	3373	240	6805	119	0	128	10	115	354	103	0	36	35	2	129	119	0
2	1.0	3	2	107	7	36	122	102	0	2898	2627	186	5524	125	0	33	2	29	101	102	0	27	26	2	102	125	0
3	0.5	2	0	106	7	45	124	102	0	2402	2169	180	4676	125	0	17	1	15	41	102	0	23	22	2	83	125	0
4	0.4	1	0	106	7	45	124	102	0	2404	2186	120	4695	118	0	14	1	12	32	102	0	23	22	1	84	121	0
5	4.5	2	1	110	8	28	124	97	0	3053	2872	240	6040	125	0	148	11	131	587	97	0	28	28	2	126	125	0
6	0.9	2	2	110	7	11	124	111	0	2415	2216	71	4786	125	0	29	1	25	289	111	0	22	21	2	118	125	0
7	5.1	2	5	110	7	11	126	16	0	2730	2545	226	5037	125	0	168	11	146	1683	16	0	25	24	2	166	125	0
8	1.0	1	6	94	8	5	130	8670	29	1756	1670	108	3240	126	0	40	3	29	753	8670	29	22	10	2	381	8691	29
9	0.7	1	9	101	7	5	128	97	0	3129	2827	240	7097	125	0	23	2	18	469	97	0	30	29	3	279	125	0
10	0.9	2	15	104	9	5	130	8673	29	3624	3303	244	7980	118	0	31	3	25	638	8673	29	41	17	6	390	8684	29
11	5.9	13	50	101	12	8	130	8673	29	-	-	-	-	30240	100	210	25	163	2770	8673	29	-	-	-	-	30240	100
12	1.0	2	15	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
13	6.8	17	41	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
14	1.0	2	3	106	9	30	120	116	0	-	-	-	-	30240	100	34	3	30	118	116	0	-	-	-	-	30240	100
15	1.2	4	4	108	8	28	120	102	0	3862	3620	300	7038	125	0	40	3	35	155	102	0	36	35	3	162	125	0
16	1.5	6	5	108	7	37	121	102	0	3806	3584	300	7002	124	0	51	3	45	149	102	0	35	34	3	111	124	0
17	0.5	0	2	110	7	25	126	97	0	3840	3600	243	7020	119	0	15	1	14	70	97	0	34	34	2	169	119	0
18	0.2	1	1	110	8	18	125	110	0	2568	2538	120	5304	127	0	6	0	5	36	110	0	23	24	1	117	127	0
19	1.2	4	3	109	8	17	126	110	0	2406	2422	120	5205	124	0	41	3	35	269	110	0	22	23	1	116	124	0
20	2.0	4	6	105	9	23	117	111	0	4325	3921	240	8008	125	0	69	6	62	310	111	0	41	41	2	200	125	0
21	3.0	6	5	104	9	14	123	110	0	3588	3313	180	7074	125	0	103	9	87	802	110	0	34	34	2	165	125	0
22	10.0	22	13	112	8	38	124	110	0	4095	3765	360	8064	124	0	323	25	289	958	110	0	37	36	3	123	124	0
23	0.5	1	8	112	10	21	128	111	0	3344	3116	300	7024	127	0	16	1	15	88	111	0	30	31	3	148	127	0
24	6.4	13	32	110	12	23	124	110	0	3967	3688	336	7792	124	0	209	24	186	990	110	0	37	41	3	139	124	0
25	0.8	2	17	108	15	24	122	111	0	3655	3321	300	7411	9259	31	26	4	23	117	111	0	36	41	3	152	9259	31
26	2.6	9	17	105	16	29	122	110	0	4092	3742	360	7305	125	0	87	14	75	323	110	0	39	46	3	148	125	0
27	0.9	2	13	105	17	20	122	110	0	2964	2880	240	5874	125	0	30	5	26	158	110	0	29	34	2	111	125	0
28	6.9	13	31	103	27	26	121	110	0	4484	4334	300	7795	124	0	242	74	206	974	110	0	44	58	3	161	124	0
29	0.9	3	4	104	25	23	123	110	0	3830	3808	300	6705	125	0	33	8	27	149	110	0	39	51	3	116	125	0
30	2.5	4	1	98	17	32	119	110	0	4826	4618	360	8012	124	0	92	16	76	278	110	0	49	56	4	143	124	0
31	1.2	1	2	103	13	25	119	116	0	3180	3101	300	5952	125	0	42	6	36	171	116	0	33	35	3	98	125	0
32	3.9	6	27	107	11	18	120	111	0	4504	4564	360	8456	125	0	130	15	116	770	111	0	46	53	3	202	125	0
33	1.0	3	41	107	16	12	121	110	0	3552	4046	180	7811	125	0	33	5	29	298	110	0	44	52	3	180	125	0
34	1.9	3	40	103	29	15	118	110	0	4800	4695	288	9463	125	0	68	24	59	497	110	0	58	73	4	233	125	0
35	1.7	5	33	101	49	9	122	13	0	3521	3920	264	7659	18	0	61	46	50	659	13	0	51	80	3	173	18	0
36	1.6	3	18	98	43	24	112	13	0	5720	4728	199	9071	18	0	58	37	50	233	13	0	68	91	4	188	18	0
37	1.2	3	14	100	59	14	127	13	0	3350	3104	227	5445	18	0	44	48	35	317	13	0	37	76	2	139	18	0
38	7.1	17	12	97	32	20	117	13	0	5644	4916	420	7774	18	0	266	102	221	1264	13	0	63	73	4	163	18	0
39	0.7	2	4	98	14	21	116	13	0	4145	3601	300	6013	18	0	26	4	22	123	13	0	43	42	3	105	18	0
40	0.6	2	6	100	14	6	122	13	0	3337	3005	120	4934	18	0	23	3	19	389	13	0	34	34	2	131	18	0
41	1.9	5	5	99	16	11	120	13	0	4722	4340	60	7387	18	0	68	11	56	602	13	0	48	49	2	153	18	0

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	6	84	16	6	111	13	0	4637	4311	60	7212	22	0	57	10	43	795	13	0	56	57	1	262	22	0
43	4.6	11	6	98	11	6	118	0	0	4602	4280	60	6923	18	0	167	19	139	2656	0	0	48	47	1	150	18	0
44	0.8	1	2	103	10	9	118	0	0	4649	4309	288	7215	13	0	28	2	24	322	0	0	46	45	3	340	13	0
45	1.6	5	1	103	10	9	118	0	0	4796	4485	263	7729	18	0	57	6	49	663	0	0	47	47	3	173	18	0
46	0.2	1	1	101	11	5	119	13	0	4523	4326	60	7404	3074	10	9	1	7	176	13	0	45	46	3	150	3074	10
47	0.9	3	3	101	10	18	116	13	0	5053	4828	107	8442	19	0	32	4	28	181	13	0	50	51	3	134	19	0
48	1.4	4	3	103	9	27	116	8	0	5036	4796	114	8362	18	0	48	4	42	186	8	0	49	50	3	113	18	0
49	1.8	4	3	104	8	64	117	8	0	5034	4803	109	8412	53	0	64	4	56	104	8	0	48	49	1	107	53	0
50	2.3	7	1	104	8	59	123	8	0	4368	4161	103	7213	18	0	79	6	67	142	8	0	42	42	1	99	18	0
51	0.9	3	0	103	10	61	120	13	0	3858	3674	201	6153	18	0	32	3	28	54	13	0	37	38	2	83	18	0
52	3.8	12	0	105	8	42	121	13	0	4533	4330	246	7496	18	0	131	11	114	331	13	0	43	43	2	135	18	0
53	1.0	3	2	109	8	28	125	13	0	2916	3004	120	5161	18	0	32	3	28	126	13	0	26	29	1	132	18	0
54	0.4	2	2	106	7	16	124	13	0	2198	2340	62	4249	18	0	14	1	12	94	13	0	21	23	1	108	18	0
55	0.5	1	3	103	7	16	118	13	0	3482	3374	120	6298	18	0	18	1	16	124	13	0	34	34	1	142	18	0
56	6.1	16	9	106	8	35	118	13	0	3821	3689	300	6862	18	0	206	16	186	620	13	0	36	37	3	119	18	0
57	0.8	3	15	105	8	22	118	13	0	3363	3222	288	6150	18	0	26	2	23	125	13	0	33	33	3	135	18	0
58	4.0	12	28	103	9	19	117	13	0	3736	3607	318	6816	18	0	141	12	124	745	13	0	37	39	3	179	18	0
59	0.9	3	13	106	9	6	122	13	0	2940	2768	240	5480	18	0	32	3	28	592	13	0	29	29	2	155	18	0
60	3.4	11	19	101	12	5	126	13	0	-	-	-	-	30240	100	120	14	97	2436	13	0	-	-	-	-	30240	100
61	0.8	2	13	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
62	7.8	20	38	89	37	14	108	29225	97	-	-	-	-	30240	100	316	168	260	2080	29225	97	-	-	-	-	30240	100
63	1.0	3	4	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
64	4.7	11	17	92	4	27	104	25479	84	-	-	-	-	30240	100	185	9	164	625	25479	84	-	-	-	-	30240	100
65	0.6	2	8	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100	-	-	-	-	30240	100
66	1.8	4	11	102	8	12	124	34	0	-	-	-	-	30240	100	63	5	52	546	34	0	-	-	-	-	30240	100
67	4.5	11	13	102	7	20	118	21	0	3484	3184	60	6522	18	0	158	11	136	808	21	0	34	32	1	156	39	0
68	0.9	2	5	105	8	18	125	8694	29	-	-	-	-	30240	100	31	2	26	186	8694	29	-	-	-	-	30240	100
69	0.5	1	4	102	7	32	118	14	0	-	-	-	-	30240	100	16	1	14	51	14	0	-	-	-	-	30240	100
70	1.9	6	3	105	8	26	117	13	0	3801	3663	93	7455	19	0	66	5	59	280	13	0	36	38	1	127	19	0
71	9.5	27	11	105	9	29	119	13	0	3809	3674	304	7466	18	0	327	28	288	1164	13	0	36	37	3	129	18	0
72	0.9	3	20	100	8	7	115	253	1	2668	2434	60	5219	979	3	31	3	27	423	253	1	27	26	1	135	979	3
73	3.4	10	3	103	8	18	119	1436	5	3673	3444	60	6746	622	2	120	9	104	685	1436	5	37	35	3	144	1441	5
74	1.1	4	1	106	9	13	124	1146	4	2658	2488	60	5281	634	2	36	4	31	297	1146	4	26	25	1	109	1151	4
75	7.8	24	8	105	8	11	118	0	0	3213	3118	300	6006	18	0	266	19	236	2541	0	0	31	32	3	151	18	0
76	1.2	2	16	102	7	6	122	0	0	2171	2053	83	4194	18	0	42	3	34	699	0	0	22	21	2	180	18	0

Table B.4: Summary statistics of data, computed using data between 19th May 2017 and 8th June 2017. This subset is discussed in sections 4.3.3 and 4.5.2. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.5 Data between April 7th 2017 and June 20th 2017

In table B.5, we show summary statistics for data collected between April 7th 2017 and June 20th 2017. This data is very similar to the subset described in appendix B.4.1, just extended by 13 days. As a result, the same conclusions made in appendix B.4.1 hold here, and the additional 13 days of data show no additional patterns.

This dataset is used to compare atypical traffic durations with NTIS incident flags in section 4.3.4.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	17	100	5	13	112	268	0	2048	1958	60	4056	553	1	26	1	24	202	268	0	20	20	1	123	553	1
1	3.6	13	27	103	7	21	114	273	0	3518	3350	194	6863	336	0	128	9	115	623	273	0	34	35	2	158	336	0
2	1.0	3	8	108	7	29	123	178	0	2808	2642	60	5766	367	0	33	2	29	333	178	0	26	26	1	113	367	0
3	0.5	2	8	106	7	18	126	423	0	2328	2189	60	4980	555	1	17	1	15	101	423	0	22	22	1	117	651	1
4	0.4	1	6	106	7	18	126	329	0	2340	2202	60	4949	355	0	14	1	11	79	329	0	22	22	1	114	376	0
5	4.5	2	18	109	8	23	128	175	0	2980	2891	180	6360	394	0	149	11	127	1800	175	0	27	28	2	135	399	0
6	0.9	2	6	110	7	11	129	641	1	2364	2216	64	4786	490	0	28	1	24	289	641	1	21	21	1	118	705	1
7	5.1	2	21	111	7	11	126	16	0	2681	2565	168	5374	441	0	166	11	146	1683	16	0	24	24	2	166	441	0
8	1.0	1	22	96	8	5	130	30392	29	1740	1674	60	3476	269	0	40	3	29	753	30392	29	21	10	1	381	30492	29
9	0.7	1	22	101	6	5	128	409	0	3103	2850	60	7097	540	1	23	2	18	469	409	0	30	29	1	279	540	1
10	0.9	2	35	102	7	4	130	13801	13	3595	3338	180	7980	266	0	31	2	25	798	13801	13	37	29	2	390	13897	13
11	5.9	13	130	102	10	3	130	32203	30	-	-	-	-	106560	100	209	20	163	7081	32203	30	-	-	-	-	106560	100
12	1.0	2	35	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100
13	6.8	17	109	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100
14	1.0	2	9	106	8	27	122	23270	22	-	-	-	-	106560	100	34	3	29	131	23270	22	-	-	-	-	106560	100
15	1.2	4	16	107	7	18	121	283	0	3816	3639	144	7516	441	0	40	2	35	232	283	0	35	36	1	162	441	0
16	1.5	6	16	108	7	24	121	165	0	3744	3581	144	7375	399	0	51	3	45	227	165	0	34	34	2	149	399	0
17	0.5	0	7	110	7	20	126	173	0	3755	3600	60	7317	264	0	15	1	14	135	173	0	33	34	1	169	277	0
18	0.2	1	5	110	8	14	127	183	0	2547	2545	60	5687	410	0	6	0	5	61	183	0	23	24	1	117	410	0
19	1.2	4	8	110	7	13	127	183	0	2400	2429	60	5388	399	0	41	3	35	411	183	0	21	23	1	116	399	0
20	2.0	4	16	106	8	14	119	403	0	4261	3925	195	8251	359	0	69	5	61	537	403	0	40	40	2	200	488	0
21	3.0	6	13	105	9	7	123	402	0	3538	3304	180	7143	884	1	102	8	87	1465	402	0	33	34	2	165	884	1
22	10.0	22	54	112	7	23	126	273	0	4035	3744	275	8335	330	0	321	22	286	1580	273	0	36	36	2	123	330	0
23	0.5	1	23	113	8	18	130	321	0	3385	3169	60	7382	391	0	16	1	14	103	321	0	31	31	1	148	391	0
24	6.4	13	97	111	10	16	125	276	0	3900	3679	240	7854	338	0	208	20	185	1446	276	0	36	39	2	154	338	0
25	0.8	2	49	108	15	5	125	284	0	3675	3284	300	7411	69739	65	26	4	22	517	284	0	36	44	3	152	69739	65
26	2.6	9	51	107	16	4	123	283	0	3973	3735	69	7980	350	0	86	14	75	2166	283	0	37	43	1	161	350	0
27	0.9	2	38	108	15	5	127	710	1	2911	2837	60	6120	865	1	29	4	25	723	710	1	27	31	1	144	985	1
28	6.9	13	94	104	23	5	121	1264	1	4475	4356	60	8247	769	1	240	58	206	4929	1264	1	43	54	2	162	1321	1
29	0.9	3	23	105	23	13	130	925	1	3799	3817	60	7260	976	1	32	8	26	258	925	1	38	50	0	151	1121	1
30	2.5	4	15	99	17	18	122	280	0	4749	4659	170	8335	402	0	91	16	74	502	280	0	48	58	2	152	402	0
31	1.2	1	12	103	12	18	120	297	0	3185	3175	146	6240	373	0	42	5	36	235	297	0	33	35	1	117	373	0
32	3.9	6	76	106	10	18	120	277	0	4504	4668	300	8854	339	0	131	13	116	770	277	0	45	52	3	214	339	0
33	1.0	3	110	108	12	5	122	291	0	3600	4189	180	8302	354	0	33	4	29	683	291	0	41	51	2	275	354	0
34	1.9	3	107	105	16	11	119	291	0	4826	4831	180	10020	365	0	67	11	59	666	291	0	54	65	2	299	365	0
35	1.7	5	104	102	29	9	122	128	0	3585	4053	264	7806	175	0	60	21	50	659	128	0	48	63	3	243	175	0
36	1.6	3	45	99	37	24	114	123	0	5726	4903	199	9269	216	0	58	29	50	233	123	0	63	83	4	207	216	0
37	1.2	3	47	101	52	14	127	116	0	3352	3179	180	5664	185	0	44	37	35	338	116	0	35	68	2	139	185	0
38	7.1	17	45	98	30	20	118	105	0	5604	5005	123	7960	153	0	263	93	219	1264	105	0	60	72	2	164	153	0
39	0.7	2	13	100	13	20	117	423	0	4095	3663	87	6808	171	0	26	4	22	130	423	0	42	42	1	108	470	0
40	0.6	2	25	102	13	6	123	543	1	3286	3009	60	4934	590	1	23	3	19	389	543	1	33	33	1	146	733	1
41	1.9	5	19	101	15	8	120	734	1	4637	4363	60	7387	509	0	67	10	56	844	734	1	47	48	2	192	827	1

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	17	96	23	6	121	751	1	4548	4333	60	7560	550	1	50	13	39	823	751	1	48	54	1	262	930	1
43	4.6	11	18	99	12	6	120	635	1	4526	4288	60	7061	465	0	166	20	137	2656	635	1	47	47	1	201	790	1
44	0.8	1	12	103	10	8	121	312	0	4587	4326	60	7215	536	1	28	2	24	399	312	0	45	45	1	344	655	1
45	1.6	5	9	103	10	8	121	0	0	4704	4517	187	7814	220	0	57	6	48	811	0	0	46	47	2	184	220	0
46	0.2	1	8	103	10	5	121	121	0	4478	4317	60	7597	3496	3	9	1	7	176	121	0	44	45	1	198	3496	3
47	0.9	3	7	101	11	15	116	112	0	4962	4855	107	8551	177	0	32	3	28	224	112	0	50	52	2	177	177	0
48	1.4	4	8	102	10	22	116	107	0	4942	4817	114	8520	239	0	48	5	42	226	107	0	49	51	1	157	239	0
49	1.8	4	8	104	9	31	121	107	0	4958	4827	109	8726	242	0	64	5	55	213	107	0	48	50	1	141	242	0
50	2.3	7	5	105	8	19	126	118	0	4295	4173	103	7371	183	0	79	6	66	437	118	0	41	42	1	170	188	0
51	0.9	3	4	105	10	9	124	136	0	3773	3675	180	6235	221	0	31	3	27	360	136	0	36	38	2	141	221	0
52	3.8	12	11	106	8	14	122	105	0	4437	4348	198	8256	167	0	130	10	113	983	105	0	42	44	2	146	167	0
53	1.0	3	7	109	8	10	127	133	0	2874	3021	84	6105	192	0	32	3	28	347	133	0	26	29	1	138	192	0
54	0.4	2	7	105	7	3	124	189	0	2158	2357	62	5520	307	0	14	1	12	478	189	0	20	23	1	142	307	0
55	0.5	1	10	104	6	5	119	128	0	3412	3348	90	7560	282	0	18	1	16	379	128	0	33	34	1	156	282	0
56	6.1	16	29	107	8	9	120	105	0	3722	3686	87	7210	159	0	205	15	182	2422	105	0	35	37	2	151	159	0
57	0.8	3	39	106	8	5	122	106	0	3290	3231	108	6397	174	0	26	2	22	542	106	0	32	33	1	156	174	0
58	4.0	12	74	104	8	7	118	106	0	3686	3632	60	7054	173	0	139	10	123	2072	106	0	36	38	3	179	173	0
59	0.9	3	44	106	9	5	130	204	0	2864	2772	60	5778	310	0	32	3	26	682	204	0	28	29	0	209	310	0
60	3.4	11	64	101	12	5	129	204	0	-	-	-	-	106560	100	120	14	95	2436	204	0	-	-	-	-	106560	100
61	0.8	2	52	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100
62	7.8	20	144	80	38	14	111	104528	98	-	-	-	-	106560	100	351	182	253	2080	104528	98	-	-	-	-	106560	100
63	1.0	3	17	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100
64	4.7	11	50	91	7	6	110	90529	85	-	-	-	-	106560	100	187	14	155	2843	90529	85	-	-	-	-	106560	100
65	0.6	2	21	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100	-	-	-	-	106560	100
66	1.8	4	37	104	8	6	124	308	0	-	-	-	-	106560	100	62	4	52	1074	308	0	-	-	-	-	106560	100
67	4.5	11	51	102	7	12	118	110	0	3388	3186	60	6694	209	0	158	11	136	1328	110	0	33	32	1	160	261	0
68	0.9	2	16	105	8	8	125	32252	30	-	-	-	-	106560	100	31	2	26	406	32252	30	-	-	-	-	106560	100
69	0.5	1	16	102	7	16	119	123	0	-	-	-	-	106560	100	16	1	14	104	123	0	-	-	-	-	106560	100
70	1.9	6	12	105	8	11	121	111	0	3670	3636	93	7782	171	0	66	5	57	628	111	0	35	37	1	151	171	0
71	9.5	27	33	106	9	16	122	119	0	3671	3626	240	7655	169	0	324	27	281	2173	119	0	34	36	2	141	169	0
72	0.9	3	33	102	8	7	118	791	1	2520	2431	60	5315	1676	2	31	3	26	423	791	1	24	25	1	135	1916	2
73	3.4	10	15	103	8	18	119	2858	3	3515	3468	60	6797	1112	1	120	9	104	685	2858	3	34	35	1	144	2944	3
74	1.1	4	12	106	9	10	124	4015	4	2655	2437	60	5281	4098	4	36	3	31	399	4015	4	25	24	1	109	5055	5
75	7.8	24	31	105	7	11	118	4244	4	3192	3063	60	6113	1502	1	266	19	236	2541	4244	4	31	30	1	151	4305	4
76	1.2	2	45	102	7	3	122	3265	3	2172	1991	60	4410	3255	3	42	3	34	1376	3265	3	22	20	1	234	3577	3

Table B.5: Summary statistics of data, computed using data between April 7th 2017 and June 20th 2017. This subset is discussed in section 4.3.4 and is used to compare the distribution of NTIS events to atypical traffic periods. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.6 Data between September 1st 2017 and November 24th 2017

In table B.6, we show summary statistics for data collected between September 1st 2017 and November 24th 2017. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. Regarding missing data, link 63 reports only missing values in this subset, where as in the entire data-set it had a small fraction of non-missing values. Regarding central tendencies, the largest differences in median speed are observed on links 42, 64 and 76. The majority of differences in median speeds are below 2 km/hr, showing little difference in an absolute sense. Median flow appears to be systematically lower across most links in table B.6 compared to table 3.1, suggesting some factor impacting the network wide behaviour. However, the majority of median flows in table B.6 are within 200 veh/hr of the values in table 3.1.

Inspecting central tendencies, links 33, 34, 35 and 37 show lower speed inter-quartile ranges in table B.6 than table 3.1. The fact these are nearby in space clearly suggest a physical factor that propagated across the network was responsible. Similarly to median flow values, inter-quartile ranges of flows are generally lower in this subset of data compared to the entire dataset.

This dataset is used in calibrate incident detection methodologies in section 4.6.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	17	100	6	0	116	32	0	2040	1973	0	4260	287	0	26	1	23	438	12333	10	21	21	0	140	293	0
1	3.6	13	23	101	7	16	115	32	0	3600	3447	0	6840	37	0	130	9	114	818	3152	3	36	36	0	156	37	0
2	1.0	3	16	106	6	11	126	0	0	2864	2750	0	5685	37	0	34	2	29	327	3120	3	27	27	0	122	37	0
3	0.5	2	2	105	7	7	124	0	0	2355	2290	0	4920	37	0	18	1	15	269	4627	4	23	23	0	121	37	0
4	0.4	1	3	105	7	7	124	0	0	2351	2269	0	4860	7823	6	14	1	12	210	3120	3	23	23	0	114	7823	6
5	4.5	2	6	108	7	11	127	0	0	3000	2988	120	6420	37	0	151	11	132	1481	8923	7	28	29	1	165	37	0
6	0.9	2	6	109	6	5	126	32	0	2377	2287	60	5038	37	0	29	2	25	624	8955	7	22	22	1	171	37	0
7	5.1	2	11	109	6	14	125	0	0	2700	2628	120	5472	37	0	168	10	147	1608	3662	3	25	24	1	193	37	0
8	1.0	1	18	89	4	0	105	484	0	1740	1638	0	4080	37	0	42	2	36	753	4146	3	20	19	1	164	1284	1
9	0.7	1	17	100	6	5	122	0	0	3168	2893	120	7098	37	0	23	1	19	469	3120	3	31	30	1	340	37	0
10	0.9	2	25	103	8	5	127	34324	28	3660	3420	180	7680	19	0	31	3	25	638	34333	28	41	17	5	390	34343	28
11	5.9	13	59	99	11	5	130	34312	28	-	-	-	-	120960	100	214	24	163	4248	34333	28	-	-	-	-	120960	100
12	1.0	2	29	-	-	-	-	120960	100	-	-	-	-	120960	100	-	-	-	-	120960	100	-	-	-	-	120960	100
13	6.8	17	61	-	-	-	-	120960	100	-	-	-	-	120960	100	-	-	-	-	120960	100	-	-	-	-	120960	100
14	1.0	2	7	105	10	8	118	32	0	-	-	-	-	120960	100	34	3	30	448	8955	7	-	-	-	-	120960	100
15	1.2	4	8	106	9	19	119	0	0	3780	3676	180	7380	37	0	40	4	36	225	4627	4	35	37	2	159	37	0
16	1.5	6	10	107	7	20	119	0	0	3727	3660	180	7260	37	0	51	4	46	275	4627	4	35	36	2	124	37	0
17	0.5	0	7	109	7	7	124	0	0	3720	3660	60	7256	20	0	16	1	14	242	3120	3	34	35	1	357	20	0
18	0.2	1	4	109	8	6	125	32	0	2520	2585	0	5640	37	0	6	0	5	110	12333	10	23	24	0	280	37	0
19	1.2	4	6	108	7	18	124	32	0	2364	2449	0	5328	37	0	41	2	36	246	12333	10	22	23	0	120	37	0
20	2.0	4	7	104	8	23	117	32	0	4191	3840	180	8100	37	0	71	6	62	314	9795	8	40	40	2	191	37	0
21	3.0	6	6	103	8	9	122	32	0	3540	3289	180	7080	37	0	104	9	88	1194	9795	8	34	34	2	243	37	0
22	10.0	22	16	111	8	29	126	32	0	4080	3819	180	8220	37	0	325	22	289	1243	9795	8	37	36	2	166	37	0
23	0.5	1	3	111	9	16	130	32	0	3333	3123	138	6623	304	0	17	1	14	116	9795	8	30	31	1	121	304	0
24	6.4	13	39	109	10	17	125	32	0	3929	3667	180	7500	37	0	211	18	184	1355	12292	10	37	38	2	140	37	0
25	0.8	2	41	108	15	11	124	81	0	3639	3461	0	7500	37	0	26	3	23	255	12341	10	35	39	0	198	86	0
26	2.6	9	39	105	17	0	123	44	0	4001	3792	0	7620	37	0	88	13	74	650	12355	10	39	43	0	174	100	0
27	0.9	2	42	106	18	0	125	32	0	2934	2851	0	5880	37	0	29	5	25	312	2946	2	30	33	0	148	3031	3
28	6.9	13	57	103	33	14	122	32	0	4550	4419	60	8068	37	0	244	112	204	1852	8955	7	46	65	1	165	37	0
29	0.9	3	12	100	26	13	122	32	0	3900	3875	0	6711	37	0	34	11	27	258	8955	7	42	55	0	134	37	0
30	2.5	4	7	97	18	0	119	0	0	4994	4761	0	8100	37	0	93	18	76	401	3120	3	51	58	0	232	253	0
31	1.2	1	16	101	11	16	119	36764	30	3180	3004	0	5820	34830	29	43	5	36	270	39801	33	34	32	0	101	36764	30
32	3.9	6	53	105	11	0	120	111	0	4680	4649	0	9300	37	0	133	13	116	819	12276	10	49	49	0	197	3871	3
33	1.0	3	113	106	14	0	121	32	0	3708	4140	0	8520	37	0	34	5	29	510	9795	8	44	50	0	190	897	1
34	1.9	3	105	103	24	0	118	32	0	4740	4802	0	10068	37	0	69	33	60	858	12288	10	58	73	2	281	59	0
35	1.7	5	156	101	44	0	119	32	0	3443	4000	0	7980	37	0	61	59	52	1022	12292	10	50	75	0	213	40	0
36	1.6	3	68	97	42	18	114	32	0	5520	4920	240	9480	37	0	60	46	50	311	12288	10	67	93	2	241	37	0
37	1.2	3	111	98	57	6	122	32	0	3160	3199	120	5700	37	0	46	49	36	711	7874	7	39	73	1	141	37	0
38	7.1	17	97	96	30	10	117	32	0	5356	5060	300	7909	37	0	269	98	221	2573	4659	4	63	74	3	225	37	0
39	0.7	2	38	97	17	0	117	32	0	3926	3690	0	6050	37	0	27	5	22	175	4659	4	44	44	0	135	239	0
40	0.6	2	54	100	17	0	126	1205	1	3180	3073	0	5100	40	0	23	4	19	389	5656	5	35	36	0	194	2208	2
41	1.9	5	41	95	18	0	117	40	0	4569	4519	0	7425	40	0	71	12	58	832	7882	7	52	53	0	180	4461	4

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	20	82	17	0	122	216	0	4504	4502	0	7320	40	0	58	11	39	795	7907	7	58	60	0	309	4279	4
43	4.6	11	28	97	13	0	121	173	0	4458	4444	0	7140	37	0	170	20	137	823	7860	6	49	48	0	180	4219	3
44	0.8	1	12	101	11	0	121	0	0	4520	4500	0	7380	38	0	28	3	24	205	7842	6	46	48	0	336	831	1
45	1.6	5	12	101	11	14	121	0	0	4730	4699	0	7980	37	0	58	6	48	418	7842	6	47	50	0	189	37	0
46	0.2	1	11	103	11	13	122	32	0	4500	4507	0	7800	47	0	9	1	7	68	7874	7	45	48	0	182	47	0
47	0.9	3	7	101	11	16	121	32	0	5040	5027	99	8684	37	0	32	4	27	203	4659	4	50	53	1	189	37	0
48	1.4	4	5	102	9	23	118	0	0	5040	5022	120	8557	37	0	48	4	42	214	4627	4	50	53	1	162	37	0
49	1.8	4	3	103	7	30	120	0	0	5068	5040	84	8613	37	0	65	3	55	221	7842	6	50	52	1	174	37	0
50	2.3	7	3	104	8	16	125	0	0	4314	4289	120	7380	37	0	79	6	66	1200	7842	6	42	43	1	237	37	0
51	0.9	3	4	101	10	13	123	32	0	3780	3736	60	6480	37	0	33	3	27	369	7874	7	38	40	1	147	37	0
52	3.8	12	17	105	8	14	124	32	0	4500	4440	120	7440	37	0	132	10	111	968	3694	3	44	45	1	231	37	0
53	1.0	3	20	108	7	15	127	32	0	2915	3107	120	5508	37	0	33	2	27	240	4659	4	27	30	1	158	37	0
54	0.4	2	24	105	7	8	130	32	0	2204	2439	0	4680	37	0	14	1	11	187	3694	3	21	25	0	161	37	0
55	0.5	1	27	101	7	6	119	104	0	3480	3428	0	6780	37	0	19	1	16	316	4659	4	35	36	0	197	109	0
56	6.1	16	30	105	8	0	119	32	0	3840	3780	0	7433	37	0	211	16	185	3128	7874	7	38	39	0	152	383	0
57	0.8	3	40	103	6	10	124	32	0	3405	3350	0	6713	37	0	26	1	22	271	7874	7	34	35	0	171	37	0
58	4.0	12	62	101	8	7	121	32	0	3821	3780	0	6960	37	0	144	13	120	2072	7874	7	38	40	0	162	37	0
59	0.9	3	46	104	10	5	130	342	0	3020	2899	0	5700	37	0	33	3	26	682	8184	7	30	30	0	145	347	0
60	3.4	11	47	100	11	0	130	32	0	-	-	-	-	120960	100	122	13	96	2436	8964	7	-	-	-	-	120960	100
61	0.8	2	35	108	11	0	122	109122	90	-	-	-	-	120960	100	28	2	25	148	111496	92	-	-	-	-	120960	100
62	7.8	20	67	104	16	0	126	109590	91	3960	3344	360	6895	110016	91	267	32	223	1477	110466	91	42	17	4	204	112992	93
63	1.0	3	15	-	-	-	-	120960	100	-	-	-	-	120960	100	-	-	-	-	120960	100	-	-	-	-	120960	100
64	4.7	11	24	93	5	42	109	106620	88	-	-	-	-	120960	100	184	10	158	405	106719	88	-	-	-	-	120960	100
65	0.6	2	10	-	-	-	-	120960	100	-	-	-	-	120960	100	-	-	-	-	120960	100	-	-	-	-	120960	100
66	1.8	4	20	102	8	6	126	42	0	-	-	-	-	120960	100	63	5	51	1074	4721	4	-	-	-	-	120960	100
67	4.5	11	35	101	6	13	117	1	0	3480	3269	0	6530	37	0	159	10	138	1239	8731	7	34	33	0	168	38	0
68	0.9	2	26	105	8	6	124	34356	28	-	-	-	-	120960	100	31	2	26	541	34908	29	-	-	-	-	120960	100
69	0.5	1	21	101	6	15	121	32	0	-	-	-	-	120960	100	16	1	13	113	8763	7	-	-	-	-	120960	100
70	1.9	6	19	103	7	0	119	32	0	3686	3617	0	7685	37	0	67	4	58	691	4711	4	36	37	0	160	104	0
71	9.5	27	31	105	8	0	119	32	0	3669	3600	60	7620	37	0	330	25	289	1716	8707	7	36	36	1	143	104	0
72	0.9	3	14	101	7	0	120	32	0	2472	2409	0	5310	37	0	31	3	26	309	8898	7	25	25	0	124	297	0
73	3.4	10	13	104	7	0	119	32	0	3360	3360	87	6720	37	0	119	8	105	724	8707	7	33	34	1	140	104	0
74	1.1	4	14	106	8	0	124	32	0	2515	2539	0	5100	37	0	36	3	31	482	8707	7	24	25	0	114	191	0
75	7.8	24	35	105	7	7	121	0	0	3168	3090	0	6291	37	0	268	18	231	3993	8731	7	32	32	0	196	37	0
76	1.2	2	66	103	7	0	126	0	0	2122	2100	0	4440	37	0	41	3	33	699	8731	7	22	21	0	189	103	0

Table B.6: Summary statistics of data, computed using data between September 1st 2017 and November 24th 2017. This subset is discussed in section 4.6 and is used to calibrate incident detection methodologies. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.7 Data between December 21st 2017 and November 1st 2018

In table B.7, we show summary statistics for data collected between December 21st 2017 and November 1st 2018. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. First, this dataset starts near Christmas, where one would clearly expect cold temperatures in the UK, and potentially lower commuter traffic. However, since it spans such a long time-period, we see this as simply a representative aspect of the data generation process, and when testing our methodology for detecting events, should be included in the validation. Secondly, if we consider missing data, we see link 31 reported data for part of the entire collection period, however is entirely missing in this subset of the data.

If we compare median speed values in table B.7 to those attained from the entire dataset in table 3.1, we see no systematic increase or decrease. The largest discrepancies are link 76 (2 km/hr lower median speed) and link 11 (1 km/hr higher median speed) however these values are small enough to consider them simply noise in the data. Median flow and density values similarly show little difference from the entire subset. Considering the inter-quartile range of speed in the entire dataset and this subset, we see the largest discrepancies occur on links 28 and 48, both of which have a larger inter-quartile range by 2 km/hr. Again, this difference is small enough to be considered negligible. The same patterns hold when inspecting the inter-quartile ranges of flow and density in the full dataset and subset considered here.

This dataset is used as unseen test data in the validation of our proposed incident detection methodology in section 4.6.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	51	100	6	0	127	3074	1	1926	1951	0	4298	33806	7	26	1	22	233	43120	10	19	20	0	133	34374	8
1	3.6	13	71	102	7	0	117	2885	1	3362	3380	0	7131	2923	1	129	9	112	935	5600	1	33	35	1	176	3616	1
2	1.0	3	46	106	6	11	126	3001	1	2662	2714	0	5943	2940	1	34	2	29	327	5634	1	25	27	0	130	3022	1
3	0.5	2	34	105	8	13	129	3184	1	2246	2280	0	5213	2988	1	18	2	14	142	11294	2	21	23	0	142	3218	1
4	0.4	1	31	105	8	13	129	3184	1	2245	2280	0	5213	8580	2	14	1	11	111	5669	1	21	23	0	140	8854	2
5	4.5	2	35	108	8	0	129	2908	1	2870	2942	0	6540	2986	1	152	12	126	1727	40351	9	26	29	0	141	3022	1
6	0.9	2	16	108	7	5	130	3087	1	2267	2308	0	5160	2960	1	29	2	24	624	40335	9	21	22	0	141	3155	1
7	5.1	2	62	110	8	0	128	2864	1	2575	2643	0	5760	3036	1	167	13	145	2301	7291	2	23	25	0	244	3440	1
8	1.0	1	106	90	5	0	116	3322	1	1669	1626	0	4358	3013	1	42	2	32	692	7168	2	19	19	0	230	4842	1
9	0.7	1	73	101	6	0	118	2892	1	3060	2896	120	7020	291067	64	23	1	20	344	6245	1	31	31	1	248	291067	64
10	0.9	2	97	103	7	6	121	11768	3	3242	3187	120	8160	33821	7	31	2	26	527	20108	4	32	33	2	217	33821	7
11	5.9	13	314	104	8	7	121	17225	4	3300	3180	60	7800	33813	7	204	17	176	2946	25565	6	32	32	1	238	33813	7
12	1.0	2	116	108	8	14	126	61197	13	3092	2985	0	7680	61299	14	34	3	29	263	69494	15	29	30	0	168	61299	14
13	6.8	17	275	105	9	8	122	33761	7	3303	3191	84	6900	61181	13	234	23	199	3041	71135	16	32	33	1	309	61181	13
14	1.0	2	47	105	11	0	122	2998	1	3060	2760	0	6120	135625	30	34	4	29	256	40345	9	29	29	0	243	135728	30
15	1.2	4	38	105	11	0	122	2864	1	3583	3683	0	7511	2985	1	41	5	35	377	11206	2	34	37	0	157	3147	1
16	1.5	6	34	105	8	0	122	2864	1	3531	3614	0	7380	2978	1	52	4	45	486	11206	2	33	36	0	184	3140	1
17	0.5	0	31	108	8	0	127	2901	1	3540	3620	0	7260	3780	1	16	1	13	179	4946	1	32	35	0	273	3948	1
18	0.2	1	32	109	9	8	130	2997	1	2406	2612	0	5820	2995	1	6	0	5	80	43141	10	22	25	0	148	3101	1
19	1.2	4	55	108	8	6	127	3037	1	2262	2479	0	5580	4872	1	41	3	35	739	43190	10	21	24	0	207	4974	1
20	2.0	4	38	103	8	8	120	2893	1	3949	3836	0	8080	2934	1	71	5	60	636	41948	9	38	40	0	243	2936	1
21	3.0	6	65	103	10	6	125	2893	1	3335	3240	0	7070	2984	1	104	11	86	1756	41948	9	32	33	0	243	2986	1
22	10.0	22	104	110	8	0	126	2891	1	3837	3720	0	8100	2934	1	329	24	286	2762	41948	9	35	36	1	292	3745	1
23	0.5	1	76	109	8	8	130	2891	1	3135	3042	0	7012	13422	3	17	2	14	272	41948	9	29	30	0	176	13422	3
24	6.4	13	275	108	11	0	125	2891	1	3720	3634	0	7723	2939	1	213	23	184	3291	43061	9	35	39	1	200	4670	1
25	0.8	2	198	108	13	0	127	2891	1	3405	3375	0	7715	25668	6	26	4	22	255	43061	9	33	39	0	206	26685	6
26	2.6	9	171	106	15	0	126	2891	1	3772	3704	0	7562	2934	1	87	14	74	614	43061	9	36	43	0	198	3951	1
27	0.9	2	193	106	16	0	126	2891	1	2828	2722	0	6262	2934	1	29	5	25	312	2920	1	28	34	0	190	2973	1
28	6.9	13	265	104	33	11	124	2891	1	4380	4260	120	8160	2933	1	242	117	201	2205	40334	9	43	66	1	207	2933	1
29	0.9	3	60	101	25	6	130	3093	1	3872	3720	0	7020	2934	1	34	10	26	559	40334	9	39	52	0	152	3136	1
30	2.5	4	41	101	19	0	122	2864	1	4844	4654	0	8241	2960	1	89	17	74	564	4754	1	48	56	0	188	4871	1
31	1.2	1	0	-	-	-	-	453600	100	-	-	-	-	453600	100	-	-	-	-	453600	100	-	-	-	-	453600	100
32	3.9	6	174	105	10	0	122	2891	1	4680	4441	0	9600	2933	1	134	14	114	1468	41979	9	47	49	2	221	5270	1
33	1.0	3	390	106	13	0	123	2911	1	3756	3907	0	8700	2941	1	34	4	28	546	41166	9	42	48	0	334	3828	1
34	1.9	3	348	103	21	11	123	2910	1	4759	4609	120	10200	2965	1	69	23	57	760	41997	9	53	68	1	309	2965	1
35	1.7	5	499	101	39	6	124	2958	1	3482	3787	0	8138	2927	1	61	49	49	1022	42050	9	46	69	0	275	2981	1
36	1.6	3	156	95	40	19	116	3034	1	5423	4666	0	9926	3343	1	60	41	49	346	42121	9	63	89	0	360	3343	1
37	1.2	3	220	98	51	10	122	3015	1	3152	3064	0	5860	3064	1	46	42	36	443	27642	6	36	67	0	197	3064	1
38	7.1	17	227	97	25	10	120	3015	1	5227	4935	0	7977	3057	1	266	75	214	2573	10700	2	57	67	0	226	3057	1
39	0.7	2	87	98	13	0	130	3089	1	3829	3649	0	6480	3104	1	26	3	20	349	10774	2	40	42	0	155	3247	1
40	0.6	2	114	100	13	0	130	4430	1	2940	3000	0	5100	41485	9	23	3	18	467	12029	3	30	34	0	158	42901	9
41	1.9	5	113	98	15	0	117	3120	1	4298	4380	0	7560	3115	1	70	10	58	1355	16439	4	46	51	0	216	6302	1

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	73	89	18	0	126	3241	1	4217	4378	0	7560	3069	1	54	11	38	954	16495	4	49	56	0	382	6022	1
43	4.6	11	96	97	12	0	128	3041	1	4213	4353	0	7560	3061	1	171	20	129	2497	28221	6	45	49	0	229	6187	1
44	0.8	1	43	101	9	0	121	3019	1	4267	4291	0	7200	39264	9	29	3	24	573	27927	6	43	46	0	344	39336	9
45	1.6	5	46	100	9	0	119	2996	1	4380	4547	0	8008	3064	1	59	5	49	1171	27860	6	44	48	0	194	3374	1
46	0.2	1	45	100	11	0	130	3017	1	4200	4359	0	7819	4046	1	9	1	7	134	27644	6	43	47	0	202	4105	1
47	0.9	3	37	100	10	0	130	3017	1	4661	4885	0	8820	3115	1	33	3	25	502	11433	3	47	52	0	354	3201	1
48	1.4	4	46	100	9	0	126	2991	1	4643	4882	0	8640	3062	1	49	4	39	704	11407	3	47	51	0	245	3125	1
49	1.8	4	40	101	9	0	129	2991	1	4641	4861	0	8798	3293	1	66	5	54	1107	27618	6	46	50	0	213	3356	1
50	2.3	7	42	102	9	0	129	3012	1	3995	4194	0	8620	3061	1	80	8	64	1036	22812	5	39	42	0	237	6148	1
51	0.9	3	26	100	10	0	126	3286	1	3500	3674	0	7260	15762	3	33	3	26	661	22260	5	36	38	0	174	25236	6
52	3.8	12	35	104	8	0	124	3017	1	4140	4385	0	7560	3061	1	132	10	111	1667	10178	2	40	44	0	306	5855	1
53	1.0	3	24	108	6	0	130	3610	1	2662	3060	0	5580	3088	1	33	2	27	588	11212	2	25	30	0	138	5383	1
54	0.4	2	31	105	7	0	130	3046	1	2003	2426	0	4680	3729	1	14	1	11	249	10629	2	19	24	0	131	4916	1
55	0.5	1	40	103	6	0	124	3642	1	3240	3424	0	6660	3539	1	18	1	15	316	11982	3	32	35	0	236	5477	1
56	6.1	16	81	105	8	0	122	3015	1	3548	3773	0	7320	3057	1	210	16	180	2737	26745	6	35	38	0	289	7491	2
57	0.8	3	158	103	6	0	129	3018	1	3148	3280	0	6441	3111	1	26	1	21	362	23059	5	31	34	0	172	3375	1
58	4.0	12	276	102	9	0	119	3015	1	3487	3620	0	7004	3057	1	142	12	122	1450	27642	6	35	39	0	160	5322	1
59	0.9	3	149	103	9	0	130	9755	2	2831	2880	0	5880	3116	1	33	3	26	682	32712	7	29	31	0	193	10698	2
60	3.4	11	232	100	14	0	130	3771	1	-	-	-	-	453600	100	122	17	95	2436	40419	9	-	-	-	-	453600	100
61	0.8	2	153	106	10	0	130	3290	1	3540	3271	0	6600	209719	46	29	3	23	606	40492	9	35	31	0	225	217215	48
62	7.8	20	422	101	12	0	129	95771	21	3628	3552	0	7440	3159	1	277	30	218	5611	120573	27	42	29	0	397	105968	23
63	1.0	3	76	92	10	0	130	281340	62	3120	2895	0	5820	210052	46	40	4	28	732	285048	63	40	18	0	340	283206	62
64	4.7	11	216	98	11	0	130	246689	54	3584	3325	0	6720	338999	75	174	19	131	2843	264580	58	37	35	0	259	339749	75
65	0.6	2	76	-	-	-	-	451453	100	-	-	-	-	453600	100	-	-	-	-	453600	100	-	-	-	-	453600	100
66	1.8	4	100	101	10	0	126	3101	1	-	-	-	-	453600	100	64	6	51	1289	13517	3	-	-	-	-	453600	100
67	4.5	11	173	100	7	0	123	3063	1	3338	3300	0	6968	3058	1	161	10	136	1789	38240	8	33	34	0	168	5031	1
68	0.9	2	83	105	10	0	130	128571	28	-	-	-	-	453600	100	31	3	25	649	134802	30	-	-	-	-	453600	100
69	0.5	1	53	101	7	0	122	3043	1	-	-	-	-	453600	100	16	1	13	329	38442	8	-	-	-	-	453600	100
70	1.9	6	55	105	8	0	122	3018	1	3431	3620	0	7560	3116	1	66	5	57	1152	11471	3	33	37	0	300	3173	1
71	9.5	27	121	104	8	0	121	3017	1	3428	3633	0	7500	3060	1	330	27	284	4289	36560	8	33	37	0	154	3122	1
72	0.9	3	36	101	8	0	119	3017	1	2297	2460	0	5340	3072	1	31	3	26	494	39630	9	23	26	0	147	3503	1
73	3.4	10	42	104	7	10	122	3015	1	3183	3380	30	6840	3060	1	120	8	102	1210	36553	8	31	34	1	158	3060	1
74	1.1	4	38	106	8	0	130	3068	1	2429	2549	0	5460	3058	1	36	3	30	772	36606	8	23	25	0	113	3819	1
75	7.8	24	156	104	7	0	127	3032	1	3060	3122	0	6088	3061	1	270	18	220	3851	39472	9	30	32	0	167	3499	1
76	1.2	2	256	98	8	0	126	3339	1	1995	2040	0	4646	3313	1	43	3	33	699	38553	8	21	23	0	200	5391	1

Table B.7: Summary statistics of data, computed using data between December 21st 2017 and November 1st 2018. This subset is discussed in section 4.6 and is used to validate incident detection methodologies. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.8 Data between September 1st 2017 and September 31st 2018

In table B.8, we show summary statistics for the subset of data between September 1st 2017 and September 31st 2018. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. If we first look to missing data, we see that link 31 has roughly 10% more data-points missing in this subset compared to the entire dataset. There are no links that report only missing values in this subset that do not do the same in the entire dataset.

Considering median speed values, the largest discrepancy is observed on link 42, with a difference of 3 km/hr compared to the main dataset. All other links have median speed values in this subset within 1 km/hr of their values in the entire dataset. Median flow values are marginally higher for the first 40 links in the dataset, with median flow values in table B.8 being up to 60 veh/hr higher than those in table 3.1. In both absolute and relative terms however, this difference is negligible.

Comparing variation in the data, we see that the largest change in inter-quartile range of speeds is observed on link 35, with a difference of 1.7 km/hr. All other values are at most different by 1 km/hr. Again, in both absolute and relative terms this difference is negligible. The largest change in inter-quartile range of flow is on link 31, with a difference of just under 100 veh/hr.

This dataset is used to fit point process models to incident data in chapter 5.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	65	100	6	0	127	3106	1	1955	1935	0	4298	66698	12	26	1	22	438	48363	9	20	20	0	140	67272	12
1	3.6	13	90	101	7	0	117	2917	1	3405	3372	0	7131	2960	1	130	9	112	935	6891	1	34	35	0	176	3653	1
2	1.0	3	61	106	6	11	126	3001	1	2692	2700	0	5943	2977	1	34	2	29	327	6893	1	25	27	0	130	3059	1
3	0.5	2	38	105	8	7	129	3184	1	2254	2279	0	5213	3025	1	18	2	14	269	13422	2	22	23	0	142	3255	1
4	0.4	1	36	105	8	7	129	3184	1	2260	2276	0	5213	16403	3	14	1	11	210	6928	1	22	22	0	158	16677	3
5	4.5	2	39	108	8	0	129	2908	1	2880	2940	0	6540	3023	1	152	11	126	1727	42950	8	27	29	0	165	3059	1
6	0.9	2	23	108	7	5	130	3119	1	2280	2285	0	5160	2997	1	29	2	24	624	42966	8	21	22	0	171	3192	1
7	5.1	2	75	110	8	0	128	2864	1	2580	2610	0	5760	3073	1	167	12	145	2301	9092	2	23	24	0	244	3477	1
8	1.0	1	117	90	5	0	116	3806	1	1680	1613	0	4358	3050	1	42	2	32	753	9453	2	19	19	0	230	6126	1
9	0.7	1	86	100	6	0	122	2892	1	3095	2862	120	7098	291104	51	23	1	19	469	7504	1	31	30	1	340	291104	51
10	0.9	2	122	103	7	5	127	57115	10	3326	3253	120	8160	61537	11	31	2	25	638	62965	11	35	32	2	390	99014	17
11	5.9	13	357	103	11	5	130	69518	12	3272	3180	60	7800	193629	34	206	21	163	4248	75380	13	31	32	1	238	193629	34
12	1.0	2	137	108	8	14	126	221038	39	3072	2970	0	7680	221140	39	34	3	29	263	226836	40	29	29	0	168	221140	39
13	6.8	17	321	104	9	8	122	193602	34	3300	3176	84	6900	221022	39	234	23	199	3041	224026	39	32	32	1	309	221022	39
14	1.0	2	56	105	10	0	122	3030	1	3060	2750	0	6120	295453	52	34	4	29	448	42976	8	29	29	0	243	295556	52
15	1.2	4	46	106	9	0	122	2864	1	3600	3644	0	7511	3022	1	40	4	35	377	13334	2	34	37	0	159	3184	1
16	1.5	6	43	106	8	0	122	2864	1	3540	3600	0	7380	3015	1	52	4	45	486	13334	2	33	36	0	184	3177	1
17	0.5	0	39	108	8	0	127	2901	1	3540	3600	0	7260	3786	1	16	1	13	242	6205	1	32	34	0	357	3954	1
18	0.2	1	34	109	8	6	129	3029	1	2414	2580	0	5820	3032	1	6	0	5	110	48384	9	22	24	0	280	3138	1
19	1.2	4	59	108	8	6	127	3069	1	2272	2460	0	5580	4909	1	41	3	35	739	48433	9	21	24	0	207	5011	1
20	2.0	4	45	103	10	8	120	2925	1	3964	3836	0	8100	2971	1	71	6	60	636	45419	8	38	39	0	243	2973	1
21	3.0	6	67	103	9	6	125	2925	1	3360	3240	0	7080	3021	1	104	10	86	1756	45419	8	32	33	0	243	3023	1
22	10.0	22	118	110	8	0	126	2923	1	3840	3720	0	8220	2971	1	328	24	286	2762	45419	8	35	36	1	292	3782	1
23	0.5	1	78	109	9	8	130	2923	1	3156	3053	0	7012	13641	2	17	2	14	272	45419	8	29	30	0	176	13641	2
24	6.4	13	294	109	11	0	125	2923	1	3732	3625	0	7723	2976	1	213	22	184	3291	48263	8	35	38	1	200	3921	1
25	0.8	2	236	108	12	0	127	2972	1	3426	3393	0	7715	3108	1	26	4	22	255	48312	8	33	39	0	206	3775	1
26	2.6	9	219	106	15	0	126	2935	1	3780	3709	0	7620	2971	1	87	13	74	650	48326	8	36	43	0	198	3652	1
27	0.9	2	233	106	16	0	126	2929	1	2826	2747	0	6262	2971	1	29	5	25	520	5872	1	28	33	0	190	6010	1
28	6.9	13	315	104	32	11	124	2923	1	4380	4273	0	8160	2970	1	242	109	201	2266	42965	8	43	65	0	207	2970	1
29	0.9	3	59	101	24	6	130	2923	1	3846	3730	0	7020	2971	1	34	10	26	559	42965	8	39	52	0	152	2971	1
30	2.5	4	46	100	18	0	122	2864	1	4833	4682	0	8241	2997	1	90	16	74	564	6013	1	48	57	0	232	3213	1
31	1.2	1	16	101	11	16	119	484603	85	3180	3004	0	5820	482669	85	43	5	36	270	487640	86	34	32	0	101	484603	85
32	3.9	6	228	105	10	0	125	3002	1	4633	4488	0	9600	2970	1	133	13	112	1266	47931	8	46	49	0	221	7230	1
33	1.0	3	525	106	13	0	123	2943	1	3719	3953	0	8700	2978	1	34	4	28	546	44637	8	42	48	0	334	4725	1
34	1.9	3	489	103	21	0	123	2942	1	4713	4663	0	10200	3002	1	69	23	57	858	47961	8	53	68	2	309	3024	1
35	1.7	5	699	101	39	0	124	2990	1	3461	3816	0	8138	2964	1	61	47	49	1022	48018	8	46	69	0	275	3021	1
36	1.6	3	234	96	40	18	116	3066	1	5418	4731	0	9926	3206	1	60	41	49	346	48085	8	63	90	0	264	3206	1
37	1.2	3	353	98	51	6	122	3047	1	3120	3088	0	5860	3101	1	46	42	36	739	30044	5	36	67	0	197	3101	1
38	7.1	17	341	97	26	6	120	3047	1	5209	4954	0	7977	3094	1	266	79	214	4288	12860	2	57	68	0	226	3094	1
39	0.7	2	134	98	15	0	130	3121	1	3813	3643	0	6480	3141	1	26	4	20	516	12934	2	41	42	0	155	3486	1
40	0.6	2	175	100	14	0	130	4800	1	2940	3011	0	5100	41525	7	23	3	18	467	14351	3	31	34	0	194	44274	8
41	1.9	5	155	97	16	0	117	3160	1	4302	4382	0	7560	3155	1	70	11	58	1355	20284	4	47	51	0	216	9168	2

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	92	85	18	0	126	3456	1	4230	4371	0	7560	3109	1	56	11	38	954	20364	4	51	58	0	382	8705	2
43	4.6	11	127	97	12	0	128	3214	1	4219	4354	0	7560	3098	1	171	21	129	2621	30609	5	45	49	0	229	8811	2
44	0.8	1	55	101	9	0	121	3019	1	4296	4336	0	7380	52857	9	29	3	24	573	30297	5	44	46	0	344	53722	9
45	1.6	5	57	101	9	0	121	2996	1	4407	4560	0	8008	3101	1	59	5	48	1171	30230	5	44	49	0	194	3348	1
46	0.2	1	55	101	9	7	130	3049	1	4215	4380	0	7819	4285	1	9	1	7	134	30046	5	43	47	0	202	4285	1
47	0.9	3	43	100	11	7	130	3049	1	4682	4905	0	8820	3152	1	33	3	25	502	13517	2	47	52	0	354	3152	1
48	1.4	4	48	100	10	7	126	2991	1	4680	4890	0	8640	3099	1	49	4	39	704	13459	2	47	52	0	245	3099	1
49	1.8	4	41	101	8	6	124	2991	1	4680	4879	0	8798	3330	1	66	5	54	1107	29988	5	46	50	0	213	3330	1
50	2.3	7	44	103	10	0	129	3012	1	4020	4200	0	8620	3098	1	80	7	64	1200	26067	5	39	42	0	237	5474	1
51	0.9	3	29	101	9	0	126	3318	1	3526	3670	0	6550	12896	2	33	3	26	661	25547	4	36	39	0	174	20125	4
52	3.8	12	55	104	8	0	124	3049	1	4173	4376	0	7560	3098	1	132	10	111	1667	11373	2	41	44	0	306	5244	1
53	1.0	3	44	108	6	0	130	3530	1	2696	3066	0	5580	3125	1	33	2	27	588	13372	2	25	30	0	158	5308	1
54	0.4	2	59	105	7	0	130	3078	1	2035	2419	0	4680	4078	1	14	1	11	249	11824	2	19	24	0	161	5265	1
55	0.5	1	67	103	8	0	124	3746	1	3262	3420	0	6780	3888	1	18	1	15	316	14142	2	32	35	0	236	5898	1
56	6.1	16	108	105	8	0	122	3047	1	3586	3773	0	7433	3094	1	210	16	180	3128	30032	5	35	38	0	289	7874	1
57	0.8	3	170	103	6	0	124	3050	1	3180	3280	0	6713	3148	1	26	1	22	362	26346	5	31	34	0	172	3412	1
58	4.0	12	318	102	9	0	121	3047	1	3540	3636	0	7004	3094	1	143	13	120	2072	30044	5	36	39	0	162	3168	1
59	0.9	3	188	103	10	0	130	7885	1	2845	2858	0	5880	3153	1	33	3	26	682	35141	6	29	30	0	193	8768	2
60	3.4	11	272	100	14	0	130	3076	1	-	-	-	-	568799	100	122	16	95	2436	43059	8	-	-	-	-	568799	100
61	0.8	2	196	106	10	0	130	112416	20	3540	3240	0	6600	369545	65	29	3	23	606	145668	26	36	30	0	225	376335	66
62	7.8	20	477	102	13	0	129	217427	38	3600	3515	0	7440	113175	20	274	32	218	5611	236901	42	42	27	0	397	230320	40
63	1.0	3	94	92	10	0	130	428508	75	3153	2886	0	5820	369878	65	40	4	28	732	431782	76	40	18	0	340	430359	76
64	4.7	11	240	97	12	0	130	391996	69	3600	3270	0	6720	498840	88	176	21	131	2843	403671	71	37	35	0	200	499590	88
65	0.6	2	89	0	0	0	0	566652	100	-	-	-	-	568799	100	-	-	-	-	568799	100	-	-	-	-	568799	100
66	1.8	4	120	101	10	0	126	3143	1	-	-	-	-	568799	100	64	6	51	1289	15667	3	-	-	-	-	568799	100
67	4.5	11	207	101	7	0	123	3064	1	3331	3290	0	6968	3095	1	160	11	136	1789	41104	7	33	34	0	168	4980	1
68	0.9	2	114	105	9	0	130	163038	29	-	-	-	-	568799	100	31	2	25	649	167648	29	-	-	-	-	568799	100
69	0.5	1	77	101	6	0	122	3075	1	-	-	-	-	568799	100	16	1	13	329	41148	7	-	-	-	-	568799	100
70	1.9	6	78	103	8	0	122	3050	1	3472	3612	0	7685	3153	1	67	5	58	1152	13690	2	34	37	0	300	3360	1
71	9.5	27	164	104	8	0	121	3049	1	3457	3604	0	7620	3097	1	330	25	284	4289	39420	7	33	37	0	154	3309	1
72	0.9	3	49	101	8	0	120	3049	1	2319	2460	0	5340	3109	1	31	3	26	494	42645	7	23	26	0	147	3905	1
73	3.4	10	59	104	7	0	122	3047	1	3197	3356	30	6840	3097	1	119	8	102	1210	39413	7	31	34	1	158	3269	1
74	1.1	4	55	106	8	0	130	3100	1	2419	2544	0	5460	3095	1	36	3	30	772	39466	7	23	25	0	114	4014	1
75	7.8	24	186	104	7	0	127	3032	1	3060	3097	0	6291	3098	1	270	18	220	3993	42320	7	30	32	0	196	3540	1
76	1.2	2	309	100	9	0	126	3339	1	2011	2043	0	4646	3409	1	42	4	33	699	41417	7	21	22	0	200	5441	1

Table B.8: Summary statistics of data, computed using data between September 1st 2017 and September 31st 2018. This subset is discussed in chapter 5 and is used to fit the self-exciting point process model. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.9 Data between September 1st 2017 and November 30th 2017

In table B.9, we show summary statistics for the subset of data between September 1st 2017 and November 30th 2017. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. Considering missing data, we see that links 11, 12 and 13 report all missing flow values in this subset, whereas there are some non-missing values in the full dataset. Whilst link 11 reports speed values for a significant fraction of the window, this is missing for links 12 and 13. Similarly, we observe a higher proportion of missing flow data between links 60 and 69 in table B.9 than in table 3.1, but this is less extreme for speed values on the same links. Note that when our point process methodology is applied to this subset, we do not use flow data, so missing values here are less impactful on this aspect of our work.

Considering median speed values, the largest discrepancies are observed on links 42 and 62, with link 42 dropping by 6 km/hr and link 62 increasing by 4 km/hr compared to the main dataset. Median flow values are generally lower across the network in this subset compared to the full dataset, typically being between 0 and 200 veh/hr lower. If we also consider variation, we see the largest changes in inter-quartile range of speeds are observed on links 35, 37 and 64. It is unclear if this is due to significant events in the dataset, or other social factors. Differences in inter-quartile range of flows are negligible across all links. Since this data is collected between September and November, we expect there to be cooler temperatures and more rain on average than one would observe over the collection range of the entire dataset.

This dataset is used to fit point process models to incident data in section 5.4.6.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	20	100	6	0	116	32	0	2040	1953	0	4260	4089	3	26	1	23	438	12333	9	21	21	0	140	4095	3
1	3.6	13	27	101	7	16	115	32	0	3555	3420	0	6840	37	0	130	9	114	818	3152	2	36	36	0	166	37	0
2	1.0	3	20	106	6	11	126	0	0	2820	2732	0	5685	37	0	34	2	29	327	3120	2	27	27	0	127	37	0
3	0.5	2	4	105	7	7	124	0	0	2340	2260	0	4920	37	0	18	1	15	269	4627	4	22	23	0	128	37	0
4	0.4	1	5	105	7	7	124	0	0	2340	2248	0	4860	7823	6	14	1	12	210	3120	2	23	23	0	158	7823	6
5	4.5	2	9	108	7	11	127	0	0	3000	2973	120	6420	37	0	151	11	132	1481	8923	7	28	29	1	165	37	0
6	0.9	2	7	109	7	5	126	32	0	2356	2260	60	5038	37	0	29	2	25	624	8955	7	22	22	1	171	37	0
7	5.1	2	12	109	6	14	125	0	0	2697	2611	120	5472	37	0	168	10	147	1608	3662	3	24	24	1	193	37	0
8	1.0	1	20	89	4	0	105	484	0	1740	1621	0	4080	37	0	42	2	36	753	4146	3	20	19	1	164	1284	1
9	0.7	1	18	100	6	5	122	0	0	3158	2877	120	7098	37	0	23	1	19	469	3120	2	31	30	1	340	37	0
10	0.9	2	28	103	8	5	127	37181	28	3660	3412	180	7680	19	0	31	3	25	638	37190	28	41	17	5	390	37200	28
11	5.9	13	62	99	11	5	130	37169	28	-	-	-	-	131038	100	214	24	163	4248	37190	28	-	-	-	-	131038	100
12	1.0	2	29	-	-	-	-	131038	100	-	-	-	-	131038	100	-	-	-	-	131038	100	-	-	-	-	131038	100
13	6.8	17	64	-	-	-	-	131038	100	-	-	-	-	131038	100	-	-	-	-	131038	100	-	-	-	-	131038	100
14	1.0	2	8	105	10	8	118	32	0	-	-	-	-	131038	100	34	3	30	448	8955	7	-	-	-	-	131038	100
15	1.2	4	9	106	9	19	119	0	0	3780	3653	180	7380	37	0	40	4	36	225	4627	4	35	37	2	159	37	0
16	1.5	6	10	107	7	20	119	0	0	3720	3637	180	7260	37	0	51	4	46	275	4627	4	35	35	2	124	37	0
17	0.5	0	8	109	7	7	124	0	0	3720	3639	60	7256	20	0	16	1	14	242	3120	2	34	34	1	357	20	0
18	0.2	1	5	109	8	6	125	32	0	2520	2580	0	5640	37	0	6	0	5	110	12333	9	23	24	0	280	37	0
19	1.2	4	6	108	6	18	124	32	0	2363	2438	0	5328	37	0	41	2	36	246	12333	9	22	23	0	120	37	0
20	2.0	4	7	104	8	23	117	32	0	4179	3826	180	8100	37	0	71	6	62	314	9795	7	40	39	2	191	37	0
21	3.0	6	6	103	8	9	122	32	0	3540	3275	180	7080	37	0	104	9	88	1194	9795	7	34	33	2	243	37	0
22	10.0	22	19	111	8	29	126	32	0	4062	3780	180	8220	37	0	325	21	289	1243	9795	7	36	36	2	166	37	0
23	0.5	1	6	111	8	16	130	32	0	3307	3115	138	6623	304	0	17	1	14	116	9795	7	30	31	1	121	304	0
24	6.4	13	43	109	9	17	125	32	0	3900	3660	180	7500	37	0	211	18	184	1355	12292	9	36	38	2	143	37	0
25	0.8	2	46	108	15	11	124	81	0	3616	3441	0	7500	37	0	26	3	23	255	12341	9	35	39	0	198	86	0
26	2.6	9	47	105	17	0	123	44	0	3965	3780	0	7620	37	0	88	13	74	650	12355	9	38	43	0	174	100	0
27	0.9	2	47	106	18	0	125	38	0	2909	2835	0	5880	37	0	29	5	25	520	2952	2	30	33	0	170	3037	2
28	6.9	13	66	103	33	11	122	32	0	4519	4409	0	8068	37	0	244	110	204	2266	8955	7	46	65	0	165	37	0
29	0.9	3	13	100	25	11	122	32	0	3899	3864	0	6711	37	0	34	10	27	305	8955	7	41	54	0	134	37	0
30	2.5	4	8	97	18	0	119	0	0	4980	4752	0	8100	37	0	93	18	76	436	3120	2	51	58	0	232	253	0
31	1.2	1	16	101	11	16	119	46842	36	3180	3004	0	5820	44908	34	43	5	36	270	49879	38	34	32	0	101	46842	36
32	3.9	6	56	105	11	0	125	111	0	4680	4641	0	9300	37	0	133	13	112	819	12276	9	49	49	0	197	3871	3
33	1.0	3	122	106	14	0	121	32	0	3703	4140	0	8520	37	0	34	4	29	510	9795	7	44	50	0	190	897	1
34	1.9	3	114	103	24	0	118	32	0	4740	4826	0	10068	37	0	69	30	60	858	12288	9	57	72	2	281	59	0
35	1.7	5	170	101	43	0	119	32	0	3439	3991	0	7980	37	0	61	56	52	1022	12292	9	50	73	0	213	40	0
36	1.6	3	70	97	42	18	114	32	0	5518	4920	240	9480	37	0	60	43	50	311	12288	9	67	93	2	241	37	0
37	1.2	3	117	98	57	6	122	32	0	3154	3202	120	5700	37	0	46	48	36	711	7874	6	39	72	1	141	37	0
38	7.1	17	107	96	29	10	117	32	0	5340	5058	300	7909	37	0	268	96	221	2573	4659	4	62	73	3	225	37	0
39	0.7	2	40	97	16	0	119	32	0	3908	3691	0	6050	37	0	27	4	22	175	4659	4	44	44	0	135	239	0
40	0.6	2	59	100	17	0	126	1205	1	3175	3078	0	5100	40	0	23	4	19	389	5656	4	35	36	0	194	2208	2
41	1.9	5	44	95	18	0	117	40	0	4560	4504	0	7425	40	0	71	12	58	832	7882	6	52	53	0	180	4461	3

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	22	82	17	0	122	216	0	4493	4493	0	7320	40	0	58	11	39	795	7907	6	58	60	0	309	4279	3
43	4.6	11	32	97	12	0	121	173	0	4440	4440	0	7140	37	0	170	20	137	823	7860	6	49	49	0	180	4219	3
44	0.8	1	13	101	11	0	121	0	0	4500	4500	0	7380	38	0	28	3	24	205	7842	6	46	48	0	336	831	1
45	1.6	5	13	101	11	14	121	0	0	4716	4689	0	7980	37	0	58	6	48	418	7842	6	47	50	0	189	37	0
46	0.2	1	11	103	11	13	122	32	0	4495	4500	0	7800	47	0	9	1	7	68	7874	6	45	48	0	182	47	0
47	0.9	3	7	101	11	16	121	32	0	5028	5020	99	8684	37	0	32	4	27	203	4659	4	50	53	1	189	37	0
48	1.4	4	5	102	9	23	118	0	0	5028	5015	120	8557	37	0	48	4	42	214	4627	4	50	53	1	162	37	0
49	1.8	4	3	103	7	30	120	0	0	5051	5040	84	8613	37	0	65	4	55	221	7842	6	50	52	1	174	37	0
50	2.3	7	3	104	8	16	125	0	0	4291	4285	120	7380	37	0	79	6	66	1200	7842	6	41	43	1	237	37	0
51	0.9	3	4	101	10	13	123	32	0	3766	3734	60	6480	37	0	33	3	27	369	7874	6	37	40	1	147	37	0
52	3.8	12	20	105	8	14	124	32	0	4484	4440	120	7440	37	0	132	10	111	968	3694	3	44	45	1	231	37	0
53	1.0	3	23	108	7	15	127	32	0	2907	3091	120	5508	37	0	33	2	27	240	4659	4	27	30	1	158	37	0
54	0.4	2	30	105	7	8	130	32	0	2197	2426	0	4680	37	0	14	1	11	187	3694	3	21	25	0	161	37	0
55	0.5	1	30	101	7	6	119	104	0	3480	3420	0	6780	37	0	19	1	16	316	4659	4	35	36	0	197	109	0
56	6.1	16	30	104	8	0	119	32	0	3835	3780	0	7433	37	0	211	16	185	3128	7874	6	38	39	0	152	383	0
57	0.8	3	41	103	6	10	124	32	0	3400	3333	0	6713	37	0	26	1	22	271	7874	6	34	35	0	171	37	0
58	4.0	12	63	101	9	7	121	32	0	3816	3780	0	6960	37	0	144	13	120	2072	7874	6	38	40	0	162	37	0
59	0.9	3	48	104	10	5	130	342	0	3013	2883	0	5700	37	0	33	3	26	682	8184	6	30	30	0	145	347	0
60	3.4	11	50	100	11	0	130	32	0	-	-	-	-	131038	100	122	13	96	2436	8964	7	-	-	-	-	131038	100
61	0.8	2	39	108	9	0	124	109126	83	-	-	-	-	131038	100	28	2	24	244	111500	85	-	-	-	-	131038	100
62	7.8	20	74	105	11	0	126	112450	86	3900	3353	360	7320	110016	84	265	24	223	3507	113326	86	42	18	4	232	115852	88
63	1.0	3	17	-	-	-	-	131038	100	-	-	-	-	131038	100	-	-	-	-	131038	100	-	-	-	-	131038	100
64	4.7	11	27	93	5	42	109	116698	89	-	-	-	-	131038	100	184	10	158	405	116797	89	-	-	-	-	131038	100
65	0.6	2	11	-	-	-	-	131038	100	-	-	-	-	131038	100	-	-	-	-	131038	100	-	-	-	-	131038	100
66	1.8	4	22	102	8	6	126	42	0	-	-	-	-	131038	100	63	5	51	1074	4721	4	-	-	-	-	131038	100
67	4.5	11	38	101	6	13	117	1	0	3461	3253	0	6530	37	0	159	10	138	1239	8731	7	34	33	0	168	38	0
68	0.9	2	27	105	8	6	124	37213	28	-	-	-	-	131038	100	31	2	26	541	37765	29	-	-	-	-	131038	100
69	0.5	1	21	101	6	15	121	32	0	-	-	-	-	131038	100	16	1	13	113	8763	7	-	-	-	-	131038	100
70	1.9	6	20	103	7	0	119	32	0	3671	3604	0	7685	37	0	67	4	58	691	4711	4	36	37	0	160	104	0
71	9.5	27	35	105	8	0	119	32	0	3660	3600	60	7620	37	0	330	25	289	1716	8707	7	36	36	1	143	104	0
72	0.9	3	15	101	7	0	120	32	0	2468	2402	0	5310	37	0	31	3	26	309	8898	7	25	25	0	124	297	0
73	3.4	10	15	104	7	0	119	32	0	3360	3360	87	6720	37	0	119	8	105	724	8707	7	33	34	1	140	104	0
74	1.1	4	16	106	8	0	127	32	0	2500	2527	0	5100	37	0	36	3	30	482	8707	7	24	25	0	114	191	0
75	7.8	24	40	105	7	7	121	0	0	3163	3080	0	6291	37	0	268	18	231	3993	8731	7	32	32	0	196	37	0
76	1.2	2	71	103	7	0	126	0	0	2119	2098	0	4440	37	0	41	3	33	699	8731	7	22	21	0	189	103	0

Table B.9: Summary statistics of data, computed using data between September 1st 2017 and November 30th 2017. This subset is discussed in section 5.4.6 and is used to fit the self-exciting point process model. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.10 Data between December 1st 2017 and February 28th 2018

In table B.10, we show summary statistics for the subset of data between December 1st 2017 and February 28th 2018. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. Considering missing data, we see that links 11, 12 and 13 report less missing values than in the subset seen in appendix B.4.9, however link 14 still does not report flow values in this subset, but has some non-missing values in the full dataset. The same missing data patterns between links 60 and 69 discussed in appendix B.4.9 are observed here. Note as in appendix B.4.9 that when our point process methodology is applied to this subset, we do not use flow data, so missing values here are less impactful on this aspect of our work.

Considering median speed values, the largest discrepancies are observed on links 30 and 62, with link 30 increasing by 4 km/hr and link 62 increasing by 5 km/hr compared to the main dataset. As in appendix B.4.9, median flow values are generally lower across the network in this subset compared to the full dataset, typically being different by less than 500 veh/hr. If we also consider variation, we see the largest changes in inter-quartile range of speeds are observed on links 35, 36 and 37. As these are neighbouring links, it is likely that this is due to a physical factor during the time-period, rather than noise in the data. Since this dataset includes the Christmas period, this may indicate that these links are significant commuter links, and the reduction in commuter traffic may have resulted in less extreme drops in speed during incident periods. Differences in inter-quartile range of flows are negligible across all links. As a result of the time this subset was collected from, we expect the coldest temperatures to be observed during it and more severe weather conditions on average than one would observe over the collection range of the entire dataset.

This dataset is used to fit point process models to incident data in section 5.4.6.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	7	98	6	0	121	0	0	1680	1800	0	4118	59684	46	27	2	22	145	3673	3	17	19	0	89	59684	46
1	3.6	13	18	101	7	23	117	0	0	3006	3300	79	6960	4	0	130	9	112	565	0	0	30	34	1	153	4	0
2	1.0	3	8	106	6	20	126	0	0	2409	2608	60	5750	4	0	34	2	29	185	0	0	23	26	1	130	4	0
3	0.5	2	12	103	9	16	122	0	0	1983	2171	60	4850	4	0	18	2	15	116	681	1	19	21	1	142	4	0
4	0.4	1	9	103	9	16	122	0	0	1991	2204	60	4855	5565	4	14	1	12	90	0	0	19	21	1	139	5565	4
5	4.5	2	5	107	7	27	127	0	0	2542	2820	60	6238	4	0	152	10	128	603	3673	3	24	27	1	140	4	0
6	0.9	2	2	108	6	11	127	0	0	2001	2158	60	4538	4	0	29	2	25	284	3673	3	19	21	1	109	4	0
7	5.1	2	12	109	7	11	126	0	0	2280	2477	60	5213	4	0	169	11	146	1610	638	0	21	23	1	168	4	0
8	1.0	1	13	89	6	0	103	187	0	1500	1541	0	3980	4	0	42	3	37	627	638	0	17	18	0	230	571	0
9	0.7	1	10	100	5	0	118	0	0	2666	2878	120	7020	69899	54	23	1	20	344	638	0	26	29	1	248	69899	54
10	0.9	2	17	102	7	6	121	17042	13	3000	3029	180	7613	58579	45	31	3	26	527	17723	14	30	30	2	217	58874	45
11	5.9	13	65	102	13	7	121	29457	23	2940	2963	60	7320	59659	46	208	28	176	2946	30138	23	28	29	1	238	59659	46
12	1.0	2	27	107	6	32	121	87046	67	2724	2846	0	6817	87050	67	34	2	30	114	87684	68	26	28	0	139	87050	67
13	6.8	17	52	105	8	8	122	59680	46	2940	3061	120	6720	87048	67	232	18	199	3041	62685	48	29	30	1	140	87048	67
14	1.0	2	17	106	9	0	122	3	0	-	-	-	-	129599	100	34	3	29	189	3676	3	-	-	-	-	129599	100
15	1.2	4	15	106	8	19	122	0	0	3120	3384	60	7229	4	0	40	3	35	225	681	1	29	33	1	146	4	0
16	1.5	6	14	106	7	23	122	0	0	3088	3360	60	7080	4	0	52	4	45	237	681	1	29	32	1	147	4	0
17	0.5	0	12	108	7	13	127	0	0	3060	3360	60	7200	99	0	16	1	13	129	0	0	28	32	1	268	99	0
18	0.2	1	8	109	7	10	129	0	0	2100	2371	60	5630	4	0	6	0	5	65	3673	3	19	22	1	148	4	0
19	1.2	4	11	108	6	11	127	0	0	1969	2266	60	5340	4	0	41	2	35	403	3673	3	18	22	1	131	4	0
20	2.0	4	9	104	8	18	120	0	0	3456	3572	120	7800	4	0	70	6	60	403	3673	3	33	36	1	189	4	0
21	3.0	6	16	103	9	12	121	0	0	2940	3040	120	6960	4	0	104	9	89	896	3673	3	28	30	1	189	4	0
22	10.0	22	30	111	7	17	126	0	0	3360	3488	120	7860	4	0	326	21	286	2120	3673	3	30	33	1	164	4	0
23	0.5	1	14	109	8	13	130	0	0	2772	2919	0	6637	4	0	17	1	14	116	3673	3	25	28	0	151	4	0
24	6.4	13	43	109	9	23	124	0	0	3270	3517	88	7260	4	0	210	16	186	1002	3673	3	30	34	1	146	4	0
25	0.8	2	34	108	9	13	127	0	0	3051	3261	0	7020	4	0	26	2	22	216	3673	3	28	33	0	206	4	0
26	2.6	9	34	107	8	16	126	0	0	3307	3586	120	7091	4	0	86	6	74	575	3673	3	31	36	1	187	4	0
27	0.9	2	38	108	11	13	126	0	0	2520	2671	60	5863	4	0	29	3	25	240	0	0	23	28	1	131	4	0
28	6.9	13	55	105	13	20	124	0	0	3840	4176	120	7605	4	0	237	31	201	1246	3673	3	36	45	1	168	4	0
29	0.9	3	16	103	17	16	130	0	0	3420	3668	120	6660	4	0	33	5	26	210	3673	3	33	43	1	152	4	0
30	2.5	4	10	103	13	19	122	0	0	4233	4620	180	8023	4	0	88	11	74	475	0	0	41	51	2	188	4	0
31	1.2	1	0	-	-	-	-	129599	100	-	-	-	-	129599	100	-	-	-	-	129599	100	-	-	-	-	129599	100
32	3.9	6	51	106	9	18	122	0	0	4206	4479	180	9240	4	0	131	11	114	774	3673	3	40	47	2	215	4	0
33	1.0	3	85	106	10	11	123	0	0	3557	3929	120	8386	4	0	33	3	28	320	3673	3	36	43	1	186	4	0
34	1.9	3	91	105	12	13	121	0	0	4392	4662	180	9926	4	0	67	8	58	529	3673	3	44	56	2	275	4	0
35	1.7	5	154	103	16	6	123	0	0	3336	3871	70	8040	4	0	60	10	50	1022	3673	3	38	50	1	208	4	0
36	1.6	3	44	97	16	21	113	130	0	5031	4860	0	9453	144	0	58	11	50	265	3803	3	54	66	0	193	144	0
37	1.2	3	77	100	22	6	122	130	0	2771	3090	0	5760	140	0	44	11	36	739	2426	2	29	42	0	139	140	0
38	7.1	17	74	100	20	6	120	130	0	4592	4939	60	7726	139	0	259	56	214	4288	811	1	47	62	1	226	139	0
39	0.7	2	35	101	13	5	130	131	0	3345	3653	0	6000	140	0	26	4	20	516	812	1	34	41	0	131	140	0
40	0.6	2	41	103	13	5	121	131	0	2640	2988	0	4920	140	0	23	3	19	467	812	1	26	33	0	156	140	0
41	1.9	5	38	100	14	8	117	132	0	3595	4208	120	7020	144	0	68	10	58	847	933	1	38	48	1	164	144	0

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	24	85	17	6	123	132	0	3540	4224	0	6955	144	0	56	10	38	795	933	1	43	56	0	341	144	0
43	4.6	11	25	98	11	6	117	132	0	3540	4156	0	6893	142	0	168	19	141	2621	2428	2	37	47	0	223	142	0
44	0.8	1	10	101	9	5	119	132	0	3720	4166	0	6840	44602	34	28	3	24	573	2428	2	38	45	0	336	44602	34
45	1.6	5	11	101	9	5	119	132	0	3662	4439	0	7658	142	0	58	5	49	1171	2428	2	37	47	0	174	142	0
46	0.2	1	9	101	12	7	126	132	0	3517	4226	0	7313	519	0	9	1	7	134	2428	2	35	45	0	171	519	0
47	0.9	3	7	101	11	8	119	132	0	3891	4767	0	8278	144	0	32	3	27	407	813	1	38	51	0	319	144	0
48	1.4	4	7	102	11	11	118	127	0	3876	4749	0	8400	142	0	48	5	41	459	808	1	38	50	0	168	142	0
49	1.8	4	5	103	9	18	120	127	0	3900	4746	0	8271	334	0	65	6	55	380	2423	2	38	49	0	210	334	0
50	2.3	7	6	104	8	0	128	127	0	3361	4064	0	7080	142	0	79	6	64	1036	1915	1	32	41	0	236	790	1
51	0.9	3	5	103	10	0	124	132	0	2918	3527	0	6111	144	0	32	3	27	331	1290	1	28	37	0	174	511	0
52	3.8	12	8	105	9	0	121	132	0	3459	4260	80	7320	142	0	131	11	114	1667	813	1	33	43	1	234	195	0
53	1.0	3	2	108	8	7	130	132	0	2206	2926	0	5328	142	0	33	3	27	545	813	1	20	28	0	115	142	0
54	0.4	2	7	105	8	6	130	132	0	1653	2295	0	4428	803	1	14	1	11	249	813	1	16	23	0	82	803	1
55	0.5	1	4	103	6	6	119	182	0	2760	3253	0	6410	537	0	18	1	16	316	863	1	26	33	0	136	587	0
56	6.1	16	13	105	8	0	122	130	0	3000	3600	0	6960	139	0	209	16	180	1564	2426	2	29	36	1	159	466	0
57	0.8	3	25	103	8	0	124	130	0	2700	3133	0	6341	140	0	26	2	22	246	2109	2	26	33	0	167	343	0
58	4.0	12	49	102	9	10	119	130	0	3051	3548	120	6715	139	0	143	12	122	1450	2426	2	30	37	1	159	139	0
59	0.9	3	31	103	10	5	128	456	0	2451	2721	0	5552	140	0	33	3	26	682	2752	2	24	29	0	193	465	0
60	3.4	11	45	101	12	0	128	147	0	-	-	-	-	129599	100	121	15	95	2436	3778	3	-	-	-	-	129599	100
61	0.8	2	40	107	10	0	129	177	0	-	-	-	-	129599	100	28	3	23	606	3808	3	-	-	-	-	129599	100
62	7.8	20	71	106	8	0	127	47653	37	3132	3388	0	7067	174	0	265	21	221	4676	48555	37	41	21	0	315	48238	37
63	1.0	3	19	-	-	-	-	129599	100	-	-	-	-	129599	100	-	-	-	-	129599	100	-	-	-	-	129599	100
64	4.7	11	45	-	-	-	-	129599	100	-	-	-	-	129599	100	-	-	-	-	129599	100	-	-	-	-	129599	100
65	0.6	2	24	-	-	-	-	129599	100	-	-	-	-	129599	100	-	-	-	-	129599	100	-	-	-	-	129599	100
66	1.8	4	28	101	9	5	125	167	0	-	-	-	-	129599	100	64	5	52	1289	968	1	-	-	-	-	129599	100
67	4.5	11	38	101	7	13	118	128	0	2880	3092	0	6660	139	0	159	11	136	1239	3540	3	28	31	0	165	139	0
68	0.9	2	22	105	9	8	127	36885	28	-	-	-	-	129599	100	31	2	26	464	37220	29	-	-	-	-	129599	100
69	0.5	1	15	100	6	5	122	133	0	-	-	-	-	129599	100	16	1	14	329	3545	3	-	-	-	-	129599	100
70	1.9	6	13	103	8	0	119	132	0	2961	3423	0	7526	143	0	67	6	58	628	933	1	29	35	1	143	283	0
71	9.5	27	33	104	8	0	121	132	0	2910	3360	120	7260	142	0	330	27	284	2860	3116	2	28	34	1	142	287	0
72	0.9	3	7	101	8	0	119	132	0	1980	2273	0	5263	142	0	31	3	26	494	3718	3	20	24	1	142	247	0
73	3.4	10	14	103	7	0	122	130	0	2768	3180	30	6403	142	0	120	8	102	954	3109	2	27	32	1	134	247	0
74	1.1	4	10	106	8	0	126	130	0	2050	2357	0	4916	139	0	36	3	31	772	3109	2	20	24	0	108	280	0
75	7.8	24	38	104	8	0	120	129	0	2580	2940	81	5700	139	0	269	19	233	3105	3715	3	26	30	1	163	143	0
76	1.2	2	53	101	8	0	126	348	0	1680	1920	0	4182	219	0	42	3	33	699	3763	3	17	19	0	200	886	1

Table B.10: Summary statistics of data, computed using data between December 1st 2017 and February 28th 2018. This subset is discussed in section 5.4.6 and is used to fit the self-exciting point process model. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.11 Data between March 1st 2018 and May 31st 2018

In table B.11, we show summary statistics for the subset of data between March 1st 2018 and May 31st 2018. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. Considering missing data, we see that links 11, 12 and 13 now report valid data at almost all times in the window, in contrast to what was seen in appendix B.4.9 and B.4.10. However link 9 does not report flow values in this subset, and link 31 reports missing data for all variables, but has a significant amount of non-missing values in the full dataset. The same missing data patterns between links 60 and 69 discussed in appendix B.4.9 and B.4.10 are observed here. Again note that as in appendix B.4.9 and B.4.10 that when our point process methodology is applied to this subset, we do not use flow data, so missing values here are less impactful on this aspect of our work.

Considering median speed values, the largest discrepancies are observed on links 41, 42, 62 and 63, with links 41 and 42 having a lower median speed by 3 and 6 km/hr respectively, and links 62 and 63 having a higher median speed by 5 and 8 km/hr respectively to the main dataset. The fact that these differences occur on two pairs of links, each pair being adjacent in space, suggests a single factor may have impacted links 41 and 42, and similarly for links 62 and 63. As in appendix B.4.9 and B.4.10, median flow values are generally lower across the network in this subset compared to the full dataset, typically being different by less than 200 veh/hr. If we also consider variation, we see the largest changes in inter-quartile range of speeds are observed on links 63 and 64. This is likely explained by the high fraction of missing speed values on these links during the subset, and hence insufficient data is available in this subset to accurately reflect the spread of the data over a wider set of times. Differences in inter-quartile range of flows are negligible across all links. As a result of the time period this subset was collected in, we expect temperatures to be increasing compared to appendix B.4.10, and less severe weather conditions than one would observe over the collection range of the entire dataset.

This dataset is used to fit point process models to incident data in section 5.4.6.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	20	100	6	18	117	1474	1	1860	1920	0	4298	1494	1	26	1	22	146	10224	8	18	20	0	107	1494	1
1	3.6	13	23	101	7	21	114	1455	1	3329	3410	64	6900	1489	1	130	9	115	623	1455	1	33	35	1	155	1489	1
2	1.0	3	15	106	6	15	121	1483	1	2640	2687	120	5848	1494	1	34	2	30	246	1483	1	25	27	1	128	1494	1
3	0.5	2	4	105	8	13	129	1474	1	2180	2258	60	5073	1498	1	18	2	14	142	2560	2	21	22	1	122	1498	1
4	0.4	1	4	105	8	13	129	1474	1	2160	2261	60	5100	1557	1	14	1	11	111	1474	1	21	22	1	123	1557	1
5	4.5	2	6	108	8	20	126	1472	1	2794	2957	120	6358	1496	1	151	12	129	814	9422	7	26	29	1	111	1496	1
6	0.9	2	4	109	8	13	127	1456	1	2184	2277	68	5160	1515	1	29	2	25	240	9406	7	20	22	1	103	1515	1
7	5.1	2	20	111	9	10	127	1434	1	2480	2640	71	5760	1509	1	166	14	145	1497	1796	1	23	24	1	215	1509	1
8	1.0	1	28	90	5	0	115	1474	1	1620	1620	0	4320	1578	1	42	2	32	627	1836	1	18	19	1	158	1965	1
9	0.7	1	21	101	6	14	114	1456	1	-	-	-	-	132479	100	23	1	21	168	1456	1	-	-	-	-	132479	100
10	0.9	2	22	103	7	17	114	1455	1	3152	3214	120	8160	1497	1	31	2	28	188	2541	2	31	33	2	159	1497	1
11	5.9	13	72	105	8	16	116	1455	1	3128	3180	74	7800	1489	1	202	15	183	1328	2541	2	30	32	1	155	1489	1
12	1.0	2	24	108	8	14	122	1474	1	2894	2947	0	7680	1565	1	34	3	30	263	2560	2	27	29	0	163	1565	1
13	6.8	17	84	104	9	8	120	1452	1	3118	3180	120	6900	1493	1	233	22	203	3041	9401	7	30	31	1	309	1493	1
14	1.0	2	7	105	11	14	121	1520	1	2940	2760	0	5940	33346	25	34	4	30	256	9443	7	28	28	0	163	33405	25
15	1.2	4	8	105	10	14	119	1434	1	3460	3707	63	7260	1495	1	41	4	36	377	2520	2	33	37	1	157	1495	1
16	1.5	6	7	105	8	14	119	1434	1	3420	3660	85	7380	1533	1	52	4	46	486	2520	2	32	36	1	184	1533	1
17	0.5	0	11	108	9	11	125	1467	1	3420	3600	0	6900	2197	2	16	1	14	179	1467	1	31	34	0	273	2199	2
18	0.2	1	11	109	9	8	129	1488	1	2316	2638	0	5820	1505	1	6	0	5	47	10239	8	21	24	0	148	1538	1
19	1.2	4	22	108	8	10	122	1493	1	2162	2510	0	5257	1489	1	41	3	36	317	10244	8	20	23	0	134	1527	1
20	2.0	4	16	103	10	8	116	1457	1	3861	3904	60	8080	1488	1	71	6	63	519	10206	8	37	39	1	241	1490	1
21	3.0	6	20	104	11	6	125	1457	1	3281	3317	0	7070	1494	1	103	11	86	1756	10206	8	31	33	0	428	1496	1
22	10.0	22	28	110	9	0	124	1455	1	3720	3780	0	8100	1489	1	328	26	291	2347	10206	8	34	36	2	292	2300	2
23	0.5	1	17	109	8	13	127	1455	1	3060	3125	0	7012	11214	8	17	2	15	139	10206	8	28	31	0	133	11214	8
24	6.4	13	78	109	11	15	125	1455	1	3600	3700	180	7620	1493	1	211	23	184	1536	10206	8	33	39	2	168	1493	1
25	0.8	2	64	108	11	14	126	1455	1	3300	3420	0	7715	1582	1	26	3	23	194	10206	8	31	39	0	187	1582	1
26	2.6	9	57	107	13	19	121	1455	1	3636	3762	60	7562	1488	1	86	13	76	481	10206	8	34	43	1	164	1488	1
27	0.9	2	60	107	15	10	121	1455	1	2760	2789	60	6262	1488	1	29	4	26	312	1455	1	26	33	1	190	1488	1
28	6.9	13	81	105	28	11	120	1455	1	4234	4327	240	8160	1488	1	240	96	208	2205	9405	7	40	63	2	207	1488	1
29	0.9	3	11	103	24	6	126	1455	1	3741	3769	0	7020	1489	1	33	9	27	559	9405	7	37	51	0	149	1489	1
30	2.5	4	11	103	18	16	119	1434	1	4688	4740	0	8100	1514	1	88	16	76	558	1434	1	45	56	0	186	1514	1
31	1.2	1	0	-	-	-	-	132479	100	-	-	-	-	132479	100	-	-	-	-	132479	100	-	-	-	-	132479	100
32	3.9	6	51	105	10	14	119	1455	1	4560	4543	187	9360	1488	1	133	13	117	995	10206	8	45	50	2	487	1488	1
33	1.0	3	114	106	13	7	121	1475	1	3610	3986	180	8700	1495	1	34	4	29	546	9424	7	41	48	2	334	1495	1
34	1.9	3	110	103	20	11	119	1474	1	4620	4713	240	10200	1519	1	68	22	59	656	10224	8	51	69	3	414	1519	1
35	1.7	5	151	101	39	8	119	1484	1	3402	3911	240	8097	1493	1	61	43	52	861	10239	8	45	70	2	275	1503	1
36	1.6	3	50	96	39	26	114	1474	1	5338	4819	243	9926	1513	1	60	38	51	217	10224	8	62	91	2	195	1513	1
37	1.2	3	70	98	49	10	117	1455	1	3044	3120	0	5780	1494	1	46	39	38	443	6689	5	35	66	0	197	1494	1
38	7.1	17	72	98	24	10	116	1455	1	5096	5000	0	7860	1488	1	264	71	222	2573	2239	2	55	67	0	222	1488	1
39	0.7	2	23	98	13	8	130	1528	1	3750	3654	0	6480	1514	1	26	3	20	349	2312	2	39	42	0	155	1561	1
40	0.6	2	32	100	13	6	121	1479	1	2660	2760	0	4958	28751	22	23	3	19	389	2263	2	27	32	0	145	28751	22
41	1.9	5	30	95	13	10	116	1478	1	4231	4320	0	7560	1496	1	71	10	58	678	3619	3	45	51	0	213	1496	1

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	17	82	15	6	124	1474	1	4148	4320	0	7380	1494	1	59	10	38	795	3615	3	51	59	0	210	1494	1
43	4.6	11	27	94	10	9	128	1434	1	4140	4320	0	7333	1489	1	175	19	129	1324	6712	5	45	49	0	229	1489	1
44	0.8	1	8	100	8	0	119	1457	1	4020	4227	0	7140	5546	4	29	3	24	220	6735	5	41	45	0	271	5618	4
45	1.6	5	10	100	8	14	119	1434	1	4302	4538	0	7920	1492	1	59	4	49	418	6668	5	43	48	0	194	1492	1
46	0.2	1	10	100	9	14	130	1455	1	4117	4310	0	7819	1853	1	9	1	7	64	6689	5	42	46	0	202	1853	1
47	0.9	3	9	100	10	13	117	1455	1	4582	4863	180	8640	1497	1	33	3	28	250	2541	2	46	52	2	189	1497	1
48	1.4	4	10	100	9	18	121	1434	1	4560	4851	240	8460	1490	1	49	4	41	274	2520	2	46	51	2	245	1490	1
49	1.8	4	7	101	9	15	124	1434	1	4538	4857	125	8798	1529	1	66	5	54	518	6668	5	45	50	1	401	1529	1
50	2.3	7	13	103	9	8	126	1455	1	3932	4144	180	8620	1489	1	80	7	66	1036	3870	3	38	42	2	308	1489	1
51	0.9	3	7	100	9	0	126	1485	1	3451	3647	0	6550	1507	1	33	3	26	661	3900	3	34	38	0	127	1791	1
52	3.8	12	10	104	8	22	122	1455	1	4058	4318	0	7320	1489	1	132	10	113	659	1817	1	39	44	0	205	1489	1
53	1.0	3	4	106	8	0	129	1455	1	2597	3013	0	5580	1516	1	33	2	27	114	2541	2	24	29	0	105	2255	2
54	0.4	2	10	105	7	0	130	1484	1	1950	2381	0	4500	1807	1	14	1	11	149	2268	2	19	23	0	94	2249	2
55	0.5	1	16	103	8	0	124	1907	1	3168	3420	0	6660	1868	1	18	1	15	150	2993	2	31	35	0	181	2733	2
56	6.1	16	28	104	8	0	118	1455	1	3480	3727	0	7200	1488	1	211	16	186	1824	6691	5	35	37	0	289	4819	4
57	0.8	3	36	103	8	8	124	1458	1	3091	3293	0	6441	1496	1	26	2	22	362	3873	3	31	34	0	172	1497	1
58	4.0	12	81	101	9	10	118	1455	1	3420	3608	139	7004	1488	1	144	13	123	1443	6689	5	35	39	2	160	1488	1
59	0.9	3	49	101	10	0	130	1983	1	2774	2827	0	5880	1501	1	34	3	26	569	7942	6	29	31	0	155	2807	2
60	3.4	11	67	99	14	0	130	1462	1	-	-	-	-	132479	100	123	17	95	2436	9412	7	-	-	-	-	132479	100
61	0.8	2	38	106	11	0	130	1626	1	3540	3211	60	6480	107381	81	29	3	23	606	9403	7	35	32	2	144	107398	81
62	7.8	20	113	106	13	0	129	44920	34	3573	3528	0	7440	1510	1	265	31	218	5611	47238	36	42	20	0	6278	47010	35
63	1.0	3	20	100	19	0	130	115674	87	3180	2874	0	5820	107718	81	36	6	28	320	116597	88	36	17	2	282	116685	88
64	4.7	11	50	97	19	0	130	114824	87	-	-	-	-	132479	100	174	23	131	1312	116612	88	-	-	-	-	132479	100
65	0.6	2	21	0	0	0	0	131364	99	-	-	-	-	132479	100	-	-	-	-	132479	100	-	-	-	-	132479	100
66	1.8	4	27	102	11	0	126	1504	1	-	-	-	-	132479	100	63	6	51	1074	3651	3	-	-	-	-	132479	100
67	4.5	11	54	101	7	0	118	1463	1	3300	3298	0	6968	1489	1	160	11	136	1610	8523	6	33	33	0	273	2582	2
68	0.9	2	23	106	11	5	130	38640	29	-	-	-	-	132479	100	31	3	25	600	39294	30	-	-	-	-	132479	100
69	0.5	1	16	101	6	15	122	1479	1	-	-	-	-	132479	100	16	1	13	141	8530	6	-	-	-	-	132479	100
70	1.9	6	23	103	7	9	119	1456	1	3420	3600	60	7560	1499	1	67	6	58	747	2542	2	33	37	1	300	1499	1
71	9.5	27	29	104	8	16	119	1455	1	3419	3583	120	7320	1488	1	330	25	288	2221	8509	6	33	37	1	151	1488	1
72	0.9	3	8	101	8	10	116	1455	1	2220	2406	0	5336	1500	1	31	3	27	307	9129	7	22	25	0	147	1500	1
73	3.4	10	12	104	7	14	118	1455	1	3129	3370	180	6840	1488	1	120	8	105	920	8509	6	30	34	2	158	1488	1
74	1.1	4	10	106	8	6	126	1474	1	2365	2515	120	5076	1489	1	36	3	31	643	8528	6	22	25	1	113	1489	1
75	7.8	24	44	104	8	10	120	1473	1	2940	3096	180	6088	1492	1	269	19	233	2795	9146	7	29	31	2	159	1492	1
76	1.2	2	60	100	10	6	124	1561	1	1928	2040	0	4560	1723	1	42	4	34	699	8613	7	20	22	0	189	1729	1

Table B.11: Summary statistics of data, computed using data between March 1st 2018 and May 31st 2018. This subset is discussed in section 5.4.6 and is used to fit the self-exciting point process model. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

B.4.12 Data between June 1st 2018 and August 31st 2018

In table B.12, we show summary statistics for the subset of data between June 1st 2018 and August 31st 2018. The differences of note between this, and the entire dataset shown in table 3.1 are as follows. Considering missing data, we see that link 31 reports missing data for all variables at all times in this subset, however it reports non-missing values for large parts of the full dataset. Recall that as in appendix B.4.9, B.4.10 and B.4.11 that when our point process methodology is applied to this subset, we do not use flow data, so missing values here are less impactful on this aspect of our work.

Considering median speed values, the largest discrepancies are observed on links 35, 36, 38 and 68, all of which have a lower median speed in this subset of around 4 km/hr compared to the full dataset. Since three of these links are very near each other in space, we again might consider a single cause of these drops in speed. Median flow values are generally higher across the network in this subset compared to the full dataset, typically being different by less than 350 veh/hr. If we also consider variation, we see the largest changes in inter-quartile range of speeds are observed on links 25, 26, 27, 28, 34 and 35. These changes are large compared to what we typically see in other subsets, having higher variations by up to 30 km/hr. This suggests a significant amount of variability in this subset of data compared to the entire dataset. When we consider why this might be, one could postulate that, because an extended school holiday occurs during this time window, this may have lead to some changes in commuter behaviour as they take holidays from work to coincide with this. Differences in inter-quartile range of flows range between increases of up to 130 veh/hr and decreases of up to 400 veh/hr. but again these are negligible relative to the absolute values the attain. In the time window this dataset is collected in, we expect temperatures to be the highest one would encounter in the UK, and therefore higher on average than in the full dataset.

This dataset is used to fit point process models to incident data in section 5.4.6.

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
0	0.7	3	11	100	6	0	127	1430	1	2176	1995	0	4140	1430	1	26	1	22	233	16701	13	22	21	0	133	1828	1
1	3.6	13	16	102	7	14	117	1430	1	3752	3396	180	7131	1430	1	128	9	112	935	1611	1	37	36	2	176	1430	1
2	1.0	3	15	106	7	11	122	1436	1	2820	2720	120	5943	1442	1	34	2	29	327	1617	1	26	27	1	123	1442	1
3	0.5	2	15	106	8	13	129	1480	1	2500	2293	0	5213	1486	1	17	1	14	142	2996	2	24	23	0	137	1486	1
4	0.4	1	15	106	8	13	129	1480	1	2489	2297	0	5213	1442	1	14	1	11	111	1661	1	24	23	0	140	1486	1
5	4.5	2	15	108	9	0	129	1436	1	3180	2988	0	6540	1486	1	152	15	126	1727	15816	12	29	30	0	141	1522	1
6	0.9	2	10	109	8	5	130	1436	1	2520	2340	0	4860	1441	1	29	2	24	624	15816	12	23	23	0	141	1441	1
7	5.1	2	28	110	9	8	127	1430	1	2873	2715	0	5628	1523	1	168	14	145	2301	1611	1	26	26	0	244	1523	1
8	1.0	1	46	90	5	5	116	1430	1	1824	1652	0	4358	1431	1	42	2	32	692	1611	1	21	20	0	190	1431	1
9	0.7	1	32	100	6	16	113	1436	1	3300	2864	300	7006	88687	67	23	1	21	147	1617	1	34	32	3	135	88687	67
10	0.9	2	47	103	7	12	115	1436	1	3376	3225	240	7920	1442	1	31	2	28	264	2952	2	33	34	2	167	1442	1
11	5.9	13	127	104	9	14	116	1436	1	3558	3177	180	7680	1442	1	204	19	183	1517	2952	2	35	34	3	148	1442	1
12	1.0	2	54	108	9	16	126	1479	1	3300	2962	0	7359	1486	1	34	3	29	230	2995	2	31	31	0	168	1486	1
13	6.8	17	101	104	12	13	119	1431	1	3540	3220	84	6900	1442	1	236	33	204	1872	15785	12	34	35	1	205	1442	1
14	1.0	2	24	103	15	14	120	1475	1	3120	2747	0	6120	1437	1	35	6	30	255	15786	12	30	31	0	243	1481	1
15	1.2	4	12	105	14	0	121	1430	1	3989	3743	0	7176	1486	1	41	6	36	234	2948	2	38	41	0	154	1648	1
16	1.5	6	11	105	11	0	119	1430	1	3918	3720	0	7020	1441	1	52	6	47	211	2948	2	37	39	0	141	1603	1
17	0.5	0	8	107	10	0	124	1434	1	3909	3720	0	7020	1434	1	16	2	14	151	1618	1	36	38	0	266	1600	1
18	0.2	1	10	109	11	8	127	1509	1	2700	2671	0	5508	1486	1	6	1	5	80	16707	13	24	27	0	145	1559	1
19	1.2	4	20	108	10	6	124	1544	1	2524	2580	0	5580	1486	1	41	4	36	739	16751	13	23	26	0	207	1550	1
20	2.0	4	13	102	9	12	117	1436	1	4366	3804	0	7901	1442	1	72	6	63	636	16629	13	43	41	0	243	1442	1
21	3.0	6	22	101	10	7	121	1436	1	3660	3243	0	7056	1486	1	106	11	89	1509	16629	13	36	34	0	241	1486	1
22	10.0	22	33	108	9	13	123	1436	1	4200	3697	120	8040	1441	1	334	28	293	2762	16629	13	39	37	2	165	1441	1
23	0.5	1	37	109	10	13	127	1436	1	3400	3022	0	6761	1950	1	17	2	15	142	16629	13	32	32	0	176	1950	1
24	6.4	13	125	107	18	7	122	1436	1	4012	3589	360	7620	1442	1	217	47	189	3291	16660	13	39	43	3	200	1442	1
25	0.8	2	87	106	33	10	123	1436	1	3615	3303	300	7365	1485	1	27	16	23	255	16660	13	37	51	3	206	1485	1
26	2.6	9	74	105	36	12	121	1436	1	4050	3604	360	7483	1442	1	88	52	76	614	16660	13	41	60	3	198	1442	1
27	0.9	2	84	106	37	0	125	1436	1	3021	2684	0	5987	1442	1	29	14	25	284	1465	1	32	45	0	178	1481	1
28	6.9	13	102	101	45	17	118	1436	1	4740	4093	360	7800	1441	1	257	174	212	1466	15816	12	50	78	3	167	1441	1
29	0.9	3	19	99	31	11	124	1436	1	4140	3572	346	6840	1441	1	36	14	27	305	15816	12	45	55	3	145	1441	1
30	2.5	4	17	97	21	16	119	1430	1	5220	4440	360	8241	1442	1	93	21	76	564	1459	1	55	56	4	157	1442	1
31	1.2	1	0	-	-	-	-	132479	100	-	-	-	-	132479	100	-	-	-	-	132479	100	-	-	-	-	132479	100
32	3.9	6	66	103	13	0	120	1436	1	4882	4231	0	9600	1441	1	137	16	116	1266	16660	13	51	50	3	221	1867	1
33	1.0	3	186	105	22	0	122	1436	1	3816	3644	0	8700	1442	1	34	10	29	520	16629	13	47	51	0	197	2329	2
34	1.9	3	165	100	50	11	123	1436	1	4906	4256	194	10063	1442	1	70	74	57	760	16660	13	62	90	2	309	1442	1
35	1.7	5	207	97	67	6	124	1474	1	3540	3458	0	8138	1430	1	63	117	49	1022	16698	13	55	100	0	233	1474	1
36	1.6	3	69	92	45	23	111	1430	1	5667	4335	0	9570	1456	1	63	50	51	346	16654	13	73	93	0	264	1456	1
37	1.2	3	85	95	55	11	116	1430	1	3375	2862	0	5860	1430	1	48	49	38	403	7939	6	42	70	0	134	1430	1
38	7.1	17	83	93	27	10	118	1430	1	5643	4664	60	7977	1430	1	277	85	219	2573	2593	2	65	66	1	208	1430	1
39	0.7	2	33	97	14	0	130	1430	1	4140	3437	0	6180	1434	1	27	4	20	244	2593	2	45	41	0	141	1530	1
40	0.6	2	39	98	15	0	130	1985	1	3376	2958	0	5100	12593	10	24	4	18	467	3062	2	36	34	0	158	13174	10
41	1.9	5	38	97	16	0	117	1510	1	4797	4328	0	7371	1475	1	70	11	58	1355	5292	4	52	50	0	216	3067	2

Link Num	Length (km)	Num Loops	Num Events	Speed (km/hr)						Flow (veh/hr)						Travel Time (seconds)						Density (veh/km)					
				Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing	Median	IQR	Min	Max	Num Missing	% Missing
42	1.3	3	25	90	21	0	119	1634	1	4691	4295	0	7253	1431	1	53	12	40	954	5351	4	53	55	0	382	2788	2
43	4.6	11	39	96	13	0	116	1475	1	4682	4286	0	7246	1430	1	171	23	142	2497	8493	6	50	48	0	222	2961	2
44	0.8	1	23	100	9	8	121	1430	1	4678	4266	0	7200	2671	2	29	3	25	358	8176	6	47	46	0	344	2671	2
45	1.6	5	22	100	9	0	116	1430	1	4874	4500	0	8008	1430	1	59	5	50	732	8176	6	49	48	0	176	1677	1
46	0.2	1	22	100	10	10	130	1430	1	4664	4322	0	7729	1573	1	9	1	7	89	7939	6	47	46	0	198	1573	1
47	0.9	3	18	100	10	7	130	1430	1	5169	4820	0	8548	1474	1	33	3	25	502	2946	2	53	52	0	354	1474	1
48	1.4	4	23	99	10	7	126	1430	1	5160	4800	0	8548	1430	1	50	5	39	704	2946	2	52	51	0	244	1430	1
49	1.8	4	22	100	10	6	123	1430	1	5152	4790	0	8580	1430	1	66	7	54	1107	7939	6	52	50	0	213	1430	1
50	2.3	7	20	102	10	0	129	1430	1	4427	4169	0	7320	1430	1	80	8	64	360	7324	6	43	41	0	137	1965	1
51	0.9	3	13	98	10	0	124	1669	1	3960	3679	0	6240	4187	3	34	4	27	290	7367	6	41	38	0	147	7216	5
52	3.8	12	16	103	8	0	122	1430	1	4590	4334	0	7560	1430	1	133	10	113	1252	2491	2	46	44	2	306	2330	2
53	1.0	3	14	108	8	0	130	1504	1	2982	3098	0	5509	1430	1	33	3	27	588	2801	2	28	30	0	138	2467	2
54	0.4	2	11	105	8	0	130	1430	1	2278	2460	0	4680	1431	1	14	1	11	187	2491	2	22	24	0	131	2176	2
55	0.5	1	15	103	6	0	123	1553	1	3638	3470	0	6540	1430	1	18	1	15	300	3069	2	36	35	0	236	2453	2
56	6.1	16	29	105	7	0	120	1430	1	3984	3805	0	7320	1430	1	209	15	182	2737	7925	6	38	39	2	258	2206	2
57	0.8	3	58	104	7	0	124	1430	1	3495	3300	0	6360	1475	1	26	2	22	192	7374	6	34	34	0	163	1535	1
58	4.0	12	108	102	10	0	118	1430	1	3840	3610	0	6660	1430	1	142	14	123	785	7939	6	38	39	2	151	1504	1
59	0.9	3	47	103	10	8	128	5104	4	3105	2880	0	5562	1475	1	33	3	27	504	11147	8	32	31	0	145	5149	4
60	3.4	11	86	96	16	0	128	1435	1	-	-	-	-	132479	100	127	19	95	2030	15789	12	-	-	-	-	132479	100
61	0.8	2	62	105	9	0	128	1487	1	3540	3240	0	6420	1490	1	29	2	23	330	15841	12	36	30	0	225	8263	6
62	7.8	20	168	98	16	0	128	12401	9	3960	3573	0	7056	1475	1	286	35	219	2806	22663	17	46	32	2	397	19217	15
63	1.0	3	32	91	11	0	130	39928	30	3173	2850	0	5700	1490	1	40	4	28	732	41799	32	41	18	0	340	40737	31
64	4.7	11	99	95	11	0	130	31065	23	3540	3070	0	6520	105721	80	180	22	131	2843	35544	27	39	34	0	200	106470	80
65	0.6	2	29	-	-	-	-	132479	100	-	-	-	-	132479	100	-	-	-	-	132479	100	-	-	-	-	132479	100
66	1.8	4	36	98	9	0	125	1430	1	-	-	-	-	132479	100	66	6	52	1289	3769	3	-	-	-	-	132479	100
67	4.5	11	64	99	6	0	123	1472	1	3626	3360	0	6705	1430	1	163	10	139	1789	15194	11	37	36	0	168	2221	2
68	0.9	2	34	101	9	0	124	38039	29	-	-	-	-	132479	100	32	3	26	649	40628	31	-	-	-	-	132479	100
69	0.5	1	19	101	7	0	118	1431	1	-	-	-	-	132479	100	16	1	14	329	15194	11	-	-	-	-	132479	100
70	1.9	6	16	105	8	6	119	1430	1	3859	3696	0	7440	1474	1	66	5	58	1152	2946	2	37	38	0	182	1474	1
71	9.5	27	47	104	9	8	120	1430	1	3900	3735	0	7320	1430	1	330	29	288	4289	13972	11	38	38	0	154	1430	1
72	0.9	3	17	101	8	0	114	1430	1	2647	2526	0	5340	1430	1	31	2	27	386	15784	12	27	27	0	129	1861	1
73	3.4	10	14	104	7	10	117	1430	1	3600	3448	180	6631	1430	1	119	8	106	1210	13972	11	35	35	2	130	1430	1
74	1.1	4	16	106	8	0	130	1464	1	2755	2608	0	5258	1430	1	36	3	30	772	14006	11	26	26	0	103	2054	2
75	7.8	24	55	103	8	0	127	1430	1	3420	3120	0	5989	1430	1	273	18	220	3851	15612	12	34	32	0	167	1868	1
76	1.2	2	100	97	8	0	114	1430	1	2315	2135	0	4440	1430	1	44	4	37	524	15194	11	25	24	0	185	2672	2

Table B.12: Summary statistics of data, computed using data between June 1st 2018 and August 31st 2018. This subset is discussed in section 5.4.6 and is used to fit the self-exciting point process model. For readability, the table has been split across two pages, and headers are included on both pages. Recall table 3.1 and the accompanying discussion for interpretation of these tables.

Appendix C

Appendix to Chapter 4

C.1 Computational aspects of our implementation

Firstly, it is possible, although rare, to find deviations from the typical behaviour contour that are low-density and high-flow. This is atypically good periods for a link, and should not be flagged. From the formulation of our system, such deviations always occur on the top left of the contour. As a result we can check, when a deviation occurs, if we have exited the contour to the left or right, and flag if appropriate. Secondly, the contours are typically made up of a single large component, and occasionally have some minor scattered regions. If these components are negligibly small relative to the total area of all contour components, we can neglect them for computational speed and simplicity. To implement this, we discard any contour components that make up less than 5% of the total area of the entire contours. Numerically, we determine the contours of typical behaviour using the ks package in R [147], and detect if they lie outside the contour using the Matlab function 'inpolygon'. We determine the distance between a data-point and the contour by defining the contour as a high-resolution set of (ρ, f) points and finding the closest point in this set to the data in question. When performing the computations, we have used scaled flow and density data, using the maximum flow we observe and the density at which this occurs for each link respectively.

C.2 The point in polygon problem

A simple way to determine if a point lies in a polygon is to use the 'crossing rule' or 'ray casting', depicted in Fig. C.1. Numerical difficulties can arise if a point is closer to the boundary than numerical precision can detect, but due to secondary thresholding this is not a problem in our application.

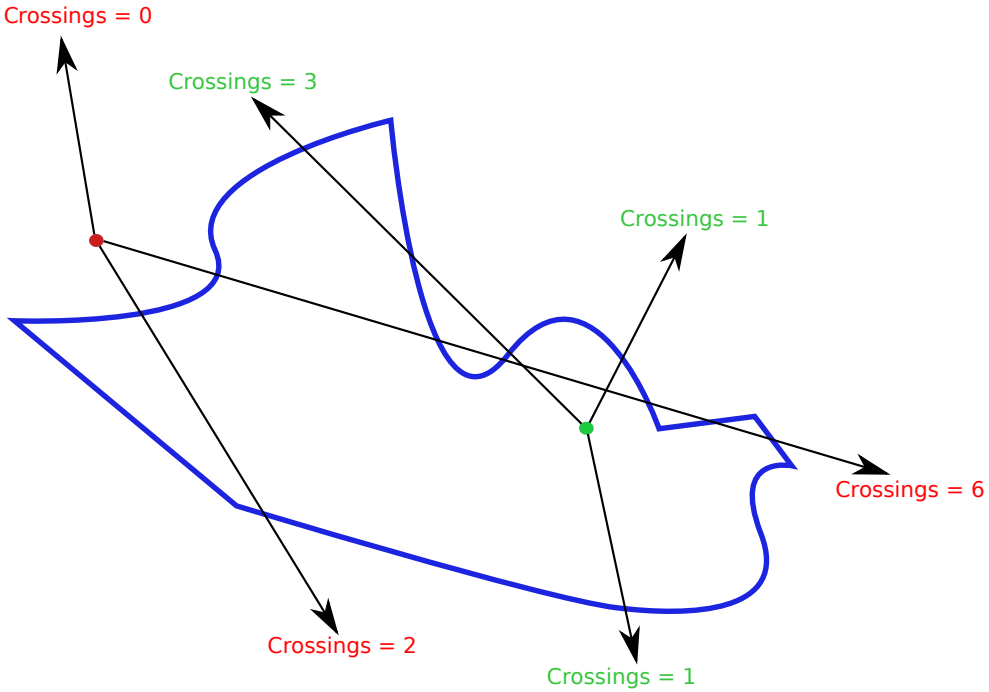


Figure C.1: An example of the crossing rule for the point in polygon problem. We cast rays from the point in question in some direction, and count how many times the ray crosses an edge of the polygon. If the number of crossings is odd, the point lies inside the polygon, otherwise it is outside. We show examples of both here, colouring by red and green to indicate outside and inside respectively. We have used an arbitrary polygon as an example, shown in blue.

C.3 Overview of hypothesis tests

C.3.1 The Wilcoxon signed rank test

The Wilcoxon signed rank test is a non-parametric test used to compare paired observations, considering ranks of variables. Firstly, one assumes that the variable investigated has some natural ordering, the data is recorded in matched pairs and that each pair is chosen randomly and independently. The null hypothesis of this test is that the median difference between pairs is 0, following a symmetric distribution. To test this, we first compute the absolute difference between paired data-points, and discard any that are equal in value. We then order the remaining data-pairs by the absolute magnitude of their difference. The test statistic is computed as

$$W = \sum_{i=1}^{N} \left[sign(x_{2,i} - x_{1,i}) \cdot R_i \right], \tag{C.1}$$

with R_i being the rank of the data-pair, and $sign(x)$ being -1 if $x < 0$, 0 if $x = 0$ and 1 otherwise. The value of W can then be compared to standard statistical

tables, and a p-value obtained. Since this test uses ranks, it is resilient to outliers. However, the distributions of the differences between data-pairs may not actually be symmetric in practice.

C.3.2 The paired sign test

Another non-parametric test used to compare paired observations is the sign test. To conduct this, we first assume that the differences between data-points are independent, each difference comes from the same continuous population and data has a natural ordering. We collect independent pairs of sample data, and then initially assume that given any pair $x_{1,i}, x_{2,i}$, $\mathbb{P}(x_{1,i} < x_{2,i}) = 0.5$. This means for any measurement pair, it is equally likely that one groups observation is larger than the other groups. If we denote the number of pairs where $x_{2,i} - x_{1,i} > 0$ as S, then under the null hypothesis S follows a binomial distribution with N trials and probability of success 0.5. Again, standard statistical tables can be used to determine a p-value given the measured test statistic.

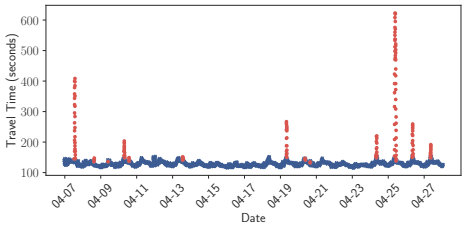
C.4 Sensitivity to threshold choices

The main question one may have in applying this method is what choice of α to make, meaning how much mass we want inside the contour of typical behaviour. This choice will directly impact the shape of the contour, and in particular it's area, meaning we will raise more incidents the larger it is made. In Fig. C.2, we quantify this by comparing the DFTB flags raised for various values of α, and no thresholding on time-quantiles applied.

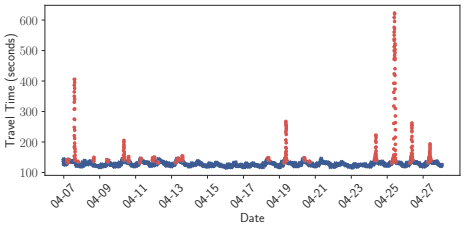
From Fig. C.2, we see a clear progression of raising more DFTB incidents as we increase α from 0.02 to 0.1. For the all cases we see the significant spikes in the series have DFTB incidents raised, however as we go from $\alpha = 0.02$ to $\alpha = 0.1$ more of the recurrent parts of the series are highlighted as deviations from typical behaviour. The use of secondary thresholding based on severity means that making some base choice of $\alpha = 0.05$ is reasonable and does not need to be considered from an operational perspective.

C.5 Loop level anomaly detection

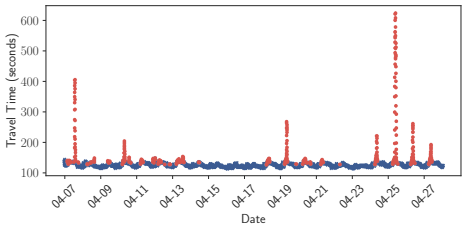
Whilst our incidents location data is only given at the link level and hence has focused our detection work on this scale of data, one could equally perform the same analysis on loop level data with no changes. To demonstrate that this is true, we construct the contour of typical behaviour at the link level, then again at the loop level for loops on the same link. Our results for doing this



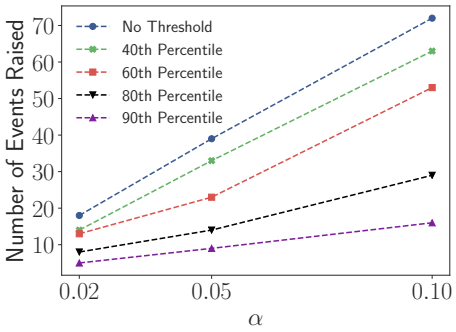
(a) $\alpha = 0.02$.



(b) $\alpha = 0.05$.

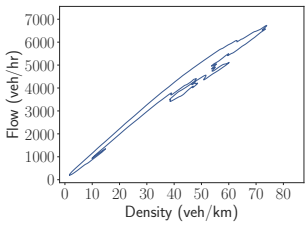


(c) $\alpha = 0.10$.

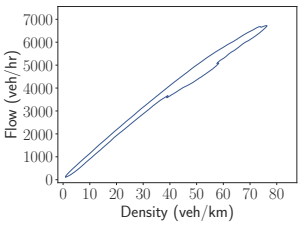


(d) Number of events varying α and temporal threshold.

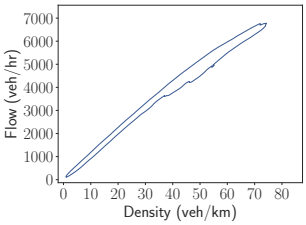
Figure C.2: Comparisons of the incident detection method using various α values to fit the contour, all on the same 3-week set of data. In (a), (b) and (c) we plot the travel time series as before, coloured and marked by the presence of a DFTB incident. In (d) we count the number of incidents raised in this 3 week window at various thresholds as a function of α.



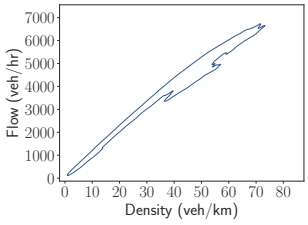
(a) Link Level.



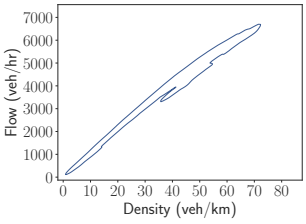
(b) Loop 1.



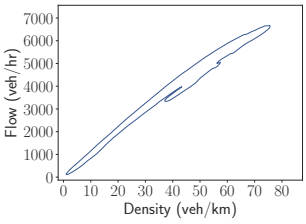
(c) Loop 2.



(d) Loop 3.



(e) Loop 4.

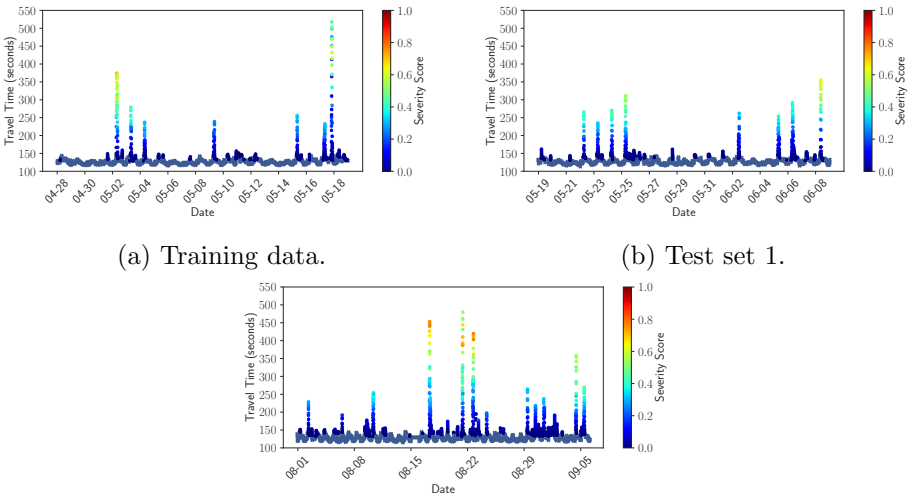


(f) Loop 5.

Figure C.3: Contour plots of 3-week segmented data. We see that the method is entirely applicable to loop and link data, and if suitable data is collected one can validate it at a higher spatial resolution than we have considered.

are shown in Fig. C.3. Using the loop level implementation, one can identify where is atypical on the network at a higher resolution than the link level.

C.6 Further example plots



(a) Training data.

(b) Test set 1.

(c) Test set 2.

Figure C.4: Severity examples on another link.

Appendix D

Appendix to Chapter 5

D.1 Derivation of weekly periodic background component

To begin, we first denote

$$w^{(w)}(t,x) = \frac{\mu_w(t)\mu_s(x)}{\lambda(t,x)}, \tag{D.1}$$

and substitute this into Eq. (5.5), giving

$$\begin{aligned}
&\sum_{i=1}^{N} w^{(w)}(t_i,x_i)\mathbb{1}_{t_i\in\cup_{k\in\mathbb{Z}}[t+km_w-\Delta t,t+km_w+\Delta t]} \\
&\approx \int_0^T\int_0^X w^{(w)}(\tau,\chi)\lambda(\tau,\chi)\mathbb{1}_{\tau\in\cup_{k\in\mathbb{Z}}[t+km_w-\Delta t,t+km_w+\Delta t]}d\chi d\tau \\
&= \int_0^T\int_0^X \frac{\mu_w(\tau)\mu_s(\chi)}{\lambda(\tau,\chi)}\lambda(\tau,\chi)\mathbb{1}_{\tau\in\cup_{k\in\mathbb{Z}}[t+km_w-\Delta t,t+km_w+\Delta t]}d\chi d\tau \\
&= \left[\int_0^T \mu_w(\tau)\mathbb{1}_{\tau\in\cup_{k\in\mathbb{Z}}[t+km_w-\Delta t,t+km_w+\Delta t]}d\tau\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&= \left[\sum_{\gamma=0}^{\lfloor\frac{T}{m_w}\rfloor}\int_0^{m_w} \mu_w(\tau)\mathbb{1}_{\tau\in[t-\Delta t,t+\Delta t]}d\tau\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&= \left[\sum_{\gamma=0}^{\lfloor\frac{T}{m_w}\rfloor}\int_{t-\Delta t}^{t+\Delta t} \mu_w(\tau)d\tau\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&\approx \left[\sum_{\gamma=0}^{\lfloor\frac{T}{m_w}\rfloor} 2\Delta t\mu_w(t)\right]\left[\int_0^X \mu_s(\chi)d\chi\right] \\
&\propto \mu_w(t).
\end{aligned} \tag{D.2}$$

Similarly to the daily case, we have the indicator function

$$\mathbb{1}_{t_i \in \cup_{k\in\mathbb{Z}}[t+km_w-\Delta t, t+km_w+\Delta t]}$$

incorporating a data-point into the estimate if it lies in some interval $[t-\Delta t, t+\Delta t]$, on any week in the dataset, where $t \in [0, m_w)$ and Δt is a small positive value. From Eq. (D.2), it follows that

$$\hat{\mu}_w(t) \propto \sum_{i=1}^{N} w_i^{(w)} \mathbb{1}_{t_i \in \cup_{k\in\mathbb{Z}}[t+km_w-\Delta t, t+km_w+\Delta t]}, \tag{D.3}$$

$$w_i^{(w)} = \frac{\mu_w(t_i)\mu_s(x_i)}{\lambda(t_i, x_i)}. \tag{D.4}$$

We can simply this expression by introducing $\tilde{t}_i = t_i - m_w \left\lfloor \frac{t_i}{m_w} \right\rfloor$, $\forall i \in \{1, 2, \ldots, N\}$, mapping the raw incident times from the domain $[0, T]$ onto the domain $[0, m_w)$. Eq. (D.3) then becomes

$$\hat{\mu}_w(t) \propto \sum_{i=1}^{N} w_i^{(w)} \mathbb{1}_{\tilde{t}_i \in [t-\Delta t, t+\Delta t]}. \tag{D.5}$$

Applying smoothing, we attain

$$\hat{\mu}_w(t) \propto \sum_{i=1}^{N} w_i^{(w)} \frac{k_{\omega_w}(t-\tilde{t}_i)}{\int_0^{m_w} k_{\omega_w}(\tau-\tilde{t}_i)d\tau} \tag{D.6}$$

where ω_w is the smoothing bandwidth specific to the weekly background component. One can incorporate potential non-negligible contributions from points either side of the domain, writing $k^{\text{lower}}_{\text{weekly}} = k_{\omega_w}\left(t - \left[t_i - m_w \left\lfloor \frac{t_i}{m_w} \right\rfloor - m_w\right]\right)$ and $k^{\text{upper}}_{\text{weekly}} = k_{\omega_w}\left(t - \left[t_i - m_w \left\lfloor \frac{t_i}{m_w} \right\rfloor + m_w\right]\right)$, finally yielding

$$\hat{\mu}_w(t) \propto \sum_{i=1}^{N} w_i^{(w)} \frac{k^{\text{lower}}_{\text{weekly}} + k_{\omega_w}\left(t - \left[t_i - m_w \left\lfloor \frac{t_i}{m_w} \right\rfloor\right]\right) + k^{\text{upper}}_{\text{weekly}}}{\int_{-m_w}^{2m_w} k_{\omega_w}(\tau-\tilde{t}_i)d\tau}. \tag{D.7}$$

D.2 Derivation of spatial triggering component

To determine the spatial triggering function, we again consider two data points $\left(\tau^{(1)}, \chi^{(1)}\right)$ and $\left(\tau^{(2)}, \chi^{(2)}\right)$. We then define $\rho(\tau^{(1)}, \chi^{(1)}, \tau^{(2)}, \chi^{(2)})$ as

$$\rho\left(\tau^{(1)}, \chi^{(1)}, \tau^{(2)}, \chi^{(2)}\right) = \begin{cases} \frac{Ag(\tau^{(2)}-\tau^{(1)})h(\chi^{(2)}-\chi^{(1)})}{\lambda(\tau^{(2)}, \chi^{(2)})}, & \text{if } \tau^{(1)} < \tau^{(2)} \text{and } \chi^{(1)} > \chi^{(2)}. \\ 0, & \text{otherwise} \end{cases} \tag{D.8}$$

If we apply this to Eq. (5.5), letting $f(\tau,\chi) = \rho(t_i, x_i, \tau, \chi)\mathbb{1}_{\chi - x_i \in [x-\Delta x, x+\Delta x]}$, then we attain

$$\begin{aligned}
&\sum_j \rho(t_i, x_i, \tau_j, \chi_j)\mathbb{1}_{\chi_j - x_i \in [x-\Delta x, x+\Delta x]} \\
&\approx \int_0^T \int_0^X \rho(t_i, x_i, \tau, \chi)\mathbb{1}_{\chi - x_i \in [x-\Delta x, x+\Delta x]}\lambda(\tau,\chi)d\chi d\tau \\
&= A\int_{t_i}^T \int_0^{x_i} \frac{g(\tau - t_i)h(\chi - x_i)}{\lambda(\tau,\chi)}\mathbb{1}_{\chi - x_i \in [x-\Delta x, x+\Delta x]}\lambda(\tau,\chi)d\chi d\tau \\
&= A\left[\int_{t_i}^T g(\tau - t_i)d\tau\right]\left[\int_0^{x_i} h(\chi - x_i)\mathbb{1}_{\chi - x_i \in [x-\Delta x, x+\Delta x]}d\chi\right].
\end{aligned} \tag{D.9}$$

If we then let $s = \chi - x_i$ we have

$$\begin{aligned}
&\sum_j \rho(t_i, x_i, \tau_j, \chi_j)\mathbb{1}_{\chi_j - x_i \in [x-\Delta x, x+\Delta x]} \\
&\approx A\left[\int_0^T g(\tau)d\tau\right]\left[\int_{-x_i}^0 h(s)\mathbb{1}_{s \in [x-\Delta x, x+\Delta x]}ds\right] \\
&= A\left[\int_{t_i}^T g(\tau)d\tau\right]\left[\int_{x-\Delta x}^{x+\Delta x} h(s)ds\right] \\
&\approx A\left[\int_{t_i}^T g(\tau)d\tau\right] 2\Delta x h(x) \\
&\propto h(x).
\end{aligned} \tag{D.10}$$

As a result

$$h(x) \propto \sum_i \sum_j \rho(t_i, x_i, t_j, x_j)\mathbb{1}_{x_j - x_i \in [x-\Delta x, x+\Delta x]} \tag{D.11}$$

and hence

$$\hat{h}(x) \propto \sum_{(i,j)} \rho_{i,j}\mathbb{1}_{x_j - x_i \in [x-\Delta x, x+\Delta x]} \tag{D.12}$$

with

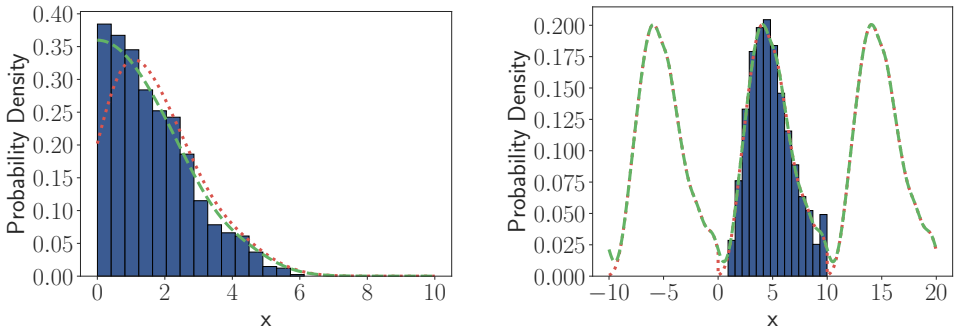
$$\rho_{i,j} = \frac{Ag(t_j - t_i)h(x_j - x_i)}{\lambda(t_j, x_j)} \;\forall\,(i,j) \text{ s.t. } t_i < t_j \text{ and } x_i > x_j. \tag{D.13}$$

As before, the estimator in Eq. (5.29) can be smoothed to give

$$\hat{h}(x) \propto \frac{\sum\limits_{(i,j)} \frac{\rho_{i,j} k_{\omega_h}(x - x_j + x_i)}{\int_0^{x_i} k_{\omega_h}(\chi - x_j + x_i)d\chi}}{\sum\limits_{i=1}^N \mathbb{1}_{x_i + x \geq 0}}. \tag{D.14}$$

D.3 Examples of boundary correction

In chapter 5, we discussed methods of boundary correction. Specifically, we use mirrored correction when the domain is truncated, and for periodic domains we include influence from points lying one period either side of the domain in question. In Fig. D.1, we show examples with and without these corrections using simulated data. In Fig. D.1a, we show a data-set truncated at 0, and fit two density estimates to it, one with and one without mirrored boundary correction. In Fig. D.1b, we show a dataset generated on the domain $[0, 10]$, and imagine that we are trying to construct a periodic function from it. Inspecting



(a) Mirrored correction example, using data generated from a half normal distribution with mean 0, standard deviation 2. Each density estimate is fit with a bandwidth of 0.75.

(b) Periodic correction example, using data generated from a gamma distribution with shape 5 and scale 1. Each density estimate is fit with a bandwidth of 0.5.

Figure D.1: Examples of boundary correction influence on simulated data. We show the data as a histogram, the non-corrected fits with (····) and the boundary corrected fits with (- - -).

Fig. D.1a, we see a clear drop in the estimate at the truncation point of 0 without correction. Additionally, from Fig. D.1b, we see a drop at the two ends of the periodic domain without any correction. These are removed when we apply our correction methods.

Appendix E

Appendix to Chapter 6

E.1 Distribution information

In table E.1 we summarise the discussed probability distributions.

Distribution	Parameters	PDF
Log-normal	$\mu \in (-\infty, \infty)$ $\sigma \in (0, \infty)$	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log(x)-\mu)^2}{2\sigma^2}}$
Log-logistic	$\alpha \in (0, \infty)$ $\beta \in (0, \infty)$	$\frac{\left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{\beta}\right)^2}$
Weibull	$\lambda \in (0, \infty)$ $k \in (0, \infty)$	$\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$
Generalised F	$\beta \in (-\infty, \infty)$ $\sigma \in (0, \infty)$ $m_1 \in (0, \infty)$ $m_2 \in (0, \infty)$	Let: $P = \frac{\frac{m_1 e^{\omega}}{m_2}}{1+\frac{m_1}{m_2}e^{\omega}}, \omega = \frac{\log(x)-\beta}{\sigma}$. Then: $\frac{1}{\sigma x B(m_1,m_2)} P^{m_1}(1-P)^{m_2}$
Mixture	$\boldsymbol{\theta}_i \forall i \in \{1, \ldots, k\}$ $\pi_i \forall i \in \{1, \ldots, k\}$	$\sum_{i=1}^{k} \pi_i \tilde{f}_i(x; \boldsymbol{\theta}_i)$

Table E.1: Summary of distributions considered throughout this work. All distributions are defined for positive x. Log-normal: Can be seen as assuming the logarithm of x is normally distributed. Generalised F: $B(m_1, m_2)$ represents the beta function. Mixture: Each mixture component has some PDF $\tilde{f}_i(x; \boldsymbol{\theta}_i)$, that depends on some parameter set $\boldsymbol{\theta}_i$, and a component weight π_i. The weighted sum of all components generates the final distribution.

E.2 Incident visualisation

Clear structure was observed incident time series in section 6.3. We show more examples in Fig. E.1.

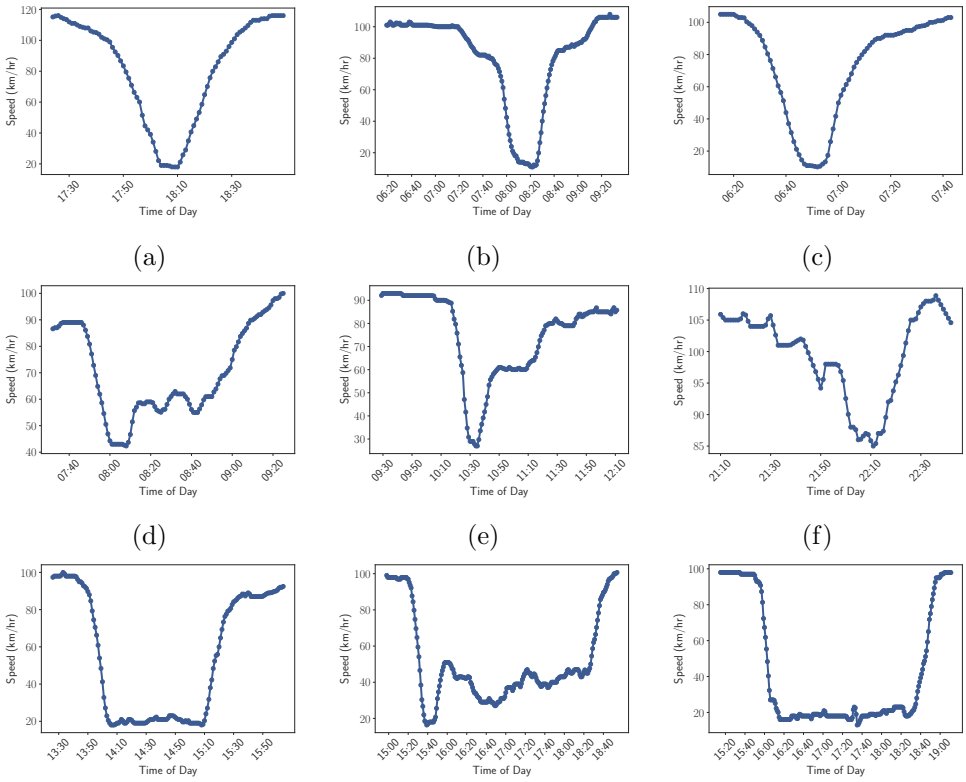


Figure E.1: Example speed time series from NTIS incidents.

E.3 Clustering analysis

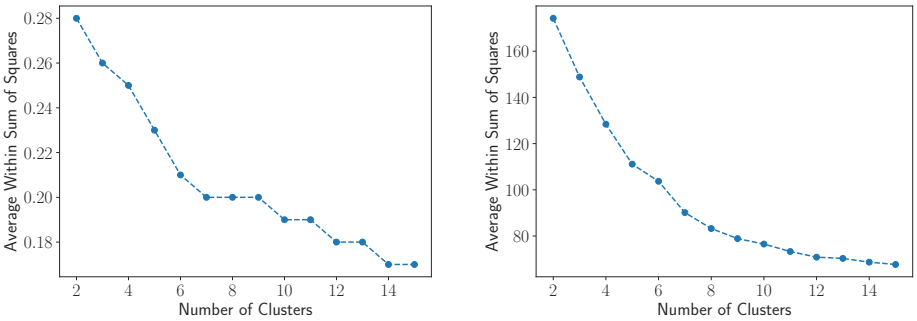
When clustering the data by their time-invariant features, we use the 'daisy' dissimilarity measure [208] as it is able to handle features of varying types using the Gower dissimilarity coefficient [209]. Inspecting the average within cluster sum of squares distance values on both clustering results, we construct 'elbow plots' shown in Fig. E.2. From Fig. E.2 we see diminishing reduction in the within cluster sum of squared distances after around 6 clusters in each case.

E.4 Models summary

In table E.2, we summarise the models considered, chosen to reflect existing work in the transportation literature and recent advancements in other disciplines

E.5 Neural network outputs

We consider a mixture of log-normal distributions as a model of incident durations, and parametrise this using a neural network. As a result, we have



(a) Time-invariant features.

(b) Time-varying features.

Figure E.2: Plots comparing the within cluster sum of squared distances when clustering on time-invariant and time-varying features. Distance measures used are very different, so the numerical values on the y-axis should not be compared.

Model	Distributional Assumptions	Complexity	Notes
AFT (LN)	Log-normal	Linear	Accelerated failure time model assuming a log-normal duration distribution, used in previous traffic study [46].
AFT (W)	Weibull	Linear	Accelerated failure time model assuming a Weibull duration distribution, used in previous traffic studies [55] and [50].
Cox	Non-Parametric	Linear	Cox regression model, baseline hazard is non-parametric, covariate effects act as an exponential of a linear regression. Used in previous traffic studies [60] and [61].
RSF	Non-Parametric	Non-Linear	Random survival forest model, using an ensemble of 1000 trees. Used in various other applications such as [71] and [72].
NN (NP)	Non-Parametric	Non-Linear	Neural network model with a non-parametric output distribution Used in healthcare [54].
NN (LN)	Mixture of log-normal	Non-Linear	Neural network model parametrising a distribution consisting of a mixture of log-normal components. Mixture used due to work in [52] suggesting these may offer predictive power for traffic incident datasets.
NN (Kernel)	Non-Parametric with smoothing	Non-Linear	Neural network model with a kernel smoothed output distribution, compromising between the strong assumption of NN (LN) and the complete freedom of NN (NP).
SW (NP)	Non-Parametric	Non-Linear	Neural network model with a sliding window CNN architecture to capture time-varying features . Output is non-parametric as in NN (NP). Used in healthcare applications [81].
SW (LN)	Mixture of log-normal	Non-Linear	Neural network model as in SW (NP), with an output layer specifying a mixture of log-normals as in NN (LN). Combines ideas in [81] with mixture observations in [52].
SW (Kernel)	Non-Parametric with smoothing	Non-Linear	Neural network model as in SW (NP), with an output layer specifying a kernel smoothed distribution. This compromises between the strong assumption of SW (LN) and the complete freedom of SW (NP).

Table E.2: Summary of models considered with key points and origins highlighted.

restrictions on the network output for certain nodes, which are accounted for using activation functions. For each mixture component, the network outputs a mixture weight, and parameters representing the mean and standard deviation of the underlying normal distribution, consistent with the implementation of the log-normal distribution in Tensorflow. A softmax activation is applied to the mixture weights to ensure their validity. A linear activation is applied to the network outputs that correspond to the underlying means, and a 'softplus' activation applied to the outputs that correspond to the underlying standard deviations, written as

$$SP(x) = \log(e^x + 1). \tag{E.1}$$

E.6 Hyper-parameter values

In table E.3 we detail the hyper-parameters considered. We apply early stopping, learning rate decay and use a batch size of 128. Models are implemented in Keras with the Adam optimiser.

Hyper-parameter	Values Considered
Dropout	0.5
Kernel size	{5, 10}
L_1, L_2 regularisation	10^{-4}
Learning rate	{ 10^{-4}, 10^{-2} }
Number of dense layers	{1, 2, 3}
Number of neurons per dense layer	{32, 64, 128, 256}
Number of mixtures	{1, 2, 3}
Number of CNN layers	{1, 2, 3}
Number of filters per layer	{4, 8, 16}
Smoothing bandwidth	3 (minutes)
Time series window size	{30, 60} (minutes)
η_σ	{0.1, 1.0}

Table E.3: Hyper-parameters considered when fitting neural network models.

E.7 Comparison of temporal convolutions to manually engineered time series features

In table E.4 we show the C-index and Brier score comparing a sliding window CNN model to one that uses the series levels and gradients. The later is denoted 'Raw (NP)'. MAPE results are shown in table E.5.

Prediction Time (minutes)	Metric	Model	Prediction Horizon (minutes)								Mean Over Horizons
			5	15	30	45	60	120	180	240	
$t = 0$	C-Index	SW (NP)	-	**0.798**	**0.743**	**0.705**	**0.682**	**0.642**	**0.637**	**0.634**	**0.692**
		Raw (NP)	-	0.755	0.701	0.669	0.646	0.599	0.592	0.588	0.650
	Brier	SW (NP)	**0.007**	**0.082**	**0.160**	**0.213**	**0.262**	**0.290**	**0.195**	**0.113**	**0.165**
		Raw (NP)	0.015	0.122	0.208	0.260	0.306	0.310	0.203	0.116	0.193
$t = 15$	C-Index	SW (NP)	**0.947**	**0.884**	**0.811**	**0.772**	**0.731**	**0.678**	**0.669**	**0.666**	**0.770**
		Raw (NP)	0.946	0.849	0.773	0.729	0.689	0.637	0.624	0.620	0.733
	Brier	SW (NP)	**0.019**	**0.069**	**0.134**	**0.180**	**0.230**	**0.259**	**0.170**	**0.096**	**0.145**
		Raw (NP)	0.022	0.091	0.163	0.210	0.257	0.272	0.175	0.098	0.161
$t = 30$	C-Index	SW (NP)	**0.960**	**0.905**	**0.803**	**0.761**	**0.733**	**0.689**	**0.681**	**0.677**	**0.776**
		Raw (NP)	0.948	0.870	0.773	0.730	0.702	0.655	0.645	0.641	0.746
	Brier	SW (NP)	**0.018**	**0.067**	**0.138**	**0.183**	**0.223**	**0.236**	**0.155**	**0.088**	**0.139**
		Raw (NP)	0.021	0.083	0.158	0.203	0.241	0.248	0.160	0.090	0.151
$t = 45$	C-Index	SW (NP)	**0.967**	**0.880**	**0.813**	**0.784**	**0.749**	**0.703**	**0.692**	**0.688**	**0.785**
		Raw (NP)	0.953	0.851	0.777	0.748	0.715	0.667	0.654	0.651	0.752
	Brier	SW (NP)	**0.019**	**0.074**	**0.138**	**0.176**	**0.217**	**0.221**	**0.148**	**0.089**	**0.135**
		Raw (NP)	0.021	0.087	0.158	0.196	0.234	0.230	0.152	0.090	0.146
$t = 60$	C-Index	SW (NP)	**0.953**	**0.903**	**0.830**	**0.787**	**0.755**	**0.711**	**0.702**	**0.699**	**0.793**
		Raw (NP)	0.938	0.870	0.799	0.763	0.729	0.679	0.667	0.665	0.764
	Brier	SW (NP)	**0.023**	**0.073**	**0.136**	**0.176**	**0.212**	**0.210**	**0.139**	**0.089**	**0.132**
		Raw (NP)	0.025	0.087	0.153	0.192	0.228	0.221	0.144	0.091	0.143
$t = 120$	C-Index	SW (NP)	**0.968**	**0.896**	**0.852**	**0.824**	**0.791**	**0.744**	**0.739**	**0.735**	**0.819**
		Raw (NP)	0.958	0.847	0.795	0.773	0.745	0.697	0.692	0.689	0.775
	Brier	SW (NP)	**0.023**	**0.082**	**0.128**	**0.157**	**0.189**	**0.183**	**0.130**	**0.089**	**0.123**
		Raw (NP)	0.026	0.101	0.159	0.180	0.207	0.194	0.135	0.091	0.137

Table E.4: Results comparing a sliding window model and a model where we input the most recent time series values and gradients. The sliding window model achieves better C-Index and Brier values across all considered prediction time and horizon pairs.

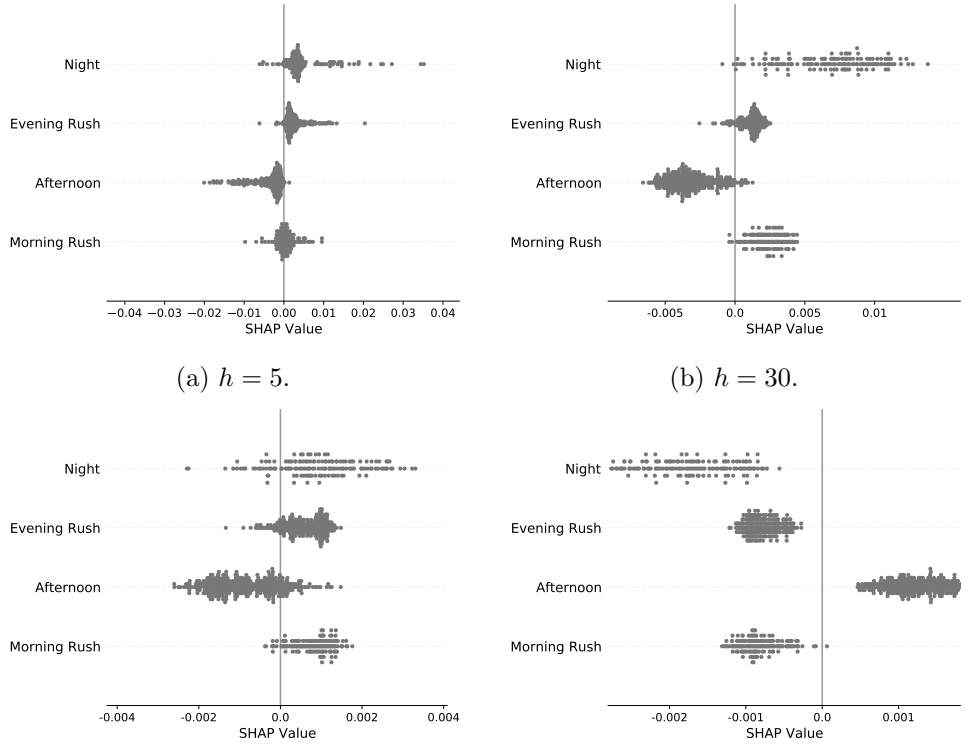
Model	MAPE Dynamic Model			
	Percentile Into Incident Prediction Made at			
	30th	50th	70th	90th
SW (NP)	**31.660**	**21.998**	17.069	10.399
Raw (NP)	33.922	23.140	**16.762**	**10.134**

Table E.5: Results comparing a sliding window model and a model where we input the most recent time series values and the gradients. The sliding window model achieves better MAPE at the 30th and 50th percentiles, but marginally worse at the 70th and 90th percentiles.

E.8 SHAP values for categorical features

In section 6.6.2, we considered SHAP values as a measure for variable importance and gave example interpretations of why the neural network model was outputting particular predictions. Here we plot further examples of what the categorical values contribute overall to the model. In Fig. E.3, we show the results for the time of day. Generally, we see that the impact of the time of day variable attaining the value 'afternoon' decreases the model output at short horizons (5, 30 minutes in Figs. E.3a, E.3b). At a horizon of 60 minutes (Fig. E.3c) the impact is mixed, and at a horizon of 180 minutes (Fig. E.3d) the model output is increased. From this we might be lead to believe that incidents occurring in the afternoon can spill over into the evening rush period and last a significant amount of time.

Impact of incident type is shown in Fig. E.4. We first note that the NTIS



(a) $h = 5$.

(b) $h = 30$.

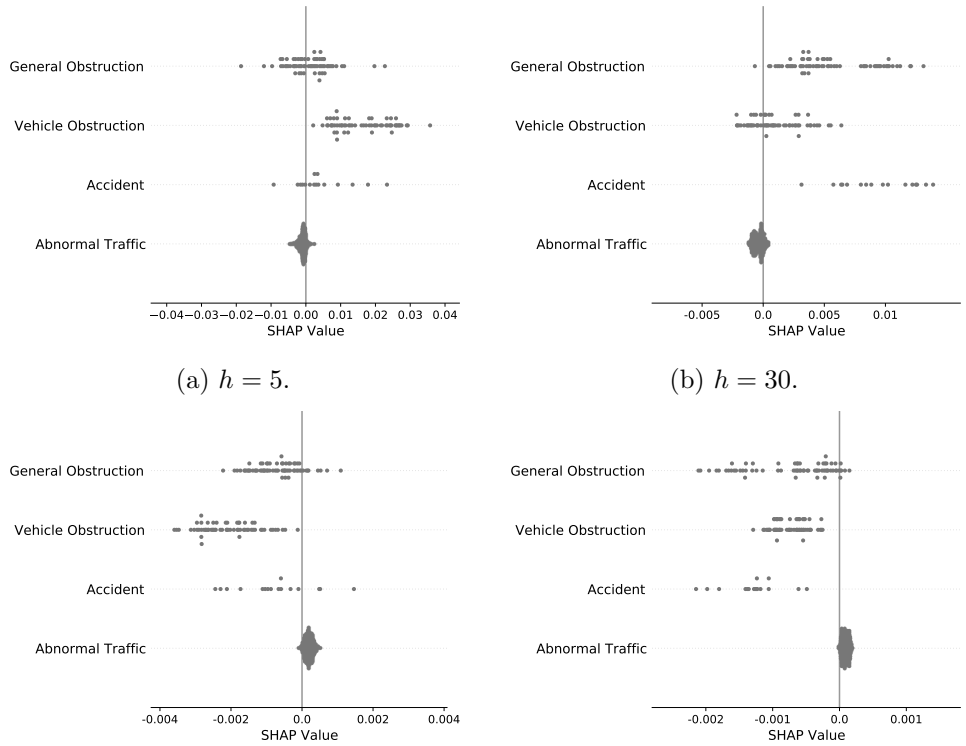
(c) $h = 60$.

(d) $h = 180$.

Figure E.3: SHAP values for the time of day feature. In each plot, we have summed the SHAP values for each one-hot encoded value, and then only plotted the resulting value in the row corresponding to each data-points true feature value.

incidents data is dominated by abnormal traffic incidents, which is clearly seen by the random sample of 1000 incidents for explanation containing a large number of abnormal traffic cases. When we inspect the data, the abnormal traffic cases are quite heavy tailed, suggesting either they are not always turned off by the operator in a timely manner, or there are some significant drops in link performance that are not directly attributed to a physical incident on the link. We see that general obstructions increase the model output at horizons of 30 minutes (Fig. E.4b) but decrease it after this. Abnormal traffic incidents typically increase the model output at higher horizons and decrease it at lower ones, however the time series and other information could then account for some of the short-lived abnormal traffic incidents we observe.

The seasons impact on the model are shown in Fig. E.5. Note that with one year of data, the season variable could capture some long term change in response time or incident severity. However, we are not told to expect this from industry experts. When we consider incidents in the winter, we see the output is mixed at a 5 minute horizon (Fig. E.5a), increased at a horizon of 30 minutes (Fig. E.5b) and decreased at the later horizons. The effect of spring



(a) $h = 5$. (b) $h = 30$.

(c) $h = 60$. (d) $h = 180$.

Figure E.4: SHAP values for the incident type feature. In each plot, we have summed the SHAP values for each one-hot encoded value, and then only plotted the resulting value in the row corresponding to each data-points true feature value.

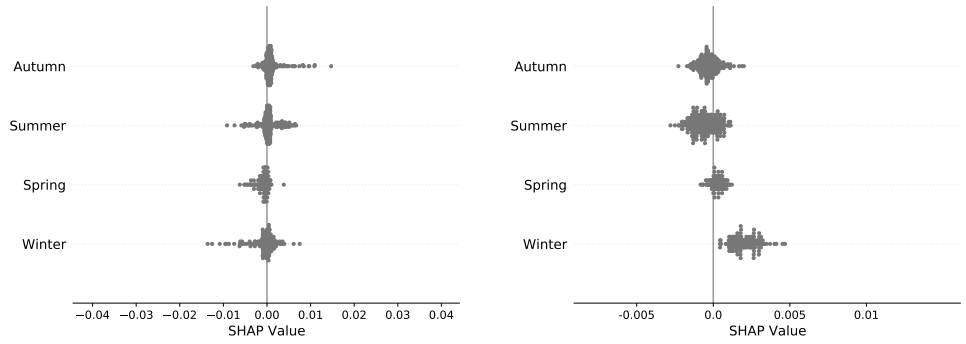
appears to be systematically smaller in size than the other seasons.

E.9 AUROC validation

Whilst we have discussed C-index and Brier score for different prediction time, prediction horizon pairs as was done in [180], we can equally look at the area under the receiver operator curve (AUROC) as a performance metric, as in [81]. We show results for this, at particular horizons, averaged over all input windows, in table E.6, computed using the methods discussed in [210] and [211].

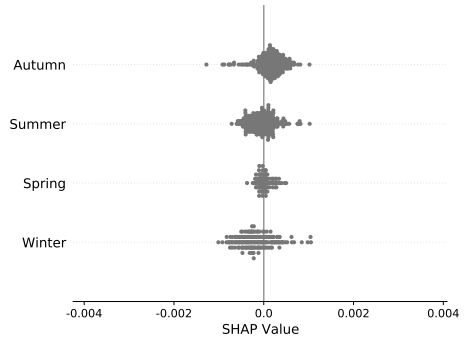
E.10 MAPE as a function of time

We have looked at the error at different percentiles into an incident because of the national criteria set by Highways England. However, we note that some other works consider error at different minutes into an incident. These are two different methods of aggregation. The first averages across different

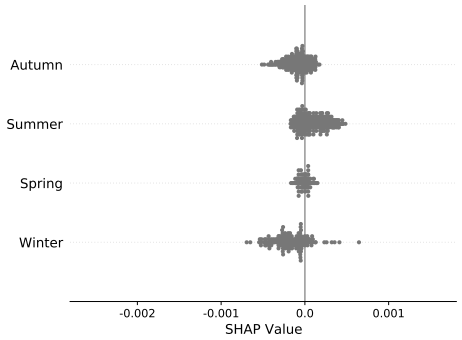


(a) $h = 5$.

(b) $h = 30$.



(c) $h = 60$.



(d) $h = 180$.

Figure E.5: SHAP values for the season feature. In each plot, we have summed the SHAP values for each one-hot encoded value, and then only plotted the resulting value in the row corresponding to each data-points true feature value.

incidents, taking predictions from potentially very different time points, but that represent the same fraction of the entire incident. The second averages across the same prediction time for every incident, but disregards the fact that this might be a very short time relative to the total duration of some incidents, and a very long time relative to others. We present results for the second averaging here to offer a clearer comparison to existing work. The results in Fig. E.6 are qualitatively similar to those in figure 2 of [83].

Model	Prediction Horizon (minutes)							Mean Over Horizons
	15	30	45	60	120	180	240	
Cox	0.693	0.716	0.655	0.615	0.558	0.557	0.540	0.619
RSF	0.805	0.745	0.692	0.653	0.582	0.585	**0.572**	0.662
SW (LN)	0.857	0.798	0.759	0.714	0.622	0.580	0.552	0.697
SW (NP)	0.862	0.808	0.771	0.731	**0.642**	**0.602**	0.568	0.712
SW (Kernel)	**0.878**	**0.818**	**0.780**	**0.735**	0.640	0.594	0.558	**0.715**
SW (Raw)	0.877	0.809	0.764	0.712	0.613	0.571	0.544	0.698

Table E.6: AUROC values for the dynamic models. Larger values indicate a better model. We see similar conclusions are present here that were made when inspecting the time-varying C-index values.

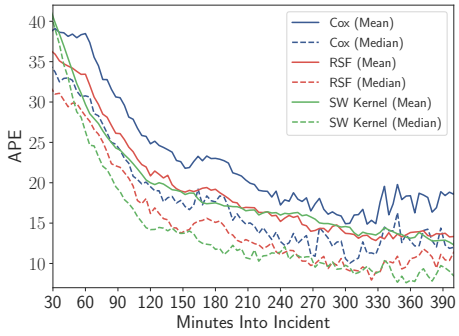


Figure E.6: APE (both mean and median shown) for some dynamic models as a function of the minutes into an incident. We show 30 minutes onwards, which is the point where only incident related information is being fed into the sliding window model, and the landmarking models are no longer fit with minor incidents, as these have already ended.

Appendix F

Appendix to Chapter 7

F.1 Emergent behaviours from a simplified model

Here we present results showing that the emergent behaviours we have observed are not dependent on the vehicle model, rather the problem formulation. A simple kinematic bicycle model is

$$\begin{aligned} \frac{dx(t)}{dt} &= v_x(t)\cos(\phi(t)) \\ \frac{dy(t)}{dt} &= v_x(t)\sin(\phi(t)) \\ \frac{d\phi(t)}{dt} &= \frac{v_x(t)}{L_V}\tan(\delta(t)) \\ \frac{dv_x(t)}{dt} &= a_x(t). \end{aligned} \tag{F.1}$$

and assumes we directly control the acceleration a_x and steering angle δ. This was used in the Apollo path planning framework in [194].

F.1.1 Vehicle following

Vehicle following behaviour is shown in Fig. F.1.

F.1.2 Vehicle passing

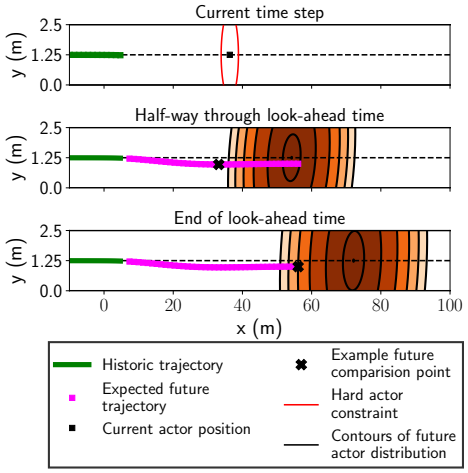
Vehicle passing behaviour is shown in Fig. F.2.

F.1.3 Vehicle overtaking

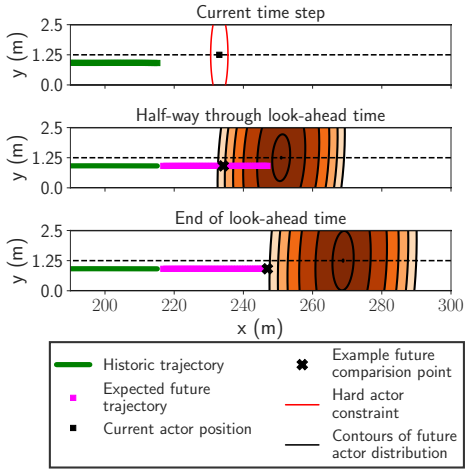
Vehicle overtaking behaviour is shown in Fig. F.3.

F.1.4 Multi-actor vehicle overtaking

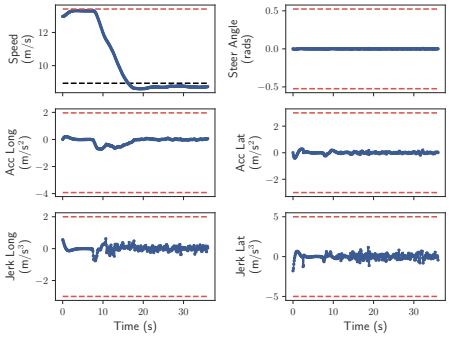
Vehicle overtaking behaviour with multiple actors is shown in Fig. F.3.



(a) Example trajectory 8 seconds into the simulation.



(b) Example trajectory 30 seconds into the simulation.

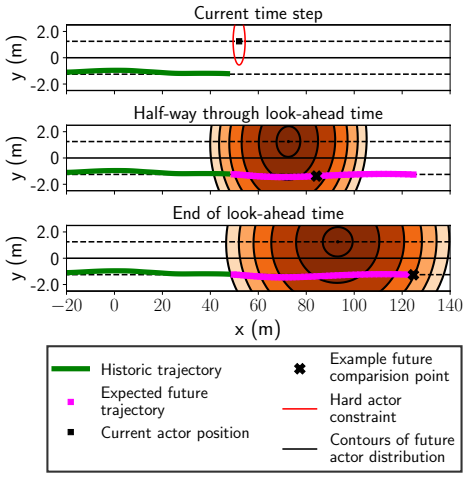


(c) Constraints summary for a vehicle following example.

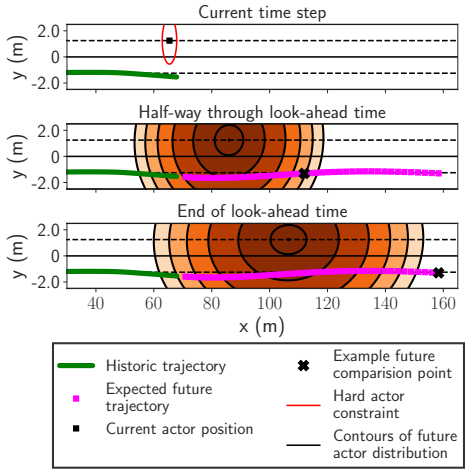
Figure F.1: Optimisation result for a vehicle following example.

F.1.5 Abnormal actor behaviour

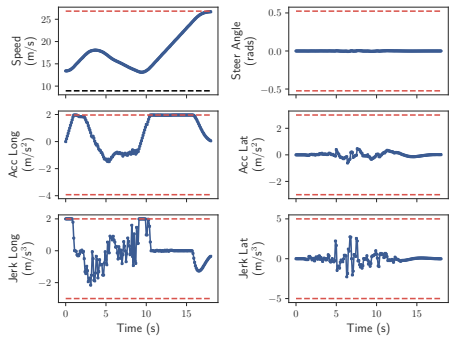
The model response to an abnormal actor braking is shown in Fig. F.5.



(a) Example trajectory 9.5 seconds into the simulation.

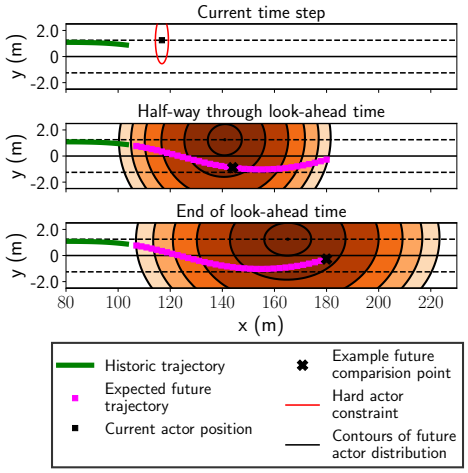


(b) Example trajectory 11 seconds into the simulation.

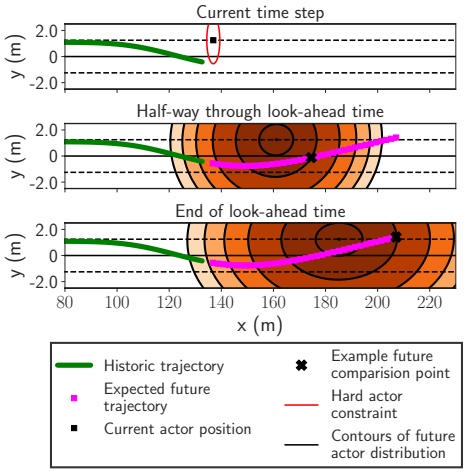


(c) Constraints summary for a vehicle passing example.

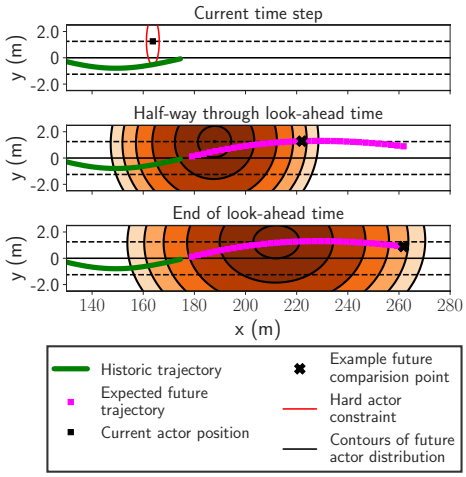
Figure F.2: Optimisation result for a vehicle passing example.



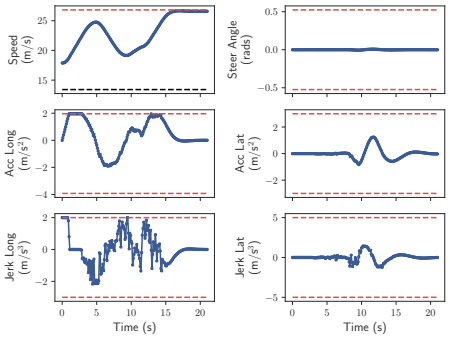
(a) Example trajectory 9.5 seconds into the simulation.



(b) Example trajectory 11 seconds into the simulation.

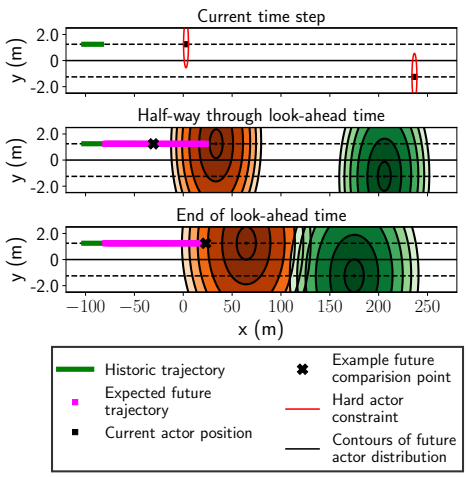


(c) Example trajectory 13 seconds into the simulation.

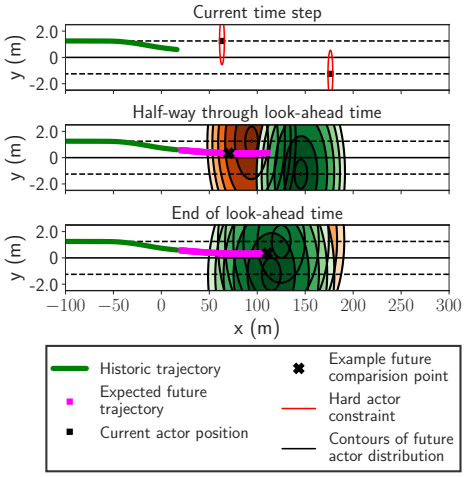


(d) Constraints summary for a vehicle overtake example.

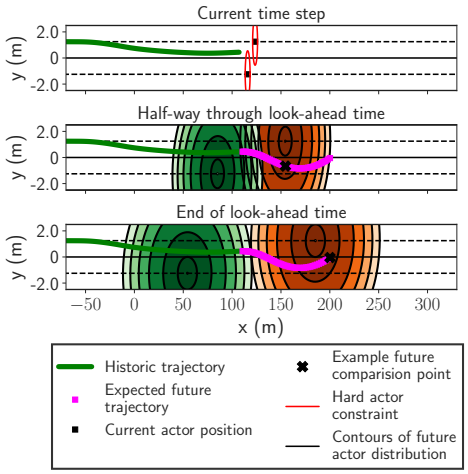
Figure F.3: Optimisation result for a vehicle overtake example.



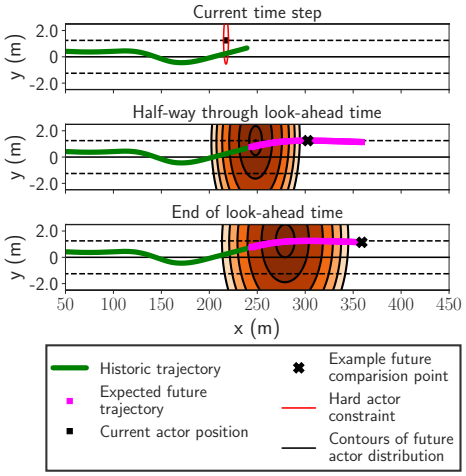
(a) Example trajectory 1 second into the simulation.



(b) Example trajectory 5.5 seconds into the simulation.

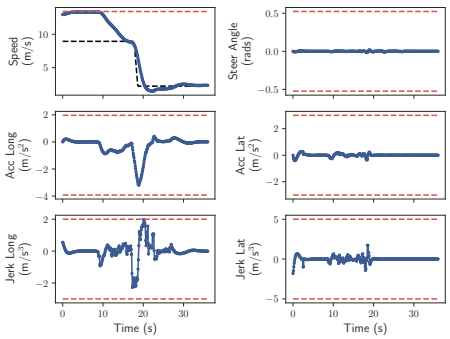


(c) Example trajectory 10 seconds into the simulation.



(d) Example trajectory 17 seconds into the simulation.

Figure F.4: Optimisation result for a vehicle overtake example with multiple actors.



(a) Constraints summary for a abnormal actor example.

Figure F.5: Optimisation result for abnormal actor behaviour scenario.

Bibliography

[1] K. Kalair and C. Connaughton. Anomaly detection and classification in traffic flow data from fluctuations in the flow-density relationship. *Transportation Research Part C: Emerging Technologies*, 127:103178, 2021.

[2] K. Kalair, C. Connaughton, and P. Alaimo Di Loro. A non-parametric Hawkes process model of primary and secondary accidents on a UK smart motorway. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70(1):80–97, 2021.

[3] K. Kalair and C. Connaughton. Dynamic and interpretable hazard-based models of traffic incident durations. *Frontiers in Future Transportation*, 2:15, 2021.

[4] K. Kalair and C. Connaughton. Large scale performance assessment of the Lighthill-Whitham-Richards model on a smart motorway. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3718–3723, Maui, HI, November 2018. IEEE.

[5] A. J. Venables, J. Laird, and H. Overman. Transport investment and economic performance: implications for project appraisal. Technical report, United Kingdom Department for Transport, 2014. URL `https://www.gov.uk/government/publications/transport-investment-and-economic-performance-tiep-report`.

[6] United Kingdom Department for Transport. Road investment strategy: for the 2015/16 - 2019/20 road period. Technical report, 2015. URL `https://www.gov.uk/government/publications/road-investment-strategy-for-the-2015-to-2020-road-period`.

[7] N. Peluffo. Strategic road network statistics. Technical report, United Kingdom Department for Transport, 2015. URL `https://www.gov.uk/government/statistics/strategic-road-network-statistics`.

[8] Highways England. Strategic road network initial report. Technical

report, 2017. URL https://www.gov.uk/government/publications/highways-englands-strategic-road-network-initial-report.

[9] A. S. Hilditch. NTIS: An innovative national traffic management system. In *Road Transport Information and Control Conference 2014 Proceedings*, pages 1–5, London, United Kingdom, October 2014.

[10] B. Hellinga and G. Knapp. Automatic vehicle identification technology-based freeway incident detection. *Transportation Research Record*, 1727 (1):142–153, 2000.

[11] L. Wei and D. Hong-ying. Real-time road congestion detection based on image texture analysis. *Procedia Engineering*, 137:196–201, 2016. Green Intelligent Transportation System and Safety.

[12] M. Amin-Naseri, P. Chakraborty, A. Sharma, S. B. Gilbert, and M. Hong. Evaluating the reliability, coverage, and added value of crowdsourced traffic incident reports from Waze. *Transportation Research Record*, 2672 (43):34–43, 2018.

[13] H. J. Payne, E. D. Helfenbein, and H. C. Knobel. Development and testing of incident detection algorithms. vol. 2, research methodology and detailed results. Technical Report FHWA-RD-76-20, Federal Highway Adminstration, Washington, D.C., 1976. URL https://rosap.ntl.bts.gov/view/dot/744.

[14] M. Levin and G. M. Krause. Incident detection: A bayesian approach. *Transportation Research Record*, 682:52–58, 1978.

[15] B. N. Persaud and F. L. Hall. Catastrophe theory and patterns in 30-second freeway traffic data - implications for incident detection. *Transportation Research Part A: General*, 23(2):103–113, 1989.

[16] A. I. Gall and F. L. Hall. Distinguishing between incident congestion and recurrent congestion: a proposed logic. *Transportation Research Record*, (1232):1–8, 1989.

[17] B. N. Persaud, F. L. Hall, and L. M. Hall. Congestion identification aspects of the McMaster incident detection algorithm. *Transportation Research Record*, (1287):167–175, 1990.

[18] F. L. Hall, Y. Shi, and G. Atala. On-line testing of the McMaster incident detection algorithm under recurrent congestion. *Transportation Research Record*, 1394:1–7, 1993.

[19] D. Sun, C. Zhang, M. Zhao, L. Zheng, and W. Liu. Traffic congestion pattern detection using an improved McMaster algorithm. In *2017 29th Chinese Control And Decision Conference (CCDC)*, pages 2814–2819, Chongqing, China, 2017.

[20] C. L. Dudek, C. J. Messer, and N. B. Nuckles. Incident detection on urban freeways. *Transportation Research Record*, 495:12–24, 1974.

[21] K. Balke, C. L. Dudek, and C. E. Mountain. Using probe-measured travel times to detect major freeway incidents in Houston, Texas. *Transportation Research Record*, 1554(1):213–220, 1996.

[22] M. Snelder, T. Bakri, and B. van Arem. Delays caused by incidents: Data-driven approach. *Transportation Research Record*, 2333(1):1–8, 2013.

[23] Y. Chung. Quantification of nonrecurrent congestion delay caused by freeway accidents and analysis of causal factors. *Transportation Research Record*, 2229(1):8–18, 2011.

[24] P. Chakraborty, C. Hegde, and A. Sharma. Data-driven parallelizable traffic incident detection using spatio-temporally denoised robust thresholds. *Transportation Research Part C: Emerging Technologies*, 105: 81–99, 2019.

[25] E. Kidando, R. Moses, T. Sando, and E. E. Ozguven. Evaluating recurring traffic congestion using change point regression and random variation Markov structured model. *Transportation Research Record*, 2672(20): 63–74, 2018.

[26] S. R. Ahmed and A. R. Cook. Application of time-series analysis techniques to freeway incident detection. *Transportation Research Record*, 841:19–21, 1982.

[27] Y-S. Jeong, M. Castro-Neto, M. K. Jeong, and L. D. Han. A wavelet-based freeway incident detection algorithm with adapting threshold parameters. *Transportation Research Part C: Emerging Technologies*, 19 (1):1–19, 2011.

[28] J. D. Crabtree and N. Stamatiadis. Dedicated short-range communications technology for freeway incident detection: Performance assessment based on traffic simulation data. *Transportation Research Record*, 2000 (1):59–69, 2007.

[29] R. L. Cheu, H. Qi, and D-H. Lee. Mobile sensor and sample-based algorithm for freeway incident detection. *Transportation Research Record*, 1811(1):12–20, 2002.

[30] B. Anbaroglu, B. Heydecker, and T. Cheng. Spatio-temporal clustering for non-recurrent traffic congestion detection on urban road networks. *Transportation Research Part C: Emerging Technologies*, 48:47–65, 2014.

[31] C. Piciarelli and G. L. Foresti. On-line trajectory clustering for anomalous events detection. *Pattern Recognition Letters*, 27(15):1835–1842, 2006. Vision for Crime Detection and Prevention.

[32] F. Porikli and X. Li. Traffic congestion estimation using HMM models without vehicle tracking. In *IEEE Intelligent Vehicles Symposium, 2004*, pages 188–193, Parma, Italy, 2004.

[33] Z. Yuan, X. Zhou, and T. Yang. Hetero-ConvLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '18, pages 984–992, London, United Kingdom, 2018.

[34] L. Zhu, R. Krishnan, A. Sivakumar, F. Guo, and J. W. Polak. Traffic monitoring and anomaly detection based on simulation of Luxembourg road network. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 382–387, Auckland, New Zealand, 2019.

[35] L. Zhu, F. Guo, R. Krishnan, and J. W. Polak. A deep learning approach for traffic incident detection in urban networks. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 1011–1016, Maui, Hawaii, 2018.

[36] J. Aguero-Valverde and P. P. Jovanis. Analysis of road crash frequency with spatial models. *Transportation Research Record*, 2061(1):55–63, 2008.

[37] Y. Fan, X. Zhu, B. She, W. Guo, and T. Guo. Network-constrained spatio-temporal clustering analysis of traffic collisions in Jianghan District of Wuhan, China. *PLOS ONE*, 13(4):1–23, 2018.

[38] J. Song, R. Wen, and W. Yan. Identification of traffic accident clusters using Kulldorff's space-time scan statistics. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 3162–3167, Seattle, WA, 2018.

[39] B. Acker and M. Yuan. Network-based likelihood modeling of event occurrences in space and time: a case study of traffic accidents in Dallas,

Texas, USA. *Cartography and Geographic Information Science*, 46(1): 21–38, 2019.

[40] M. Mehdi Moradi and J. Mateu. First- and second-order characteristics of spatio-temporal point processes on linear networks. *Journal of Computational and Graphical Statistics*, 0(0):1–21, 2019.

[41] P. M. Dixon. *Ripley's K Function*. American Cancer Society, 2014.

[42] M. Kulldorff. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1):61–72, 2001.

[43] S. J. Rey. Space-time patterns of rank concordance: Local indicators of mobility association with application to spatial income inequality dynamics. *Annals of the American Association of Geographers*, 106(4): 788–803, 2016.

[44] R. Li, F. C. Pereira, and M. E. Ben-Akiva. Overview of traffic incident duration analysis and prediction. *European Transport Research Review*, 10(2):22, 2018.

[45] T. F. Golob, W. W. Recker, and J. D. Leonard. An analysis of the severity and incident duration of truck-involved freeway accidents. *Accident Analysis & Prevention*, 19(5):375–395, 1987.

[46] Y. Chung and B-J. Yoon. Analytical method to estimate accident duration using archived speed profile and its statistical analysis. *KSCE Journal of Civil Engineering*, 16(6):1064–1070, 2012.

[47] H. Zhang and A. J. Khattak. Analysis of cascading incident event durations on urban freeways. *Transportation Research Record*, 2178(1): 30–39, 2010.

[48] W. Junhua, C. Haozhe, and Q. Shi. Estimating freeway incident duration using accelerated failure time modeling. *Safety Science*, 54:43–50, 2013.

[49] A. T. Hojati, L. Ferreira, S. Washington, P. Charles, and A. Shobeirinejad. Modelling total duration of traffic incidents including incident detection and recovery time. *Accident Analysis & Prevention*, 71:296–305, 2014.

[50] A. A. Kaabi, D. Dissanayake, and R. Bird. Response time of highway traffic accidents in Abu Dhabi: Investigation with hazard-based duration models. *Transportation Research Record*, 2278(1):95–103, 2012.

[51] I. Ghosh, P. T. Savolainen, and T. J. Gates. Examination of factors affecting freeway incident clearance times: a comparison of the generalized

F model and several alternative nested models. *Journal of Advanced Transportation*, 48(6):471–485, 2014.

[52] Y. Zou, K. Henrickson, D. Lord, Y. Wang, and K. Xu. Application of finite mixture models for analysing freeway incident clearance time. *Transportmetrica A: Transport Science*, 12(2):99–115, 2016.

[53] K. W. Smith and B. L. Smith. Forecasting the clearance time of freeway accidents. Technical report, 2001.

[54] C. Lee, W. R. Zame, J. Yoon, and M. van der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the 32-nd AAAI Conference on Artificial Intelligence*, pages 2314–2321, Palo Alto, CA, 2018.

[55] D. Nam and F. Mannering. An exploratory hazard-based analysis of highway incident duration. *Transportation Research Part A: Policy and Practice*, 34(2):85–102, 2000.

[56] Y. Chung, L. F. Walubita, and K. Choi. Modeling accident duration and its mitigation strategies on South Korean freeway systems. *Transportation Research Record*, 2178(1):49–57, 2010.

[57] Y-S. Chung, Y-C. Chiou, and C-H. Lin. Simultaneous equation modeling of freeway accident duration and lanes blocked. *Analytic Methods in Accident Research*, 7:16–28, 2015.

[58] D. Zeng and D. Y. Lin. Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*, 102(480): 1387–1396, 2007.

[59] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

[60] B. Oralhan and Z. Göktolga. Determination of the risk factors that influence occurrence time of traffic accidents with survival analysis. *Iranian journal of public health*, 47(8):1181–1191, 2018.

[61] J. Wu, R. Subramanian, M. Craig, M. Starnes, and A. Longthorne. The effect of earlier or automatic collision notification on traffic mortality by survival analysis. *Traffic Injury Prevention*, 14(sup1):S50–S57, 2013.

[62] R. Li and M. Guo. Competing risks analysis on traffic accident duration time. *Journal of Advanced Transportation*, 49(3):402–415, 2014.

[63] A. J. Khattak, J. L. Schofer, and M-H. Wang. A simple time sequential procedure for predicting freeway incident duration. *Journal of Intelligent Transportation Systems*, 2(2):113–138, 1995.

[64] A. Garib, A. E. Radwan, and H. Al-Deek. Estimating magnitude and duration of incident delays. *Journal of Transportation Engineering*, 123 (6):459–466, 1997.

[65] J. Weng, W. Qiao, X. Qu, and X. Yan. Cluster-based lognormal distribution model for accident duration. *Transportmetrica A: Transport Science*, 11(4):345–363, 2015.

[66] C. Ding, X. Ma, Y. Wang, and Y. Wang. Exploring the influential factors in incident clearance time: Disentangling causation from self-selection bias. *Accident Analysis & Prevention*, 85:58–65, 2015.

[67] A. J. Khattak, J. Liu, B. Wali, X. Li, and M. Ng. Modeling traffic incident duration using quantile regression. *Transportation Research Record*, 2554(1):139–148, 2016.

[68] W. J. J. Knibbe, T. P. Alkim, J. F. W. Otten, and M. Y. Aidoo. Automated estimation of incident duration on dutch highways. In *2006 IEEE Intelligent Transportation Systems Conference*, pages 870–874, Toronto, Canada, 2006.

[69] C. Zhan, A. Gan, and M. Hadi. Prediction of lane clearance time of freeway incidents using the M5P tree algorithm. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1549–1557, 2011.

[70] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S Lauer. Random survival forests. *Annals of Applied Statistics*, 2(3):841–860, 2008.

[71] S. Dietrich, A. Floegel, M. Troll, T. Kühn, W. Rathmann, A. Peters, D. Sookthai, M. von Bergen, R. Kaaks, J. Adamski, C. Prehn, H. Boeing, M. B. Schulze, T. Illig, T. Pischon, S. Knüppel, R. Wang-Sattler, and D. Drogan. Random survival forest in practice: a method for modelling complex metabolomics data in time to event analysis. *International Journal of Epidemiology*, 45(5):1406–1420, 2016.

[72] F. Miao, Y. Cai, Y. Zhang, and C. Li. Is random survival forest an alternative to cox proportional model on predicting cardiovascular disease? In *6th European Conference of the International Federation for Medical and Biological Engineering*, volume 45, pages 740–743, Dubrovnik, Croatia, 2015.

[73] H. Wang and G. Li. A selective review on random survival forests for high dimensional data. *Quantitative bio-science*, 36(2):85–96, 2017.

[74] W-W. Wu, S-Y. Chen, and C-J. Zheng. Traffic incident duration prediction based on support vector regression. In *11th International Conference of Chinese Transportation Professionals (ICCTP)*, pages 2412–2421, Nanjing, China, 2011.

[75] B. N. Araghi, S. Hu, R. Krishnan, M. Bell, and W. Ochieng. A comparative study of k-nn and hazard-based models for incident duration prediction. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1608–1613, Qingdao, China, 2014.

[76] W. Wang, H. Chen, and M. C. Bell. Vehicle breakdown duration modelling. *Journal of Transportation and Statistics*, 8(1):75–84, 2005.

[77] C-H. Wei and Y. Lee. Applying data fusion techniques to traveler information services in highway network. *Journal of the Eastern Asia Society for Transportation Studies*, 6:2457–2472, 2005.

[78] Y. Lee and C-H. Wei. A computerized feature selection method using genetic algorithms to forecast freeway accident duration times. *Computer-Aided Civil and Infrastructure Engineering*, 25(2):132–148, 2010.

[79] G. Valenti, M. Lelli, and D. Cucina. A comparative study of models for the incident duration prediction. *European Transport Research Review*, 2 (2):103–111, 2010.

[80] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology*, 18(1), 2018.

[81] D. Jarrett, J. Yoon, and M. van der Schaar. Dynamic prediction in clinical survival analysis using temporal convolutional networks. *IEEE Journal of Biomedical and Health Informatics*, 24(2):424–436, 2020.

[82] C-H. Wei and Y. Lee. Sequential forecast of incident duration using artificial neural network models. *Accident Analysis & Prevention*, 39(5): 944–954, 2007.

[83] R. Li, F. C. Pereira, and M. E. Ben-Akiva. Competing risk mixture model and text analysis for sequential incident duration prediction. *Transportation Research Part C: Emerging Technologies*, 54:74–85, 2015.

[84] B. Ghosh, M. T. Asif, J. Dauwels, U. Fastenrath, and H. Guo. Dynamic prediction of the incident duration using adaptive feature set. *IEEE Transactions on Intelligent Transportation Systems*, 20(11):4019–4031, 2019.

[85] Y. Lin and R. Li. Real-time traffic accidents post-impact prediction: Based on crowdsourcing data. *Accident Analysis & Prevention*, 145: 105696, 2020.

[86] X. Li, J. Liu, A. Khattak, and S. Nambisan. Sequential prediction for large-scale traffic incident duration: Application and comparison of survival models. *Transportation Research Record*, 2674(1):79–93, 2020.

[87] D. González, J. Pérez, V. Milanés, and F. Nashashibi. A review of motion planning techniques for automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 17(4):1135–1145, 2016.

[88] M. Likhachev and D. Ferguson. Planning long dynamically feasible maneuvers for autonomous vehicles. *The International Journal of Robotics Research*, 28(8):933–945, 2009.

[89] D. Dolgov, S. Thrun, M. Montemerlo, and J. Diebel. Path planning for autonomous vehicles in unknown semi-structured environments. *The International Journal of Robotics Research*, 29(5):485–501, 2010.

[90] R. Kummerle, D. Hahnel, D. Dolgov, S. Thrun, and W. Burgard. Autonomous driving in a multi-level parking structure. In *2009 IEEE International Conference on Robotics and Automation*, pages 3395–3400, Kobe, Japan, 2009.

[91] S. M. LaValle and J. J. Kuffner Jr. Randomized kinodynamic planning. *The International Journal of Robotics Research*, 20(5):378–400, 2001.

[92] L. Labakhua, U. Nunes, R. Rodrigues, and F. S. Leite. Smooth trajectory planning for fully automated passengers vehicles: Spline and clothoid based methods and its simulation. In *Informatics in Control Automation and Robotics: Selected Papers from the International Conference on Informatics in Control Automation and Robotics 2006*, volume 15, pages 169–182, Setúbal, Portugal, 2008.

[93] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, T. Strauss, C. Stiller, T. Dang, U. Franke, N. Appenrodt, C. G. Keller, E. Kaus, R. G. Herrtwich, C. Rabe, D. Pfeiffer, F. Lindner, F. Stein, F. Erbs, M. Enzweiler, C. Knoppel, J. Hipp, M. Haueis, M. Trepte, C. Brenk, A. Tamke, M. Ghanaat, M. Braun, A. Joos, H. Fritz, H. Mock, M. Hein, and E. Zeeb.

Making bertha drive - an autonomous journey on a historic route. *IEEE Intelligent Transportation Systems Magazine*, 6(2):8–20, 2014.

[94] J. H. Lee. Model predictive control: Review of the three decades of development. *International Journal of Control, Automation and Systems*, 9(3):415, 2011.

[95] P. Falcone, F. Borrelli, J. Asgari, H. E. Tseng, and D. Hrovat. Predictive active steering control for autonomous vehicle systems. *IEEE Transactions on Control Systems Technology*, 15(3):566–580, 2007.

[96] Y. Yoon, J. Shin, H. Jin Kim, Y. Park, and S. Sastry. Model-predictive active steering and obstacle avoidance for autonomous ground vehicles. *Control Engineering Practice*, 17(7):741–750, 2009.

[97] C. Philippe, L. Adouane, B. Thuilot, A. Tsourdos, and H. Shin. Safe and online mpc for managing safety and comfort of autonomous vehicles in urban environment. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 300–306, Maui, HI, 2018.

[98] K. Lee and D. Kum. Collision avoidance/mitigation system: Motion planning of autonomous vehicle via predictive occupancy map. *IEEE Access*, 7:52846–52857, 2019.

[99] A. A. Pereira, J. Binney, G. A. Hollinger, and G. S. Sukhatme. Risk-aware path planning for autonomous underwater vehicles using predictive ocean models. *Journal of Field Robotics*, 30(5):741–762, 2013.

[100] Q. Zhang, S-G. Yue, Q-J. Yin, and Y-B. Zha. Dynamic obstacle-avoiding path planning for robots based on modified potential field method. In *International Conference on Intelligent Computing: Intelligent Computing Theories and Technology*, volume 7996, pages 332–342, Nanning, China, 2013.

[101] Y. Rasekhipour, A. Khajepour, S. Chen, and B. Litkouhi. A potential field-based model predictive path-planning controller for autonomous road vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 18(5):1255–1267, 2017.

[102] H. Wang, Y. Huang, A. Khajepour, Y. Zhang, Y. Rasekhipour, and D. Cao. Crash mitigation in motion planning for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 20(9):3313–3323, 2019.

[103] B. Németh, P. Gáspár, and T. Hegedűs. Optimal control of overtaking maneuver for intelligent vehicles. *Journal of Advanced Transportation*, 2018(1):1–11, 2018.

[104] P. Petrov and F. Nashashibi. Planning and nonlinear adaptive control for an automated overtaking maneuver. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 662–667, Washington, DC, 2011.

[105] Y. Wang, T. Qu, J. Chu, S. Yu, and H. Guo. Trajectory planning and tracking control of vehicle obstacle avoidance based on optimization control. In *2019 Chinese Control Conference (CCC)*, pages 3157–3162, Guangzhou, China, 2019.

[106] Q. Wang, B. Ayalew, and T. Weiskircher. Predictive maneuver planning for an autonomous vehicle in public highway traffic. *IEEE Transactions on Intelligent Transportation Systems*, 20(4):1303–1315, 2019.

[107] Y. Pan, C-A. Cheng, K. Saigol, K. Lee, X. Yan, E. A. Theodorou, and B. Boots. Imitation learning for agile autonomous driving. *The International Journal of Robotics Research*, 39(2-3):286–302, 2020.

[108] J. Chen, B. Yuan, and M. Tomizuka. Deep imitation learning for autonomous driving in generic urban scenarios with enhanced safety. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2884–2890, Macau, China, 2019.

[109] F. Codevilla, M. Mäller, A. López, V. Koltun, and A. Dosovitskiy. End-to-end driving via conditional imitation learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4693–4700, Brisbane, Australia, 2018.

[110] J. E. Chacón and T. Duong. *Multivariate Kernel Smoothing and Its Applications*. CRC Press, Taylor & Francis Group, Boca Raton, Florida, 2018.

[111] B. W. Silverman. *Density estimation for statistics and data analysis*. CRC Press, Taylor & Francis Group, Boca Raton, Florida, 1998.

[112] M. C. Jones, J. S. Marron, and S. J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American statistical association*, 91(433):401–407, 1996.

[113] J. E. Chacón and T. Duong. Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test*, 19(2):375–398, 2010.

[114] A. Reinhart. A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, 33(3):299–318, 2018.

[115] J. Zhuang, Y. Ogata, and D. Vere-Jones. Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth*, 109(B5):17, 2004.

[116] J. Zhuang. Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth, Planets and Space*, 63(3):5, 2011.

[117] T. Kumazawa and Y. Ogata. Nonstationary ETAS models for nonstandard earthquakes. *Annals of Applied Statistics*, 8(3):1825–1852, 2014.

[118] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, 2011.

[119] G. Mohler. Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*, 30(3): 491–497, 2014.

[120] S. Meyer and L. Held. Power-law models for infectious disease spread. *Annals of Applied Statistics*, 8(3):1612–1639, 2014.

[121] F. P. Schoenberg, M. Hoffmann, and R. J. Harrigan. A recursive point process model for infectious diseases. *Annals of the Institute of Statistical Mathematics*, 71(5):1271–1287, 2019.

[122] E. W. Fox, M. B. Short, F. P. Schoenberg, K. D. Coronges, and A. L. Bertozzi. Modeling e-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*, 111(514):564–584, 2016.

[123] L. Lesage. A Hawkes process to make aware people of the severity of COVID-19 outbreak: application to cases in France. Research report, Université de Lorraine; University of Luxembourg, 2020. URL `https://hal.archives-ouvertes.fr/hal-02510642`.

[124] J. Zhuang and J. Mateu. A semiparametric spatiotemporal Hawkes-type point process model with periodic background for crime data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(3): 919–942, 2019.

[125] J. Zhuang. Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(4):635–653, 2006.

[126] L. Zhongping, C. Lirong, and C. Jianhui. Traffic accident modelling via self-exciting point processes. *Reliability Engineering and System Safety*, 180:312–320, 2018.

[127] K. W. Lim, W. Wang, H. Nguyen, Y. Lee, C. Cai, and F. Chen. Traffic flow modelling with point processes. In *Proceedings of the 23rd World Congress on Intelligent Transport Systems*, pages 1–12, Melbourne, Australia, 2016.

[128] A. LeNail. NN-SVG: publication-ready neural network architecture schematics. *Journal of Open Source Software*, 4(33):747, 2019.

[129] M. T. Ribeiro, S. Singh, and C. Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 1135–1144, San Francisco, CA, 2016.

[130] A. Shrikumar, P. Greenside, and A. Kundaje. Learning important features through propagating activation differences. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 3145–3153, Sydney, Australia, 2017.

[131] S. M. Lundberg and S. Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*, pages 4765–4774, Los Angeles, CA, 2017.

[132] S. Lundberg. SHAP (SHapley Additive exPlanations). `https://github.com/slundberg/shap`, 2017. Accessed: 14-12-2019.

[133] F. A. Potra and S. J. Wright. Interior-point methods. *Journal of Computational and Applied Mathematics*, 124(1):281–302, 2000. Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.

[134] R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. Wiley, Hoboken, NJ, USA, 3 edition, May 2019.

[135] R. J. A. Little. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83(404):1198–1202, December 1988.

[136] S. French. Uncertainty and imprecision: Modelling and analysis. *The Journal of the Operational Research Society*, 46(1):70–79, January 1995.

[137] S. French, S. Haywood, D. H. Oughton, and C. Turcanu. Different types of uncertainty in nuclear emergency management. *Radioprotection*, 55: S175–S180, May 2020.

[138] Highways England. Safer driving on motorways. https://highwaysengland.co.uk/road-safety/safer-driving-on-motorways/, 2021. Accessed: 12-02-2021.

[139] B. Harbord, J. White, K. Mccabe, A. Riley, and S. Tarry. A flexible approach to motorway control. In *Proceedings of the 13th ITS World Congress 2006*, pages 37–42, London, United Kingdom, 2006.

[140] V. Knoop, S. P. Hoogendoorn, and H. van Zuylen. Empirical differences between time mean speed and space mean speed. In *Traffic and Granular Flow '07*, pages 351–356, 2009.

[141] J. Jiang, J. W. Polak, J. Barria, and R. Krishnan. On the estimation of space-mean-speed from inductive loop detector data. *Transportation Planning and Technology*, 33(1):91–104, 2010.

[142] J. G. Wardrop. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, Part II:325–378, 1952.

[143] C. F. Daganzo. *Fundamentals of transportation and traffic operations*. Emerald Group Publishing Limited, Bingley, United Kingdom, September 1997.

[144] R. D. Kühne. Greenshields' legacy: Highway traffic. In *75 Years of the Fundamental Diagram for Traffic Flow Theory: The Fundamental Diagram*, number E-C149, pages 3–10, Woods Hole, MA, 2008.

[145] M. J. Cassidy. Bivariate relations in nearly stationary highway traffic. *Transportation Research Part B: Methodological*, 32(1):49–59, 1998.

[146] M. J. Cassidy and B. Coifman. Relation among average speed, flow, and density and analogous relation between density and occupancy. *Transportation Research Record*, 1591(1):1–6, 1997.

[147] T. Duong. *ks: Kernel Smoothing*, 2018. URL https://CRAN.R-project.org/package=ks. R package version 1.11.3.

[148] Y-C. Chen, C. R. Genovese, and L. Wasserman. Density level sets: Asymptotics, inference, and visualization. *Journal of the American Statistical Association*, 112(520):1684–1696, 2017.

[149] Systematics. Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation. Technical report, 2005. URL https://ops.fhwa.dot.gov/congestion_report/. Federal Highway Administration, Texas Transportation Institute.

[150] J. S. Ren, W. Wang, J. Wang, and S. Liao. An unsupervised feature learning approach to improve automatic incident detection. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 172–177, Anchorage, AK, 2012.

[151] R. L. Cheu, D. Srinivasan, and E. T. Teh. Support vector machine models for freeway incident detection. In *Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems*, volume 1, pages 238–243, Shanghai, China, October 2003.

[152] B. M. Williams and A. Guin. Traffic management center use of incident detection algorithms: Findings of a nationwide survey. *IEEE Transactions on Intelligent Transportation Systems*, 8(2):351–358, 2007.

[153] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[154] B. Trawiński, M. Smętek, Z. Telec, and T. Lasota. Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. *International Journal of Applied Mathematics and Computer Science*, 22(4):867–881, 2012.

[155] A. Gupta, H. P. Gupta, B. Biswas, and T. Dutta. A divide-and-conquer-based early classification approach for multivariate time series with different sampling rate components in IoT. *ACM Transactions on Internet of Things*, 1(2):21, 2020.

[156] J. Hamidzadeh and N. Namaei. Belief-based chaotic algorithm for support vector data description. *Soft Computing*, 23(12):4289–4314, 2019.

[157] World Health Organization. *Global status report on road safety 2018*. World Health Organization, 2018.

[158] Department for Transport. Reported road casualties in Great Britain: main results 2018. Technical report, 2019. URL `https://www.gov.uk/government/statistics/reported-road-casualties-great-britain-main-results-2018`.

[159] S-T. Chiu. Boundary adjusted density estimation and bandwidth selection. *Statistica Sinica*, 10(4):1345–1367, 2000.

[160] P. Hall and L-S. Huang. Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics*, 29(3):624–647, 2001.

[161] INRIX. INRIX reveals the UK's worst traffic jams over the past year, 2017. URL `http://inrix.com/press-releases/uk-traffic-jams/`. Accessed: 23-12-2019.

[162] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[163] E. N. Brown, R. Barbieri, V. Ventura, R. E. Kass, and L. M. Frank. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 14(2):325–346, 2002.

[164] R. B. Cleveland, W. S. Cleveland, and I. Terpenning. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73, 1990.

[165] G. Jogesh Babu and C. R. Rao. Goodness-of-fit tests when parameters are estimated. *Sankhyā: The Indian Journal of Statistics (2003-2007)*, 66(1):63–74, 2004.

[166] D. J. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, pages 359–370, Seattle, WA, 1994.

[167] H. Izakian, W. Pedrycz, and I. Jamal. Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence*, 39:235–244, 2015.

[168] P. Montero and J. A. Vilar. TSclust: An R package for time series clustering. *Journal of Statistical Software*, 62(1):1–43, 2014.

[169] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

[170] D. Kumar and B. Klefsjö. Proportional hazards model: a review. *Reliability Engineering & System Safety*, 44(2):177–188, 1994.

[171] D. G. Kleinbaum and M. Klein. *Survival Analysis*. Statistics for Biology and Health. Springer-Verlag, New York, 3rd edition, 2012.

[172] T. M. Therneau. *A Package for Survival Analysis in S*, 2015. URL `https://CRAN.R-project.org/package=survival`. version 2.38.

[173] B. Efron. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72(359):557–565, 1977.

[174] C. Jackson. flexsurv: A platform for parametric survival modeling in R. *Journal of Statistical Software*, 70(8):1–33, 2016.

[175] H. Ishwaran and U. B. Kogalur. Random survival forests for R. *R News*, 7(2):25–31, 2007.

[176] D. Faraggi and R. Simon. A neural network model for survival data. *Statistics in Medicine*, 14(1):73–82, 1995.

[177] M. Luck, T. Sylvain, H. Cardinal, A. Lodi, and Y. Bengio. Deep learning for patient-specific kidney graft survival analysis. *Arxiv preprint*, 2017. arXiv: 1705.10245.

[178] C. Lee. Deephit. `https://github.com/chl8856/DeepHit`, 2020. Accessed: 01-10-2020.

[179] U. Dafni. Landmark analysis at the 25-year landmark point. *Circulation: Cardiovascular Quality and Outcomes*, 4(3):363–371, 2011.

[180] C. Lee, J. Yoon, and M. van der Schaar. Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Transactions on Biomedical Engineering*, 67 (1):122–133, 2020.

[181] F. E. Harrell Jr, K. L. Lee, and D. B. Mark. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15 (4):361–387, 1996.

[182] Y. Yuan. *Prediction Performance of Survival Models*. PhD thesis, Waterloo, Ontario, Canada, 2008.

[183] Y. Qi and H. Teng. An information-based time sequential approach to online incident duration prediction. *Journal of Intelligent Transportation Systems*, 12(1):1–12, 2008.

[184] IBI Group UK. Highways England's provision of information to road users. Technical report, Office of Rail and Road and Highways England, 2019. URL `https://www.orr.gov.uk/media/10534`.

[185] H. C. van Houwelingen and H. Putters. *Dynamic Prediction in Clinical Survival Analysis*. Monographs on Statistics and Applied Probability 123. CRC Press, Boca Raton, Florida, 2012.

[186] L. Claussmann, M. Revilloud, D. Gruyer, and S. Glaser. A review of motion planning for highway autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 21(5):1826–1848, 2020.

[187] D. Nistér, H. Lee, J. Ng, and Y. Wang. An introduction to the safety force field. White paper, NVIDIA, 2019.

[188] S. Shalev-Shwartz, S. Shammah, and A. Shashua. On a formal model of safe and scalable self-driving cars. *Arxiv preprint*, 2017. arXiv: 1708.06374.

[189] Baidu Apollo Team. Apollo: Open source autonomous driving. `https://github.com/ApolloAuto/apollo`, 2017. Accessed: 14-11-2019.

[190] United Nations. Uniform provisions concerning the approval of: Vehicles with regard to steering equipment, 2008.

[191] R. Rajamani. *Vehicle Dynamics and Control*. Mechanical Engineering Series. Springer New York, , 2 edition, 2012.

[192] J. A. Matute, M. Marcano, S. Diaz, and J. Perez. Experimental validation of a kinematic bicycle model predictive control with lateral acceleration consideration. *IFAC-PapersOnLine*, 52(8):289–294, 2019. 10th IFAC Symposium on Intelligent Autonomous Vehicles IAV 2019.

[193] R. Pepy, A. Lambert, and H. Mounier. Path planning using a dynamic vehicle model. In *2006 2nd International Conference on Information Communication Technologies*, volume 1, pages 781–786, Damascus, Syria, 2006.

[194] R. He, J. Zhou, S. Jiang, Y. Wang, J. Tao, S. Song, J. Hu, J. Miao, and Q. Luo. TDR-OBCA: A reliable planner for autonomous driving in free-space environment, 2020. arXiv: 2009.11345.

[195] J. Kong, M. Pfeiffer, G. Schildbach, and F. Borrelli. Kinematic and dynamic vehicle models for autonomous driving control design. In *2015 IEEE Intelligent Vehicles Symposium (IV)*, pages 1094–1099, Seoul, South Korea, 2015.

[196] P. Polack, F. Altché, B. D'Andrea-Novel, and A. de La Fortelle. Guaranteeing consistency in a motion planning and control architecture using a kinematic bicycle model. In *2018 Annual American Control Conference (ACC)*, pages 3981–3987, Milwaukee, WI, 2018.

[197] R. H. Byrd, J. C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1):149–185, 2000.

[198] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

[199] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. S. Torr, and M. Chandraker. Desire: Distant future prediction in dynamic scenes with interacting agents. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2165–2174, Honolulu, HI, 2017.

[200] K. Leung, E. Schmerling, and M. Pavone. Distributional prediction of human driving behaviours using mixture density networks. Technical report, Stanford University, 2016.

[201] Y. Zhao, R. Yang, G. Chevalier, R. C. Shah, and R. Romijnders. Applying deep bidirectional lstm and mixture density network for basketball trajectory prediction. *Optik: International Journal for Light and Electron Optics*, 158:266–272, 2018.

[202] J. E. Naranjo, C. Gonzalez, R. Garcia, and T. de Pedro. Lane-change fuzzy control in autonomous vehicles for the overtaking maneuver. *IEEE Transactions on Intelligent Transportation Systems*, 9(3):438–450, 2008.

[203] P. Petrov and F. Nashashibi. Modeling and nonlinear adaptive control for autonomous vehicle overtaking. *IEEE Transactions on Intelligent Transportation Systems*, 15(4):1643–1656, 2014.

[204] G. Usman and F. Kunwar. Autonomous vehicle overtaking - an online solution. In *2009 IEEE International Conference on Automation and Logistics*, pages 596–601, Shenyang, China, 2009.

[205] S. Shalev-Shwartz, S. Shammah, and A. Shashua. Safe, multi-agent, reinforcement learning for autonomous driving. *Arxiv preprint*, 2016. arXiv: 1610.03295.

[206] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[207] Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, Cambridge, 2008.

[208] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, 1990.

[209] J. C. Gower. A general coefficient of similarity and some of its properties. *Biometrics*, 27(4):857–871, 1971.

[210] S. Pölsterl. scikit-survival: A library for time-to-event analysis built on top of scikit-learn. *Journal of Machine Learning Research*, 21(212):1–6, 2020.

[211] J. Lambert and S. Chevret. Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. *Statistical Methods in Medical Research*, 25(5):2088–2102, 2016.