# AGQ CDT Placement Impact Report

Tudor-Ioan Caba, Lucia Noelle

August 2026

THE OFFICE OF

THE SCOTTISH ROAD WORKS COMMISSIONER

Algebra Geometry & Quantum Fields

# Contents


**1 Background** **3**
1.1 The Scottish Roadworks Register ... 3
1.1.1 The National Street Gazetteer ... 4
1.1.2 Description of current system ... 4

**2 Problem statement** **6**
2.1 Current challenges ... 6
2.2 Research questions ... 6

**3 Data sources** **8**
3.1 The Scottish Road Works Register ... 8
3.2 INRIX ... 9
3.3 TomTom ... 10
3.4 Glasgow Traffic API ... 10
3.5 Drakewell ... 11
3.6 OpenStreetMap ... 11
3.7 Bus Data ... 11
3.8 Waze ... 11
3.9 UBDC Origin-Destination Matrices ... 11

**4 Data-driven analysis** **12**
4.1 Predictive relationship between traffic data (INRIX) and roadworks ... 12
4.1.1 Manual inspection ... 13
4.1.2 Anomaly detection methods ... 17
4.1.3 Random Forest ... 25
4.1.4 Transformer-based methods ... 27
4.2 Incorporating traffic counts ... 29
4.3 Predictive relationship between traffic data (INRIX) and bus disruptions ... 31

**5 A quantitative model for traffic sensitivity** **34**
5.1 The simplest criticality model: All-or-Nothing ... 34
5.2 Traffic Assignment with the Frank-Wolfe Algorithm ... 37
5.3 Case Study: The road network in Glasgow ... 38

**6 Conclusion and recommendations** **47**
6.1 Discussion of experiments ... 47
6.1.1 Traffic data analysis ... 47
6.1.2 Network criticality analysis ... 48
6.2 Practical suggestions for the Scottish Roadworks Register ... 49
6.3 Data-related suggestions ... 50

**7 Further work** **52**
7.1 Refining and building on the traffic sensitivity analysis ... 52
7.2 Revisiting the traffic data ... 52
7.3 Simulating traffic in a small neighbourhood ... 52
7.4 Learning conflict resolution ... 53
7.5 Leveraging the bus delay data ... 53
7.6 Bayesian methods ... 53

**Appendices** **55**

**A A note on model evaluation** **56**
A.1 Metrics ... 56

**B Importance of OD attachment nodes** **58**

# Chapter 1

# Background

The Scottish Roadworks Commissioner is in charge of overseeing the planning, coordination and execution of road works in Scotland. Roadworks in Scotland are regulated by the New Roads and Street Works Act (1991). There are three types of actors that have the power to undertake roadworks. Utilities companies, local authorities and others. The former two have statutory powers but need to give notice of their works through the Scottish Roadworks register. The latter need permission from the local authorities to undertake their works. The Scottish Roadworks register is the centralised database that records all roadworks in Scotland.

## 1.1 The Scottish Roadworks Register

When companies give notice of their planned works via the register they need to provide information on the exact location of the works, timing, traffic management etc.

Roadworks fall into the following categories:

- Emergency
- Urgent
- Minor works
- Minor works without excavation
- Standard works
- Major works
- Remedial works

There are different noticing requirements. Emergency/Urgent works do not need advance notice, notice only needs to be given within 2 hours of work starting. Clearly these works cannot be planned for. Minor works without excavation do not need notice at all but are still supposed to be recorded on the register as non-notifiable works. Other works require several advance notices (three months, one month, one week in advance).

The seven-day and 24-hour notices have a validity period that is equal to the length of the notice itself. Meaning that if a seven-day notice is recorded, then the undertaker has seven days from the proposed start date to commence the works. There is a possibility to request an extension to the period if required. One-month and three-month notices have an expiry date. Works that are not started within the validity period/before the expiry date are considered abandoned and a new notice is required.

Currently the impact of road works is determined from the type of traffic management, reinstatement category of the road and traffic sensitivity. Traffic sensitivity is intended to capture local geometry. The reinstatement category is a measure of how much traffic the road takes on a regular basis.

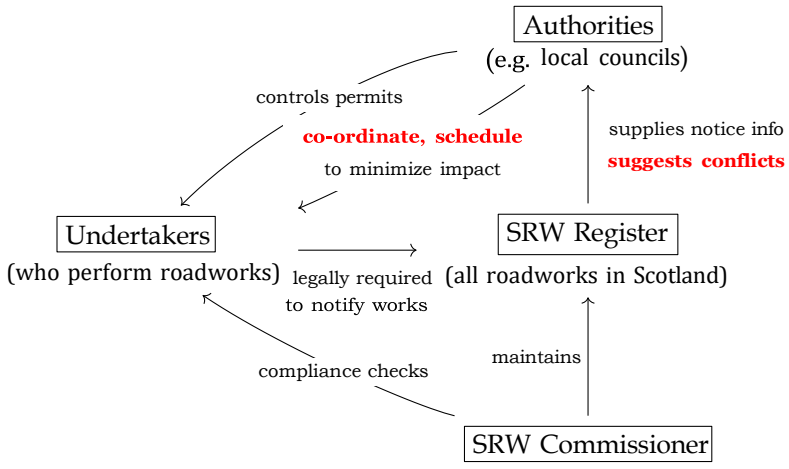
It is the local authorities' duty to coordinate all the roadworks in their area. By default works go ahead as proposed by undertakers on the register, it is only if the local authorities see a conflict and ask the undertakers to change their plans. They have the right to issue directions to the undertakers:

- Section 115: Direction to the Timing of works – Direction for when works should be carried out.
- Section 115A: Direction to the Placing of Apparatus – Alternate route.
- Section 125(3): Avoidance of unnecessary delay or obstruction – Direction to get off the road.

### 1.1.1 The National Street Gazetteer

Roads data is contained in the Gazetteer. Roads are classified by reinstatement categories which serve as a proxy for how busy the road is. Reinstatement is measured in terms of million standard axles, so a measure of how many vehicles the road needs to withstand over a given period of time. There is additional information recorded: roads may be classified as traffic-sensitive which seeks to capture how significant the road is in its local context. Moreover special engineering difficulties may be recorded for a road. Roads that have just been resurfaced may be designated as protected/embargoed by local authorities which stops any works taking place for a certain period of time.

### 1.1.2 Description of current system




Some of the duties of the Commissioner include:

- To maintain the Scottish Road Works Register (SRWR), which centralizes information about all past, ongoing, and planned road works in Scotland,
- To enforce noticing practices and to ensure that works are executed satisfactorily, with the legal ability to issue fines,
- To ensure that the impact of roadworks is minimized.

While undertakers are legally required to submit various types of notices to the Register along the life cycle of a work, the SRWC does *not* issue permits to undertakers, and are *not* involved in the

roadworks scheduling process: these duties fall on the road owners (typically various local councils).

As a result, good communication between the local authorities and the undertakers is crucial for effective coordination and minimising impact on road users. The register provides some functionalities for that, allowing authorities to communicate about specific notices, and approve requests.

In order to facilitate the coordination of roadworks, the Commissioner can offer *recommendations* to the authorities using the Register. Currently, the recommendations consist of the following:

**Current conflict checks**.

1. **Proximity** conflict check: whenever two works at a distance less than 50m (500$m$ in traffic sensitive areas) from each other are happening / planned to happen at the same point in time, the Register raises a flag.
2. **Same bus line** conflict check (upcoming): Whenever the system detects two or more works that are happening / scheduled to happen at the same point in time *along the same bus line*, the Register raises a flag.

These flags are intended to help human operators sift and prioritize the incoming planned roadworks for the purpose of scheduling, in order to minimize local impact.

They are also shown when works notices are created so undertakers can see potential conflicts and assess whether their plans need to change.

The SRWR provides information to local authorities but it is their duty to check all notices and make all decisions independently. Therefore any type of information (in particular conflict flagging) within the register is for information purposes only and the SRWC is not liable for any consequences.

# Chapter 2

# Problem statement

## 2.1 Current challenges

**Problem.** The system flags too many conflicts, most of them irrelevant, which makes them unhelpful in prioritizing and aiding the scheduling process.

Different locations experience challenges in different guises:

- In big cities, with a well-connected internal road system, there are many smaller-scale roadworks, which makes it harder to sift through the incoming information and decide what deserves priority and what not. There is a high volume of works, and their cross-interactions are complex.
- In small cities, coordinating becomes the most complicated issue. As explained to us by Falkirk city council, if one of the arterial roads through Falkirk has important road-works, it becomes quite hard to coordinate the rest of the small road works along the route as to not completely disrupt traffic, and also maintain some capacity for possible urgent roadworks.
- In rural areas, roads are generally narrower so traffic management options are limited, and road closures become more likely. A lane closure or stop/go boards can effectively become a road closure. Furthermore, the low density of the road network means diversions can be very long, or even impossible.
- On motorways, the impact of roadworks is not as localised, so the cumulative effect of roadworks can be very high, even if these works are spaced apart.

Currently, local authorities use an ad-hoc system for dealing with the incoming works within their area. These systems rely heavily on common sense and local knowledge. Some authorities (such as Traffic Scotland) complement their knowledge with quantitative solutions utilizing historic traffic data, live traffic monitoring, as well as traffic models. However, these methods tend to be expensive, not always accurate, and only deployed for major works on trunk roads.

## 2.2 Research questions

The current conflict checking mechanism is very coarse, and does not utilize any information apart from what is contained within register notices. Given that traffic data is collected for many purposes, the question is whether it could be used to obtain a more granular understanding of roadworks' impact on real-life traffic *locally*, i.e. along a road.

On a larger scale, a question is to determine the significance of a road within the larger network as the importance of a road determines the potential impact roadworks on it can have. This needs to take into account the geometry of the road network and the demand placed on it in real life. Quantifying the traffic sensitivity of a road could be used as a criterion for ranking and prioritizing roadworks, thereby providing more useful and realistic information to Register users.

Based on these considerations, we investigate the following questions:

1. What data sources are available, and how could they potentially be used to address the challenges?
2. Is there a predictive relationship between traffic conditions (as captured by various data sources) and roadworks?
3. Is there a quantitative way of capturing traffic sensitivity in order to identify works that are most likely to cause significant disruption?

We address Question 1 in Chapter 3, where we provide a description of all data sources identified. In Chapter 4 we use various statistical and machine learning methods to evaluate the relation between traffic data and roadworks, answering Question 2. We treat Question 3 in Chapter 5, where we implement a method for capturing traffic sensitivity based on traffic assignment algorithms, and exemplify this method on the road network of Glasgow. In Chapter 6 we discuss the results of our research, and make recommendations for possible improvements. We propose further research directions in Chapter 7. Further technical details can be found in the Appendices.

# Chapter 3

# Data sources

The following are the data sources that we explored, with a description of their basic functionality as well as the features we found to be the most useful in our research. Locating and handling quality data proved to be one of the touchstones within the project. Obtaining parts of this data was a challenge, and we actively engaged with third parties and stakeholders in the process. For completeness and future reference, data sources that we did not use in our analyses are included as well.

## 3.1 The Scottish Road Works Register

The Scottish Road Works Register is the online database that centralizes information about past, present and future roadworks in Scotland. It contains data starting in 2008.

It is a legal requirement for undertakers to report roadworks to the register. When companies give notice of their planned works via the register, they need to provide various types of information about the work. We highlight here the most important categories of information for our purposes:

1. Road work type. Some of the categories include
    - Emergency
    - Urgent
    - Minor works
    - Minor works without excavation
    - Standard works
    - Major works
    - Remedial works
    - Works entirely on the footway
2. Traffic management, including (for the traffic-impacting works)
    - Road Closure
    - Lane Closure
    - Contraflow
    - Traffic Lights
    - Traffic Signal
    - Give and Take

3. Time and duration information. Depending on the nature of the roadwork, there are various requirements for advance noticing. For us, the most important parts were the Actual Start and Actual End tags, which are supposed to encode the actual timespan of the work (within a margin of a 2 hours).

4. Location information. This contains both information such as road name etc. as well as WKT ("Well-Known Text") representations of the geographical region on the map on which they are taking place (e.g. part of a road segment).

   Currently, the data can be visualized using the platform provided by Symology. There is no API; data can be downloaded manually by geographical region from the interface. There are usually many distinct entries for a particular work during its timespan, and we performed our own filtering when using the data.

   Apart from storing roadwork data, the register also functions as a tool that councils use to stay up to date with and manage roadworks in their area. For example, there is a comment system which is sometimes used as a means of communication between the undertakers and the managing entity. Meaning that notices often contain a lot of detail.

The Scottish Roadworks Register is hosted by Symology, a UK based provider of Asset Management Software.

## 3.2 INRIX

INRIX is a traffic analytics platform. The SRWC has been using INRIX since 2025, Traffic Scotland has been working with them for several years. We obtained access to INRIX's roadway analytics portal through Traffic Scotland.

INRIX collects connected vehicle data (and a very small amount of mobile phone data). Currently they estimate to capture about 10-15% of vehicles on the road. The raw data is processed and offered via a portal. The user can either select a larger region or a corridor and download average speeds and travel times for a time period. Roads are split up into segments of varying sizes (ranging from a few metres to a few kilometres depending on the local geometry).

Information available per segment are average speed, historical average speed, reference speed (free-flow speed), average travel time for vehicles traversing this segments as well as confidence scores for the observations and a road closure binary variable. They are available in 1, 5, 15, 60 minute intervals. The values in the larger time intervals are weighted averages of observations within the time interval. The coarser the time intervals the more information is potentially lost. INRIX will only show a value for the average speed in a 5 minute interval if it has values for every minute within that interval. Similarly for every 15 minute, 60 minute interval. Intervals start on the hour, so incomplete intervals are discarded as granularity is reduced. Hence 60 minute intervals are not useful to work with as they only appear if the system captured data for every single minute that hour. We have been working with 5 or 15 minute intervals.

Speeds are only given as integers so precision is lost with rounding. Travel times are given with two decimal digits precision. Travel times and speeds are captured in separate processes, so they are not related by the usual formula. The proportion of vehicle types that INRIX tracks is unknown. For instance if there are a lot of HGVs in a road segment then the average speed might be lower than reflective of traffic flow at that time because HGVs tend to travel more slowly.

Reference speed is understood as free flow speed. This value is calculated as the 66th percentile of observed speeds. The rationale given for this is that 1/3 of the day is congested, 1/3 of the day is

'normal', and 1/3 of the day is uncongested. Congestion is reported as a percentage of free flow speed which is dependent on the traffic conditions and may change for consecutive segments.

If the system does not record enough data for a road segment over a given time period then the system fills it in with historic averages or free flow speed. There is an option to disable that when downloading data. Data is available going back over a decade but the real data is patchy, From about mid-2023 there is real-time data available for most of the day (6am-8pm). During weekends and bank holidays not a lot of real data is recorded due to reduced traffic flow.

If vehicles are stationary for a long time the system assumes that they are parked and discards the data which can lead to congestion/traffic jams being underrecorded. As of July 2026, there is a fix being applied for this for all roads. The system will take into account whether there are a lot of vehicles stationary in the vicinity to distinguish between parked cars and slow traffic. This means that data from 07/2026 might display fundamentally different patterns than before.

Overall the data is very averaged (average speed across a road segment averaged across all vehicles traversing the segment during a given time period), which can be helpful for getting the general picture of the traffic situation.

## 3.3 TomTom

TomTom is another traffic analytics platform. The data collection is fleet-based, the same as INRIX, but it is processed in different ways and the end analytics products are different.

It is not currently being used by the Commissioner; we learned about it from the Scottish Safety Camera Project staff, who offered us samples of the traffic data recorded along a single road and in the surrounding region. We later obtained access to a free 30-day trial of TomTom Move, offering us full functionality but with data restricted to the month of August 2024 (and later July 2026).

The traffic data is significantly less granular than INRIX. It comes as statistics over a selected period of time (e.g. mean, average, standard deviation, percentiles and sample sizes of speed and travel time over time intervals of at least 1 hour). TomTom offers however a very useful tool for obtaining Origin-Destination (OD) matrices: one can draw polygonal regions on the map, choose a time interval, and obtain a matrix containing the numbers of trips recorded by TomTom between each pair of regions during that time interval. Note however that since statistics regarding what percentage of traffic in that time period was captured by TomTom are absent, it is unclear how these OD matrices relate to the real-life number of trips.

TomTom is more user friendly than INRIX and has built-in support for both importing and exporting GeoJSON data.

## 3.4 Glasgow Traffic API

Glasgow City Council have an online portal Glasgow Open Data that offers access to various traffic-related APIs. GCC staff arranged access to a business account. The traffic API gives access to data from over 500 traffic counters across the city. Every counter offers hourly flows as well as historical statistics (mean, maximum, minimum etc.). We had some trouble consistently querying the API for larger quantities of data and found that the data required a lot of

cleaning.

We were able to access a cleaned dataset of traffic counts ranging over 4 years (October 2019-Octover 2023) from the Urban Big Data Centre at the University of Glasgow.

## 3.5 Drakewell

Drakewell is a system of automatic traffic counters employed by Transport Scotland. For every site there are hourly traffic counts available. There are 3651 locations of which 1688 are still active. Historic data going back to 2000 but all of the sites that are still active are located along trunk roads.

We decided not to use this dataset because the locations of the counters did not pertain to our studies.

## 3.6 OpenStreetMap

OpenStreetMap is an online, open source map data provider. The data comes from volunteers. We used OpenStreetMap for extracting road networks of cities, as well as for the adjacent road information provided, such as road lane counts and road directionality (for one-way roads).

## 3.7 Bus Data

Lothian Buses captures a lot of data. We obtained access to LB Roads, their own online platform where they manually curate bus-impacting roadworks ( from the Register), collect comments about traffic obstructions from bus drivers and track diversions. It is a good resource for inspecting various bus routes, but we did not find that it contains any other useful bulk data.

They also collect data and statistics for all individual bus trips (not available on LB Roads). Via discussions with Lothian Buses, we obtained 4 months of data (1 March - 30 June 2026) for Route 21. There is scope to obtain the same type of data for a longer period of time / multiple routes, although we have been unable to due to time constraints.

## 3.8 Waze

Waze for cities is an interactive platform that monitors traffic. We obtained access but could not work out how to gather and analyse pertinent traffic data. The only data that was readily available were traffic incidents and potholes.

## 3.9 UBDC Origin-Destination Matrices

The Urban Big Data Centre in Glasgow hosts various traffic-related datasets. We have identified this dataset providing an OD matrix for Edinburgh, assembled from trips between approx. 200 zones using data from a provider known as Huq over 4 years (2019 - 2023). We would hope the quality of this dataset is higher and closer to real-life counts than the OD matrices provided by TomTom. However we were unsuccessful in sourcing the data from UBDC despite persistent efforts during the placement programme and that have continued in order to help support subsequent placement researchers.

# Chapter 4

# Data-driven analysis

In this chapter we seek to quantify the traffic impact of roadworks. First we would like to establish a predictive relationship between the INRIX traffic data and roadworks as this serves as a proxy for impact. Then we add in traffic counts to improve the predictive power of the model. Finally we investigate the relationship between the traffic data and bus delays as bus delays provide an alternative notion of roadworks impact.

## 4.1 Predictive relationship between traffic data (INRIX) and roadworks

As a case study we chose at Gilmerton Road in Edinburgh. This is a fairly straight road connecting the ring road to the city centre. On INRIX it is made up of seven segments totalling 4.45km. We chose it because of its traffic volume, length and location.

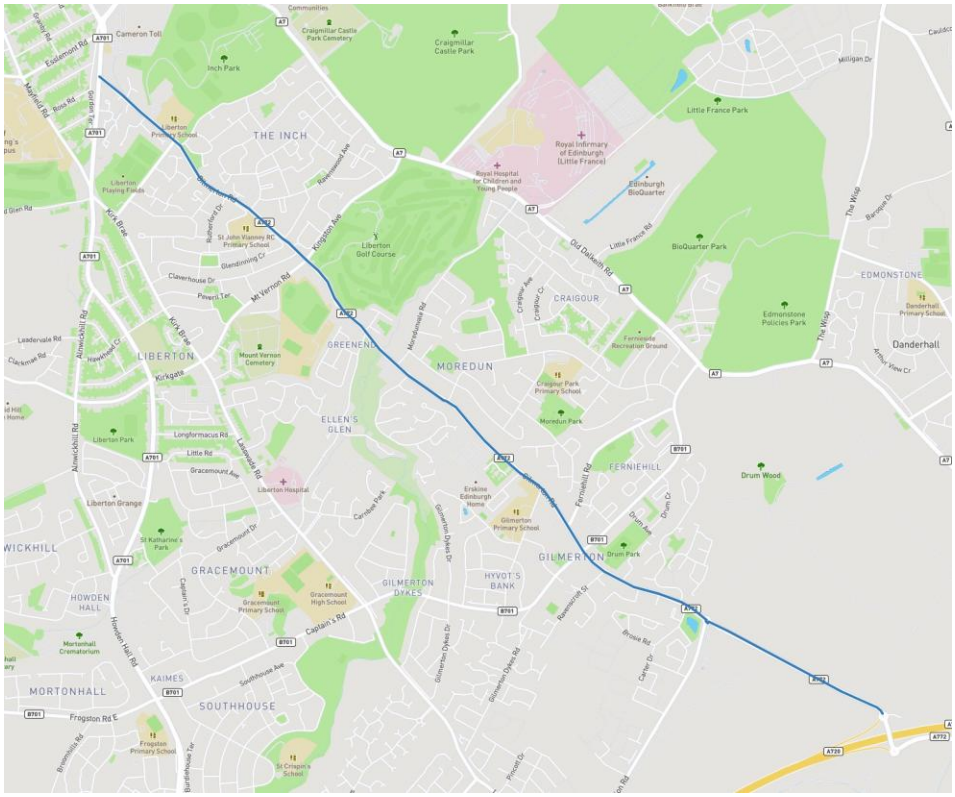



Figure 4.1: Location of Gilmerton Road.

### 4.1.1 Manual inspection

Due to data availability the data considered are weekdays starting from August 2023. The average total travel time along the road towards the city centre is between 7 and 8 minutes depending on time of day. The distributions of travel times per hour is skewed to the right.

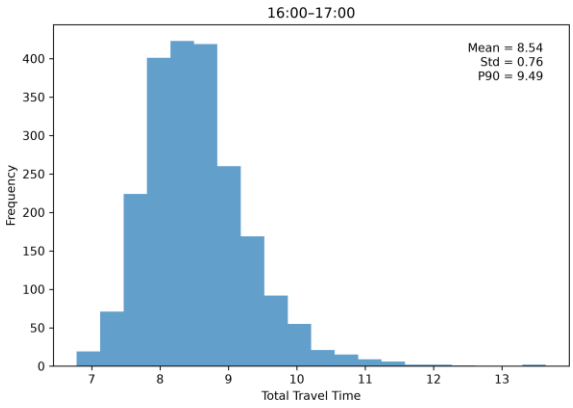



Figure 4.2: Histogram of average total travel time (in minutes) along Gilmerton road between 4 and 5pm on weekdays.

In a first instance we looked for outliers in the data, i.e. time periods where the average travel

time was unusually long compared to what it would usually be at that time of day (more than three standard deviations away from the mean). A manual comparison with the roadworks records did not bring up anything. Apparently none of the roadworks led to unusual delays that were apparent on INRIX. The fact that INRIX discards data of cars that are stationary for too long based on the assumption that they are parked might be obfuscating the impact that roadworks have.

We found a striking example of that when trying to understand outliers. On the 12th June 2024 there was a crash on Liberton road leading to a road closure. According to news reports the crash happened around 12.30pm leading to significant queues due to diversions in the surrounding area including Gilmerton road. [1] The travel time data (see 4.3) available for midday only shows delays of a 1-2 minutes relative to average. There is data missing between 1 and 2.30pm.

It is only around 4-5pm that delays increase significantly. The speed data for the afternoon reflects stuck traffic between 4 and 5 pm but not before. We would surmise that the queues and delays mentioned in the news were not visible in the system due to data being discarded. Interestingly, the missing data between 1 and 2.30pm arises from only one segment missing data. This is the segment located next to the incident so one could surmise that this segment is missing data because so many cars were stuck and there was not enough traffic flow to register. The congestion that was visible in the afternoon might have been the incident slowly dissolving combined with an influx of rush hour traffic.

[1] News report

| Date Time | Total Time |
|---|---|
| 2024-06-12 12:00:00 | 8.03 |
| 2024-06-12 12:15:00 | 8.73 |
| 2024-06-12 12:30:00 | 8.55 |
| 2024-06-12 12:45:00 | 8.99 |
| 2024-06-12 13:00:00 | 8.76 |
| 2024-06-12 14:30:00 | 10.42 |
| 2024-06-12 14:45:00 | 9.09 |
| 2024-06-12 15:00:00 | 9.68 |
| 2024-06-12 15:15:00 | 9.48 |
| 2024-06-12 15:30:00 | 9.17 |
| 2024-06-12 15:45:00 | 8.98 |
| 2024-06-12 16:00:00 | 11.47 |
| 2024-06-12 16:15:00 | 10.33 |
| 2024-06-12 16:30:00 | 11.51 |
| 2024-06-12 16:45:00 | 11.50 |
| 2024-06-12 17:00:00 | 10.08 |
| 2024-06-12 17:15:00 | 9.32 |
| 2024-06-12 18:45:00 | 7.60 |
| 2024-06-12 19:00:00 | 7.09 |
| 2024-06-12 19:15:00 | 6.74 |

Figure 4.3: Total travel time along Gilmerton road on 12th June 2024 after 12pm. There is data missing between 1-2.30pm. Source: INRIX

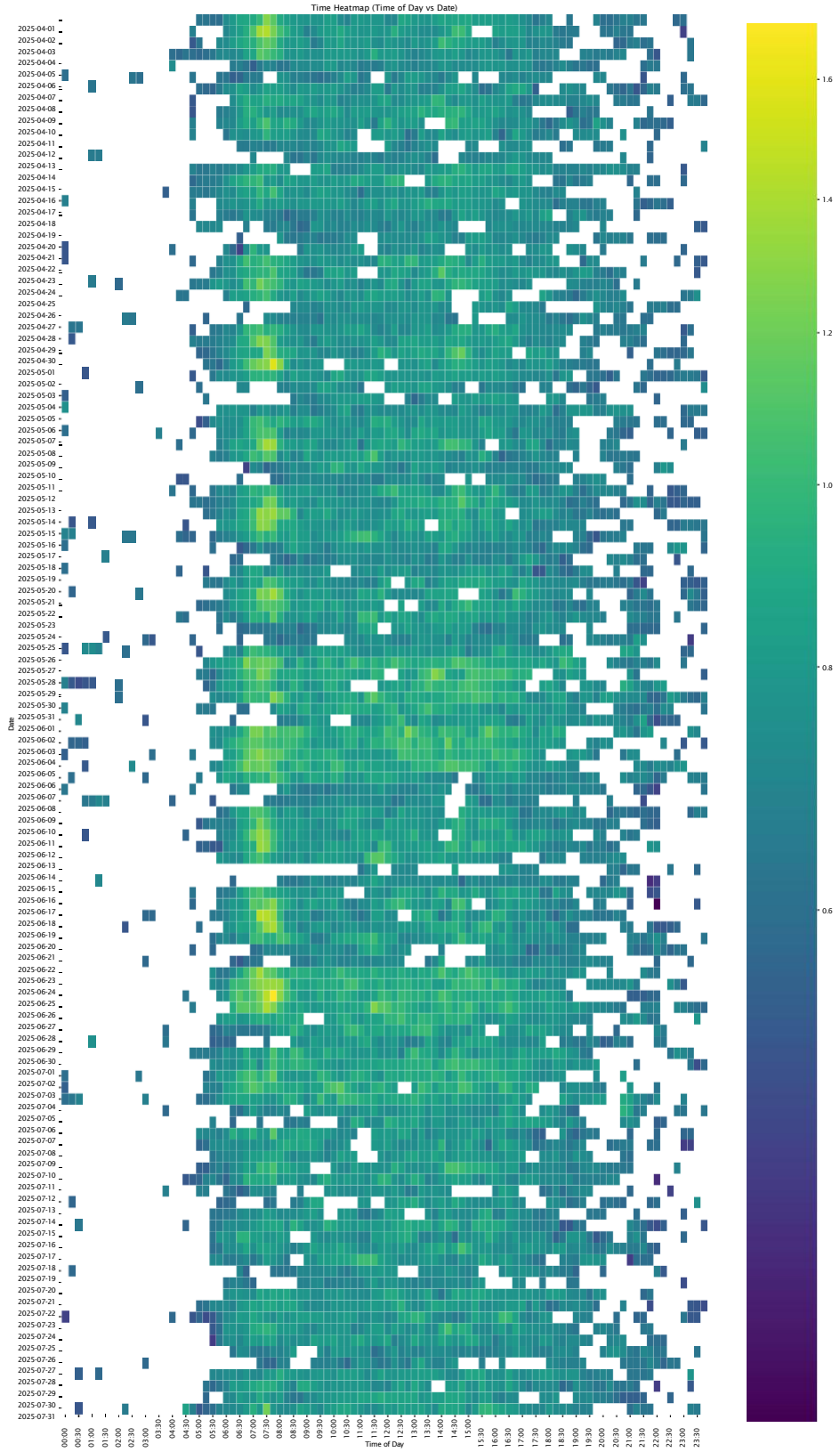



Figure 4.4: Plot of 15 minute interval travel time (as a percentage of average travel time per time of day) on a segment of Gilmerton Rd. between Apr1i6l and August 2025. Blank cells signal lack of data. Travel times Data source: INRIX.

### 4.1.2 Anomaly detection methods

If roadworks cause a disturbance then this should lead to unusual traffic behaviour. Our next approach is to detect anomalies i.e. unusual traffic behaviour and to compare them with recorded roadworks. For our investigations, take Ferry road in Edinburgh as a case study.

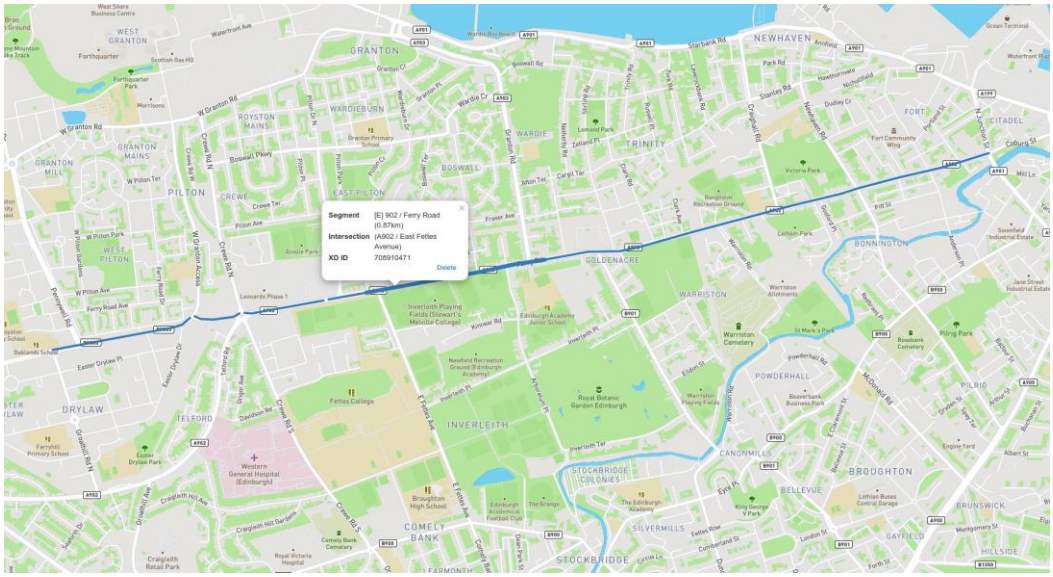



Figure 4.5: Ferry Road in Edinburgh. Data below is presented for the highlighted segment

Since traffic fluctuates heavily throughout the day and across different segments of a road, we limit our analysis to individual segments, as provided by INRIX. We present results for segment with ID 708910471.

As input take real time traffic data split into 15-minute bins between 07:00 - 22:00 for weekdays between 1 September 2023 and 24 June 2026. For every timeslot the following metrics are recorded per segment:

- Travel time
- Average speed
- Historical average speed

The goal is to detect time intervals during which these metrics deviate significantly from their 'usual' behaviour.
There are two approaches. The first one is purely statistical, based on deviation from a baseline measure of travel time along each segment. The second one is isolation forest which is an unsupervised learning method.

#### Statistical anomaly analysis

Statistical scoring is as follows: fix a segment. For every 15-minute timeslot, look at the travel times during the same timeslot on the same day in the past 60 days. Compute the following values:

- median travel time

- A z-score, given by the formula

$$\text{z-score} = \frac{\text{75th percentile value} - \text{25th percentile value}}{1.349}$$

  Both percentile values are computed from the same day of week, same timeslot data over the past 60 days.

The z-score is also known as pseudo-standard deviation; the 1.349 factor comes from the fact that the interquartile range for a standard normal distribution is 1.349. This is more robust to outliers influencing the standard deviation.
[2]

Based on this, define the anomaly score of a 15-minute timeslot by

$$\text{anomaly score} = \text{abs}\left(\frac{\text{observed} - \text{median}}{\text{z-score}}\right)$$

where abs denotes the absolute value. This measures how far the observed travel time is from the median, in terms of standardized units. The final output is anomaly percentile for that segment:

$$\text{anomaly percentile} = \text{percentile of anomaly score across all anomaly scores}$$

(here, anomaly scores are taken across all timeslots). This is meant to further normalize the anomaly score, and is used to find the top 5% most anomalous time intervals.

**Segment-level results** For segment 708910471 there is a heatmap of the anomaly percentile, and a heatmap of the top 5% most anomalous 15-minute timeslots

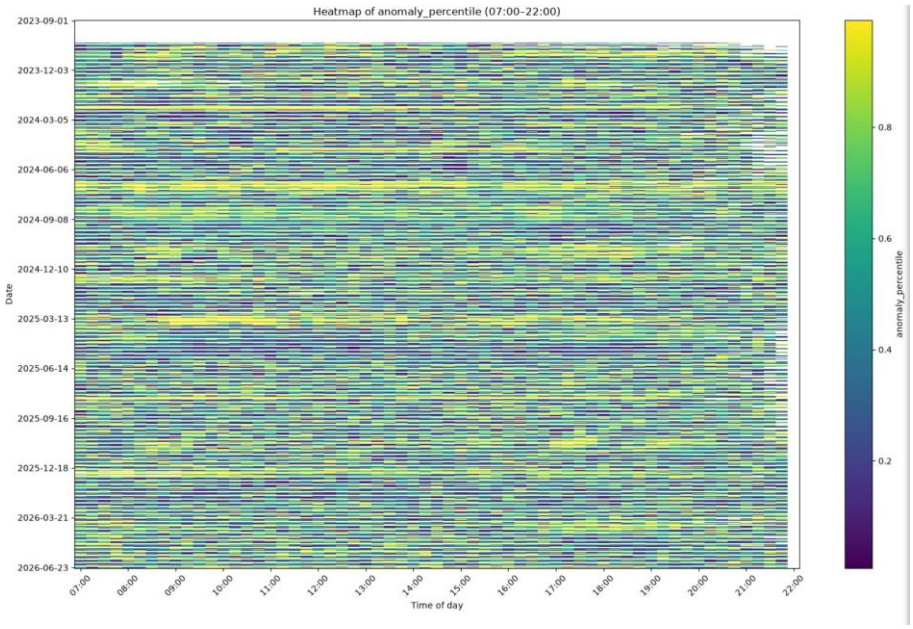



Figure 4.6: Heatmap of anomaly percentile. Yellow means more anomalous. Notice that there are a few longer yellow periods.

[2]Since our travel time data is right-skewed, and our quartile range is symmetric, our version of anomaly detection is skewed towards flagging unusually high (as opposed to low) travel times. We have tried other scoring methods, such as Median Absolute Deviation (MAD) [6] or computing the above z-score on the log distribution, but these performed significantly worse at matching with roadworks on the single segment, and had similar performance on the entire road.

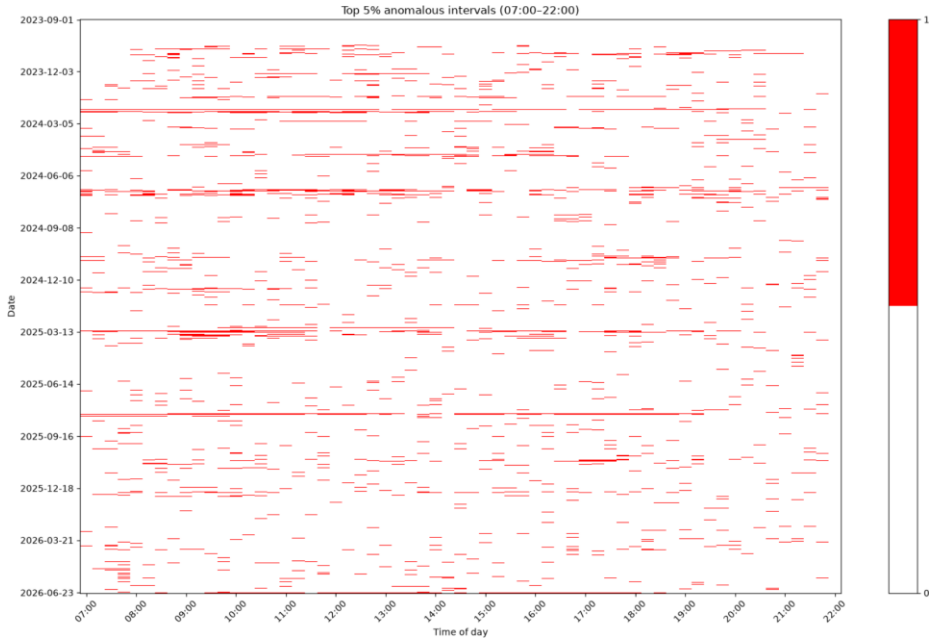



Figure 4.7: Top 5% most anomalous timeslots in red. Some are isolated and some cluster around certain days.

In order to detect longer anomalous periods, track periods of at least 3 consecutive timeslots whose anomaly score is in the top 5%:

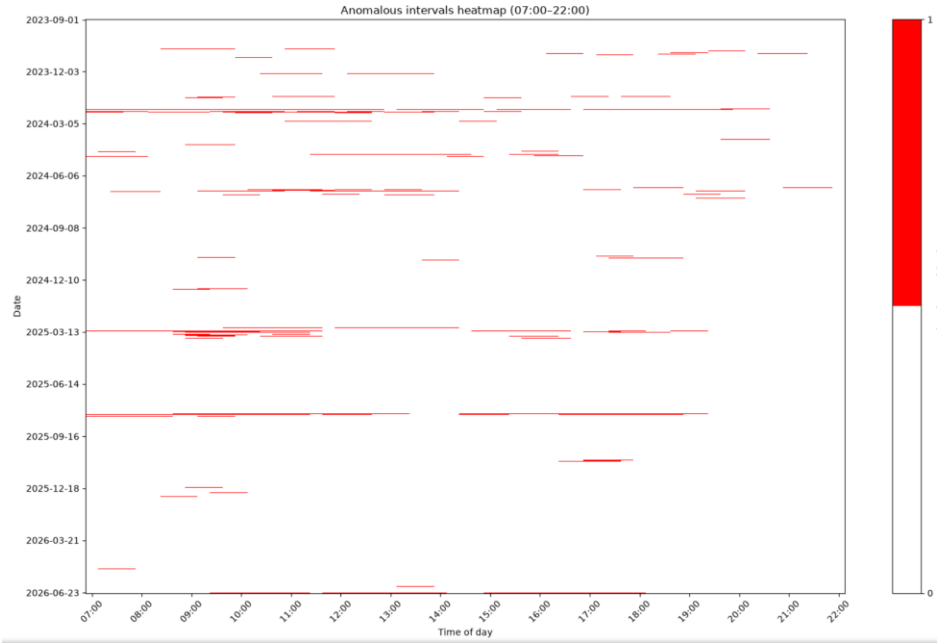



Figure 4.8: Periods of at least 3 consecutive top 5% anomalous interval.s

Compare longer anomalous periods against roadwork data for the entire road. Count how many

anomalous periods intersect with a roadwork with traffic-impacting travel management.

Table 4.1: Anomalous periods to roadworks

| # of periods | # intersecting works | # not intersecting works |
|---|---|---|
| 125 | 92 (73.60%) | 33 (26.40%) |

For completeness, plot a combined graph of roadworks (as determined by actual start and actual end metrics in the Register) and anomalous periods:

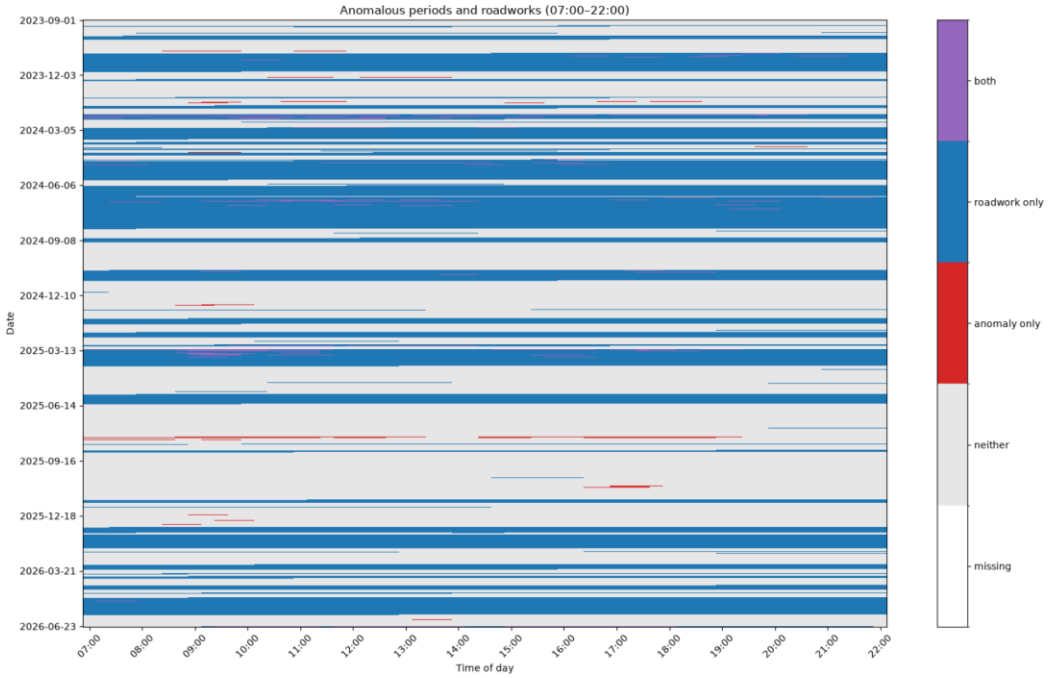



Figure 4.9: Combined graph of anomalous periods and active roadworks

**Road-level** To pass to road level repeat this analysis, but sum up travel times for all segments on the road. The definition of the anomaly score and percentile remain the same. We obtained the following distribution of anomalous intervals:

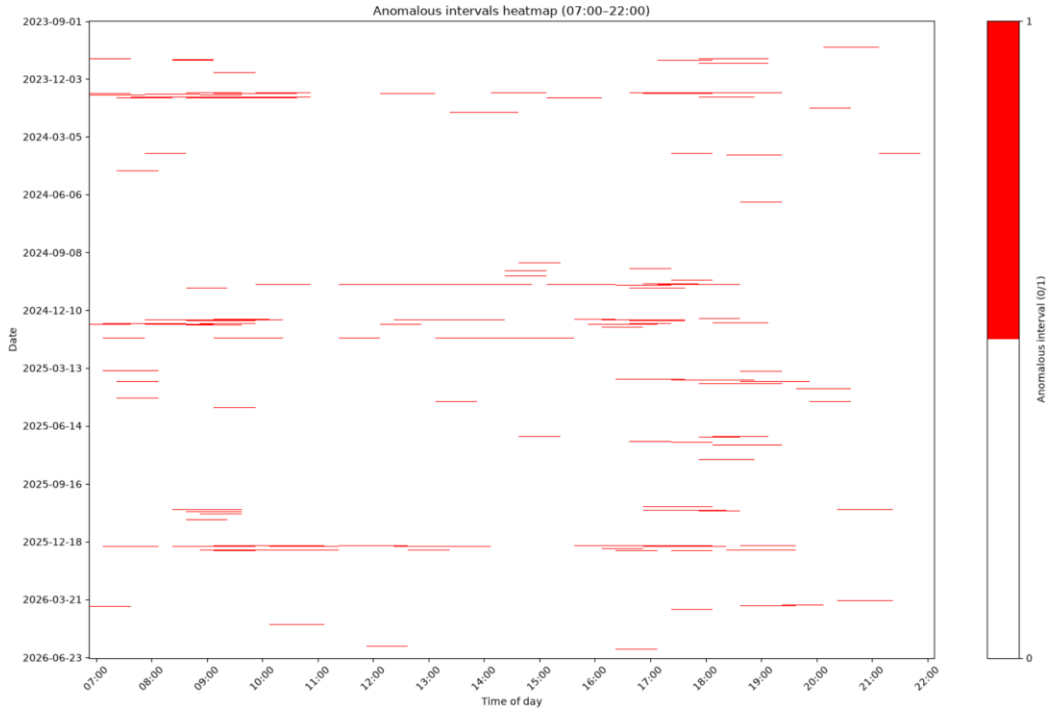



Figure 4.10: Entire-road anomalies

The matching at the level of the entire road performed worse - only 30% of works intersected with an anomalous period:

Table 4.2: Anomalous periods to roadworks, entire road

| # of periods | # intersecting works | # not intersecting works |
|---|---|---|
| 160 | 47 (29.38%) | 113 (70.62%) |

### Isolation Forest

Isolation Forest is an unsupervised machine learning algorithm that detects anomalies in a data set $X$ [3]. Unsupervised means that one does not need to know in advance what anomalies look like. This makes it a good approach for this problem as a priori one does not know what traffic disruptions due to roadworks look like.

Isolation forest works as follows: A finite list of features $F_i$ is fixed to begin with. Then an isolation tree is created. At the $i$-th step a feature $F_i$ as well as a split value $s_i$ between for the selected feature are chosen at random and the data is partitioned into two sets according to the split value. This procedure is repeated until all data points are individual leaves on the tree (or a maximum depth is reached). The path length of a data point $x$ is the number of steps required until $x$ is in a set by itself. The more anomalous a data point $x$ is relative to the other points in the sample the shorter its path $h(x)$ length will be.

[3]For full details, see [4]

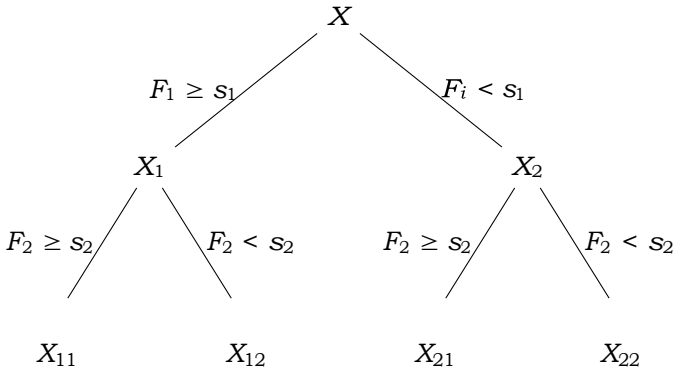



Figure 4.11: The first two steps of an isolation tree. At the first step, a feature $F_1$ (e.g. travel time) and split value $s_1$ are chosen. The set $X_1$ contains those data points that have $F_1$ larger than $s_1$. At the second step, a feature $F_2$ and split value $s_2$ are chosen. The set $X_{11}$ contains those data points of $X_1$ that have $F_2$ larger than $s_2$. We continue until the sets contain a single point. Anomalous points tend to have shorter path lengths. For instance, if the set $X_{12}$ consists of a single data point $x$ then the path length $h(x)$ is equal to 2.

This isolation tree is created a number of times (we chose 200). Since the features and splits are chosen at random these yield different partitions of the data. The anomaly score of a data point $x$ is calculated from the average of path lengths across the 200 isolation trees. In a dataset of size $n$ the average path length is

$$c(n) = 2H(n-1) - 2\frac{(n-1)}{n}$$

where $H(i)$ is the $i$-th harmonic number. This is used to normalise path lengths. The anomaly score of a data point $x$ in a data set of size $n$ is

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}}$$

where $E(h(x))$ is the average path length $x$. An anomaly score $s(x, n)$ close to 1 indicates anomalies, $s(x, n)$ close to or below 0.5 indicates 'normal' behaviour (relative to the data set). Experiments were run using Scikit-Learn.

**Segment-level results** For every segment and every 15-minute time interval on a weekday we want to find anomalous behaviour. The following features were used:

- average travel time to traverse the segment
- average speed
- difference between average speed and historical average speed
- ratio between average speed and historical average speed

The 5% most anomalous segment time intervals were designated as anomalies. In order to get a notion of incident, we only considered instances where at least 3 consecutive time intervals (i.e. minimum 45 minutes) were flagged as anomalies. The results are plotted in a heatmap with colours indicating how many segments were affected simultaneously.

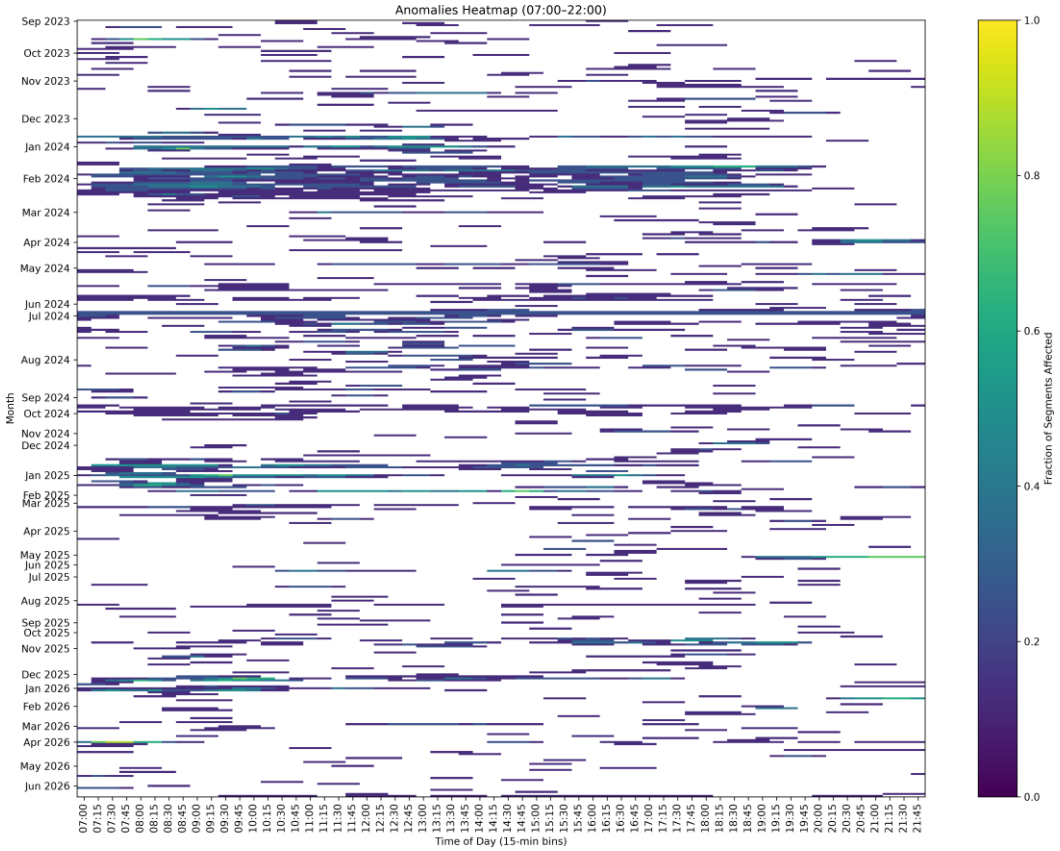



Figure 4.12: Heatmap of longer anomalous periods ($\geq$45 mins). Colours denote number of segments affected.

We also obtained a list of longer anomalous periods affecting at least three segments of road simultaneously. This yielded 78 periods in total. By inspection we determined that a number of them occurred on bank holidays or during the Christmas/New Year's period. Other events that were visible were storm 'Eowyn' in January 2025 as well as a major accident. After removing these we were only left with a handful of anomalous periods for which we could not find an external explanation. Some of them might be related to roadworks but drawing a causal relationship proved impossible.

We concluded that requiring several segments to be anomalous simultaneously was too restrictive to capture day-to-day traffic behaviour and that we would be left only with extreme incidents. Therefore we tried an other experiment to capture anomalies along the entire road.

**Road-level** An alternative notion of road-level incidents used as features the total travel time along the road and congestion (measured as percentage of historical average speed) per segment. Total travel time was weighted more strongly as a feature as this was deemed the most impactful predictor. For every 15-minute interval the total travel time is compared all other occurrences of the 15-minute interval and congestion per segment during a 15-minute interval is compared to congestion on the same segment during the same 15-minute interval. This is to ensure that temporal and spatial differences in traffic behaviour do not flag as anomalies. The fifth percentile of anomaly scores was used to determine anomalies i.e. the 5 % most anomalous scores are declared to be anomalies.

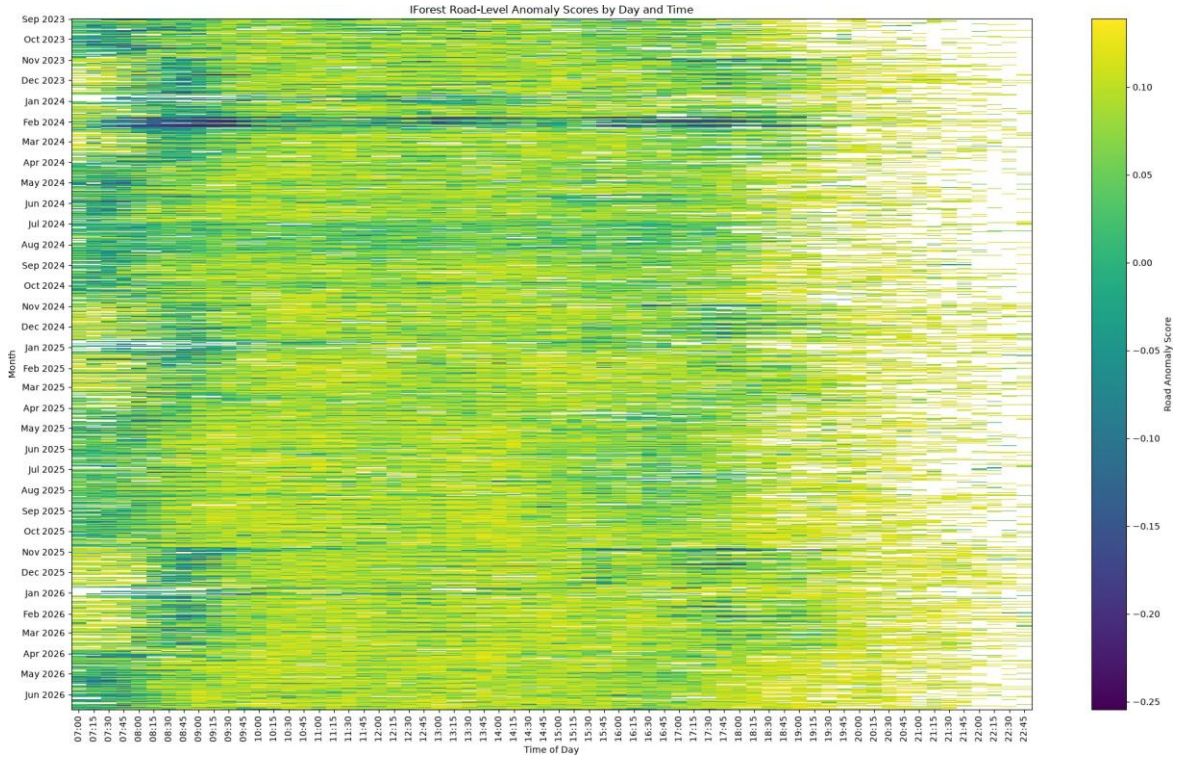



Figure 4.13: Heatmap of anomaly scores on road-level. Darker colours indicate anomalies.

Comparing metrics between anomalous and normal segments showed that the mean speed and mean travel time per segment are about the same between normal and anomalous segment. However the standard deviation of speed and travel time is larger indicating that anomalous segments have a greater spread of values for both speed and travel time.

As before we obtained a list of longer anomalous periods. These contained about 85% of the anomalous periods from the previous experiment. After filtering out the Christmas/New Year's period we were left with about 150 anomalies. Comparing with the roadworks data, we found that 54% of longer anomalies happened whilst a roadwork was recorded on the road, whereas 25% of roadworks intersected with at least one longer anomaly. However this does not establish a meaningful causal relationship between detected anomalies and roadworks.

**Discussion** The anomaly detection methods proved effective in detecting fundamentally different traffic behaviour (e.g. bank holidays, major incidents or extreme weather). These experiments show that the INRIX data contains enough detail to reflect major changes in traffic behaviour. There are many complex factors that may impact traffic and the most anomalous behaviour might not be related to roadworks at all. Even if an anomalous period intersects with a roadwork this is not proof that the roadwork caused the anomaly. An other issue is that how much of the data is labelled as anomalous is decided in advance (in our case 5%). Therefore we would need to estimate the prevalence of disruptions from roadworks and find a way of separating these out from other types of disruptions.

Next we move on to supervised learning. These are methods where the model tries to establish a relationship between patterns in traffic data and roadworks.

### 4.1.3 Random Forest

When trying to determine correlations between traffic data and roadworks using advanced learning methods, one important concern is *interpretability*, i.e. understanding which input features the model uses to make its predictions. Random forests are a highly interpretable class of models used for supervised classification.

#### Training set-up

We implement a Random Forest Classifier as provided by scikit-learn to predict from INRIX data along Ferry road whether a traffic-impacting roadwork is happening on some segment of the road. Prepare the raw data as follows:

- INRIX data along Ferry Road Edinburgh ( see 4.5) between 7AM - 22PM during the period August 2023 - June 2026. Drop any segment that has a density of measurements less than 0.8.
- Roadworks data from this period pulled from the register. Restrict to roadworks with both actual start and actual end time and with traffic management that is likely to impacting roadworks, leaving a total of 118 roadworks.
- Encode the geometry of the INRIX segments by hand (as linestrings) for location information.

Using the WKT geometry data from the register roadworks are matched to segments. The (input, output) pair for the classification task was structured as follows:

- Input: for every (segment, day) pair, a list containing 184 entries including:
  - All measured speeds and times on that segment on that day
  - A *mask* for every timeslot (1 or 0, depending on whether there was a recorded measurement for that timeslot)
  - Day of the week, one-hot encoded (i.e. an entry for every day of the week, where the actual day is set to 1 and the rest are set to 0)
  - Segment, one-hot encoded.
- Output: a single bit (0 or 1) depending on whether there was an active roadwork on the segment during that day or not.

#### Model description

The random forest has 400 estimators, unbounded max depth and square root max features. The model works as follows:

- Initialise 400 decision trees. Each of them sees the same number of training samples but sampled randomly with replacement, so each tree has a slightly different training set.
- Each decision tree picks, at each node, around $\sqrt{184} \approx 13$ features at random. It tests a number of thresholds for each feature, and splits on the one which is "best" at distinguishing roadworks from no roadworks.
- The positive class in the training data is low: only 4.2% of samples are roadwork-positive. This is taken into account when splitting: the trees are incentivised to try to separate out the positive class as much as possible.
- Each tree keeps growing until it is stopped by restraints. The end output is that the tree has separated the input data into a number of subsets (its "end leaves") which in a perfect world would be as pure as possible (i.e. all positive or all negative). However, in practice, this is not always the case.

After training, when the forest is presented with a new sample, each of the 400 trees makes a decision and places the sample in one of their end leaves, reporting the percentage of positive roadwork samples in that leaf. These percentages are averaged across all trees. The result is a single percentage, the vote of the random forest on how likely the sample is to be a roadwork. In order to decide whether the sample is declared to be a roadwork, a *validation* set is used. This is an unseen portion of data which the forest is given and a threshold percentage is chosen in order to maximise the number of accurate predictions on the validation set.

**Results**

Use the most recent 30% of data for each segment for testing. Of the remaining 70%, the latest 20% are for validation and the remaining for training. The threshold is chosen using an $f_\beta$ score of 2 (meaning that recall is weighted over precision - this is due to the low prevalence of the positive class on validation). The test data consists of 1760 samples with a positive roadwork prevalence of 4%. The resulting statistics are :

| Statistic | Score |
|---|---|
| Precision | 0.078 |
| Recall | 0.743 |
| F2 | 0.2750 |
| PR-AUC | 0.0727 |
| ROC-AUC | 0.7125 |

The "confusion matrix" for predictions on test:

| | Predicted no roadwork | Predicted roadwork |
|---|---|---|
| No roadwork | 1037 | 649 |
| Yes roadwork | 19 | 55 |

The model correctly identified 55 out of 74 roadwork-positive days as actually having a roadwork, but at the cost of identifying roadworks on 649 other days on which there was no roadwork happening. The PR AUC score is 1.72 times larger than the baseline prevalence, which indicates that the forest consistently outputs higher votes for roadwork days than for non-roadwork days. The forest is detecting a non-trivial relationship, but it is weak.

**Feature importance and interpretability**

Feature importance is an other outcome of the random forest model. Each input feature is assigned *feature importance*, a number which indicates how useful this feature is for splitting the data into positive / negative class across all trees in the forest. The cumulative importance per feature class in our forest is:

| Feature type | Cumulative importance |
|---|---|
| Travel time | 0.572647 |
| Speed | 0.350788 |
| Segment ID | 0.069124 |
| Mask | 0.005245 |
| Day of week | 0.002195 |

Travel time and speed related features clearly dominate importance, with travel time features being significantly more important than speed. The top 15 most important times are:

| Time of day | Importance |
|---|---|
| 09:45 | 0.027713 |
| 18:00 | 0.026855 |
| 08:00 | 0.025030 |
| 07:15 | 0.024390 |
| 17:45 | 0.024192 |
| 17:30 | 0.023333 |
| 07:45 | 0.022330 |
| 12:00 | 0.022148 |
| 09:30 | 0.021238 |
| 17:15 | 0.020691 |
| 19:00 | 0.020156 |
| 14:00 | 0.019604 |
| 17:00 | 0.019090 |
| 10:30 | 0.018718 |
| 07:30 | 0.018120 |

These concentrate around morning (8-9 AM) and evening (17-18 PM).

**Discussion**

Feature importance is the most conceptually interesting outcome of the random forest: it confirms the intuition that the impact of roadworks is most prevalent around rush hour, and manifests itself predominantly as increases in travel time. The random forest also detects some signal, being able to consistently rank roadwork days higher than non-roadwork days.
The difference in roadwork day prevalence between training, validation and test sets is likely affecting the experiment. This could be remedied by training on a larger dataset, possibly over a bigger region with more streets.

### 4.1.4 Transformer-based methods

Traffic data comes as a sequence of observations that change over time. When detecting anomalies we want to take this time-dependence into account, i.e. look at how speed and travel time change over time. Transformer architectures are very well-suited for learning short-term as well as long-term patterns in data. Transformers contain attention layers, information does not just traverse the network it can be passed back (self-attention, see for instance [7]) which allows the model to learn a sequence of observation.

**Transformer**

The set-up of traffic data and roadworks is the same as for random forest. The last 30% are for test, and the last 30% of training are for validation. Speed and travel time are on different scales, which is a possible issue for transformer-type architectures. Hence z-normalize speed and travel time per timeslot. That is, for every timeslot (say 9:00 AM), compute mean speed and travel time as well as their standard deviations and set

$$\text{standardized}_{9:00\ \text{AM}} = \frac{\text{actual} - \text{mean}_{9:00\text{AM}}}{\text{st. dev}_{9:00\ \text{AM}}}$$

for speed and travel time (and similarly for the other tine slots). Fill any missing slots with 0. These statistics are computed only for the training set; validation and test samples are normalized from the training statistics.

Explaining the structure of a transformer is beyond the scope of this report (see for instance [7]). The main structural features in our implementation are:

- A log loss function, weighted appropriately to offset the positive roadwork days imbalance in the training data. [4]
- Model parameters are:

| Model parameter | Value |
|---|---|
| Embedding dimension | 32 |
| Number of layers | 3 |
| Attention heads (per layer) | 2 |
| Dropout rate | 0.1 |

This is a relatively small model chosen by sweeping on the embedding dimension, number of layers and dropout rate (i.e. trying multiple parameters for each), training 10 different instances of the model, and choosing the architecture which maximizes the PR AUC score on validation with $\beta = 2$. The above architecture achieved the following values over 10 runs with randomized seeds:

| Statistic | Score (mean $\pm$ std) |
|---|---|
| Precision | $0.0775 \pm 0.0197$ |
| Recall | $0.7770 \pm 0.2463$ |
| F2 | $0.2557 \pm 0.0322$ |
| PR-AUC | $0.1618 \pm 0.0350$ |
| ROC-AUC | $0.7437 \pm 0.0433$ |

**Discussion**

The statistics are similar to the random forest, with one notable exception: the PR-AUC score has increased dramatically (2.2x), and this increase is stable across all 10 runs (with a std dev of 0.035). This means that the transformer model is better at ranking actual roadwork days above non-roadwork days in the test data than the random forest. It is difficult to pinpoint exactly where this improvement comes from, since transformer models are not interpretable. Despite the PR-AUC increase, the correlation detected by the model is still too weak to be useful to determine the impact of roadworks on traffic. It is unclear whether this could be improved by training on more data.

**Gated Transformer Network**

Gated Transformer Networks have been used for multivariate time series classification problems [5]. The time series data we are given contains two variables (average speed and travel time) which are related. A GTN learns both the time dependence of the variables as well as the relationship between the variables at every time step.

GTNs are supervised learning algorithms i.e. start with labelled traffic data. For every 15 minute time interval on every road segment there is a binary variable that encodes whether or not there is a roadwork nearby. Roadworks were mapped onto the nearest two segments. We only considered roadworks of traffic management type: Portable Traffic Lights, Give and Take Traffic Control, Road Narrowing, Road Closure, Lane Closure as they are the most potentially impactful. Traffic data is restricted to between 7am and 10pm on weekdays. About 17% of time intervals contain a roadwork which is a class imbalance.

[4] Another option would be to artificially replicate the positive samples in the training data until equal prevalence is reached equal prevalence, but we did not attempt this.

A Gated Transformer Network with 3 layers and 4 attention heads is coded from PyTorch. The transformer considers the previous 2 hours of time and speed observations in order to predict whether or not a roadwork was active. Features used: time of day, day of the week, travel time, average speed, speed ratio (average speed as ratio of historical average speed, a measure of congestion) and location (segment ID). Due to the class imbalance in the training set the model is penalised for mistakes on roadwork events.

The model is trained on a temporal train-validate-test split and training stops when validation loss stops decreasing in order to avoid overfitting. The model produced the following metrics:

| Statistic | Score |
|---|---|
| Precision | 0.192 |
| Recall | 0.394 |
| F1 | 0.258 |
| PR-AUC | 0.205 |
| ROC-AUC | 0.587 |

The ROC-AUC indicates that the model is only slightly better than guessing randomly (that would have a ROC-AUC of 0.5). The PR-AUC is 16% above baseline which is the positive rate 0.177. This indicates that the model learns something but the signal is not very strong.

## 4.2 Incorporating traffic counts

In addition to travel time and speed, traffic counts are a useful metric of traffic behaviour. In this section we investigate whether adding traffic counts in the model improves predictions. As a case study consider Great Western Road in Glasgow since it is a major traffic corridor and has traffic counters in both directions. We use the UBDC traffic counts dataset and INRIX traffic data during the period 10/2019-10/2023. This period includes Covid which had an impact on both roadworks and traffic behaviour so the results are not expected to be particularly accurate. The point is to determine whether the traffic counts improve relative model performance.

We look at eastbound traffic on Great Western Road between Anniesland cross and St. George's cross. On INRIX this part of the road made up of 4 segments. There are 11 traffic counter sites along this road but we select one per segment, the east-most (counting the number of vehicles entering the segment).

Traffic counts are hourly, whereas we use 15min traffic data (as the hourly INRIX data is too sparse). Assign the same traffic count to all 15min intervals within the hour. Time intervals that do not have a traffic count assigned to them are discarded. We consider time intervals between 7am and 10pm.

In order to create labelled training data 4.15 each roadwork on Great Western road during this time is mapped to the nearest segment and a binary roadworks column is added to the traffic data file. This column is 1 if a roadwork was happening on a segment during a given 15-minute interval and 0 otherwise.

To learn patterns in the data we use a random forest classifier and extreme gradient boosting (XGBoost). The XGBoost classifier learns temporal patterns in the data by looking at every observation within the context of the previous 2 hours of observations.

**Results** The random forest classifier had the following results

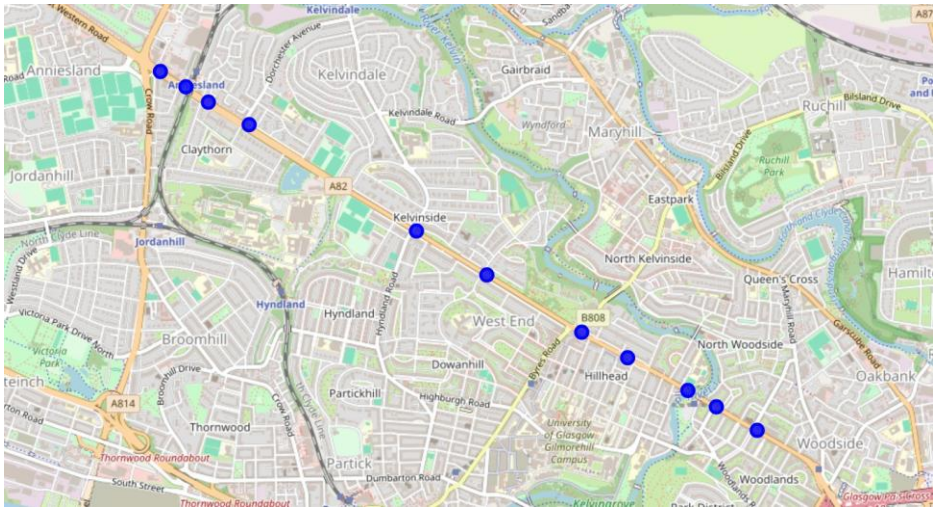



Figure 4.14: Traffic counters along Great Western Road.

| Date Time | Segment ID | Speed(km/ hour) | Hist Av Speed(km/ hour) | Ref Speed(km/ hour) | Travel Time(Minutes) | Corridor/Region Name | flow | roadworks |
|---|---|---|---|---|---|---|---|---|
| 2023-09-26T07:00:00Z | 708897730 | 25.0 | 19 | 32 | 2.22 | Great Western Road Eastbound | 226.0 | 1 |
| 2023-09-26T07:15:00Z | 708897730 | 26.0 | 19 | 32 | 2.22 | Great Western Road Eastbound | 226.0 | 1 |
| 2023-09-26T07:30:00Z | 708897730 | 24.0 | 19 | 32 | 2.33 | Great Western Road Eastbound | 226.0 | 1 |
| 2023-09-26T07:45:00Z | 708897730 | 22.0 | 19 | 32 | 2.55 | Great Western Road Eastbound | 226.0 | 1 |
| 2023-09-26T08:00:00Z | 708897730 | 25.0 | 21 | 32 | 2.23 | Great Western Road Eastbound | 194.0 | 1 |

Figure 4.15: Head of the training data including both traffic counts and travel time/speed data per segment. The roadworks column indicates that there was a roadwork in the vicinity of the segment during this time. The model is trying to learn the relationship between the traffic data and the roadworks column.

| Statistic | Score |
|---|---|
| Precision | 0.46 |
| Recall | 0.48 |
| F1 | 0.47 |
| ROC-AUC | 0.683 |

and confusion matrix

| | Predicted no roadwork | Predicted roadwork |
|---|---|---|
| No roadwork | 24180 | 7453 |
| Yes roadwork | 6972 | 6344 |

The model correctly identified about half of active roadworks and for every correctly identified roadwork, there were 1.18 false positives.
Travel time and traffic counts are the two most important features. Adding in the traffic counts improved model performance.

XGBoost had the following results

| Statistic | Score |
|---|---|
| Precision | 0.50 |
| Recall | 0.59 |
| F1 | 0.54 |
| ROC-AUC | 0.732 |

and confusion matrix

| | Predicted no roadwork | Predicted roadwork |
|---|---|---|
| No roadwork | 18270 | 6168 |
| Yes roadwork | 4365 | 6157 |

The model identified about 60% of roadworks correctly and for every correct roadwork period it gave one false positive. The most important features are travel time and traffic counts and how they change over the 2 hours prior to an observation.

Traffic counts were important features for the models. This holds true across different model architectures. Therefore traffic flow appears to contain information that is relevant for the detection of roadworks. The best model achieved a ROC-AUC of 0.732. There is some predictive relationship between the traffic data and the roadworks variable but it is noisy. When there are traffic counts available they should be taken into account.

## 4.3 Predictive relationship between traffic data (INRIX) and bus disruptions

We obtained 4 months (1st of March to 30th of June 2026) of bus delay data from Lothian Buses for bus route 21. The data consists of

- A list of all bus stops, with names and coordinates;
- A list of all bus 21 trips that took place in the aforementioned interval. The data per trip consists of the stops visited (in order), the expected (i.e. pre-planned) time of arrival for each stop, the actual time of arrival and the actual time of departure.

Quantify bus disruptions by computing, for each pair of adjacent stops, the ratio between the actual time the bus took between stops and the expected time:

$$\text{delay ratio}_i = \frac{\text{arrival time}_i - \text{arrival time}_{i-1}}{\text{expected time}_i - \text{expected time}_{i-1}}$$

"Travel time" between stops is the difference between arrival times. This takes into account "dwell time", i.e. time the bus spent stationary at the previous stop. A more accurate travel time measure would be time of arrival at stop $i$ minus time of departure from stop $i-1$. When computing the delay ratio using this notion of travel time, we found that the mean delay ratio among all trips was 0.78, and the 75th percentile was 0.88. With the above definition of delay ratio, the corresponding mean value was 1.03, with 75th percentile 1.21, which is much more reasonable. This suggests that Lothian Buses take into account dwell time when coming up with their expected bus times.

For manual analysis, consider a bus to be disrupted between two adjacent stops if the above delay ratio was greater than 1.3, and severely disrupted if it was greater than 2. Bus disruptions tend to be localized along the bus route (i.e. there were not many examples of a bus being delayed at successive stops).

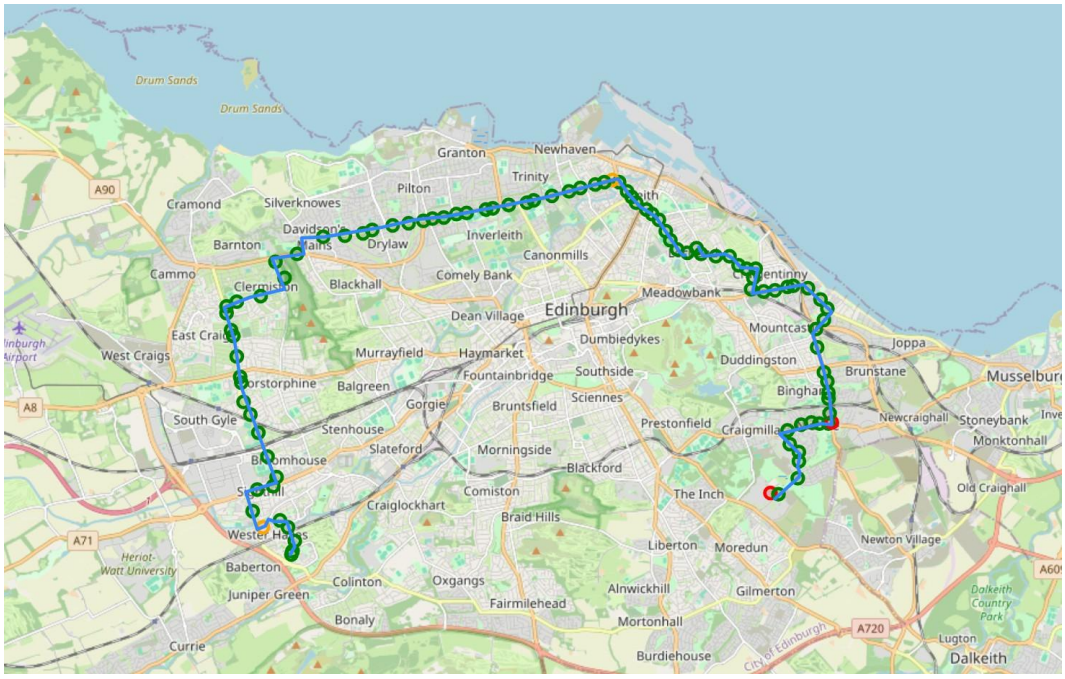



Figure 4.16: Example of a bus trip along route 21. A stop is green if its delay ratio is less than 1.3, yellow if it is between 1.3 and 2, and red if it is greater than 2.

We test whether the INRIX traffic data correlates with bus disruptions. Data gathering and preprocessing consists of

- Extracting INRIX traffic data along the bus route at 5 minute intervals. The route consists of 94 INRIX segments.
- Restricting the INRIX data and the bus data to 7AM - 22PM for better coverage, and dropping two segments with coverage less than 50%
- Restricting to bus trips going East to West (due to directionality of the INRIX segments)
- Assigning each bus stop to its nearest INRIX segment

There are many choices involved in setting up the classification problem. The set up used depends on two parameters window$\in$ {0, 1 }and threshold $\in${1.2, 1.3, 1.4, 1.5 }(window controls how large a window of time around a bus arrival we mark as disrupted, and we decide disruptions based on having delay ratios above the threshold).

- Input consists of all traffic data (speed, travel time, historic average speed) across an entire day on a single segment, with segment id and day of week one-hot encoded.
- Add a binary "bus passing" marker to every 5-minute slot in the input data. If window = 0, for every stop of a bus trip identify the 5-minute slot in which the bus arrived and mark that time interval on the corresponding segment **and** every segment between the stop and the previous one as "bus crossing". If window = 1 do the same and extend the marking to include the 5-minute timeslots before and after the bus crosses as well.
- Output consists is a single bit per timeslot marked positive for a bus passing. The bit is 1 if the delay ratio for that stop was greater than threshold, and 0 otherwise.

Fill in missing traffic data whenever necessary by repeating data from neighbouring cells (average data coverage was 96%). For each segment remove any days marked positive for road closures (this accounted to about 2% of training samples). Normalize the traffic data using a z-score. Split the data into training, validation and test batches. We compute the statistics (mean, standard deviation) for the z-score on the training data only, and apply them to normalize the validation and test sets.

For the classification task implement a transformer with parameters

| Model parameter | Value |
|---|---|
| Embedding dimension | 64 |
| Number of layers | 2 |
| Attention heads (per layer) | 4 |
| Dropout rate | 0.1 |

for a total of 100865 parameters. Sweep over window size and threshold. Penalize the loss function based on positive class prevalence in the test data.

**Results**

These are the results:

- Every model with window = 1 beat every model with window = 0.
- Best overall performance for threshold = 1.3.
- For the best-performing model, recall is 79% (the model correctly flagged 83% of bus delays) with a precision of 58% - about half of the delays are false alarms. The ROC AUC score is 0.74 - so the model is slightly better than random.

The PR AUC score increased consistently (relative to baseline prevalence in the test data) as the threshold increased (for the window = 1 models):

| Threshold | PR AUC lift |
|---|---|
| 1.3 | 2.06 |
| 1.4 | 2.29 |
| 1.5 | 2.61 |
| 1.6 | 2.89 |

This means that the model becomes better as we restrict to more and more severe bus disruptions.

**Discussion**

The traffic data has mild predictive power on bus delays, and severe delays are more predictable from the traffic data than the less severe ones. It would be interesting to see how these results change with more training data available. Further research into the bus delays as a measure of disruption would be helpful and beneficial to future AGQ CDT cohorts.

# Chapter 5

# A quantitative model for traffic sensitivity

In this chapter, we describe an approach for quantifying traffic criticality of roads in a city. This is not a new problem and has been considered at length. We took initial inspiration from [1] (traffic criticality in York) and [3] (similar problem for the city of Changchun, located in northeastern China).

At a minimum, any setup for large-scale traffic criticality determinations in a city requires the following data:

1. The road network of the city, usually in the form of a directed graph,
2. Free flow speed and travel time data for each road in the network,
3. A measure of traffic flow across the network.

Item 3 is usually given by an Origin-Destination (OD) matrix. This splits the city into a number of regions (OD zones) and for each pair of zones, counts the number of trips between them.

## 5.1 The simplest criticality model: All-or-Nothing

Given a network of roads, together with travel time information and an OD matrix for given regions we can model traffic across the network. The simplest way of doing this is all-or-nothing which is illustrated on the road network of St. Andrews:

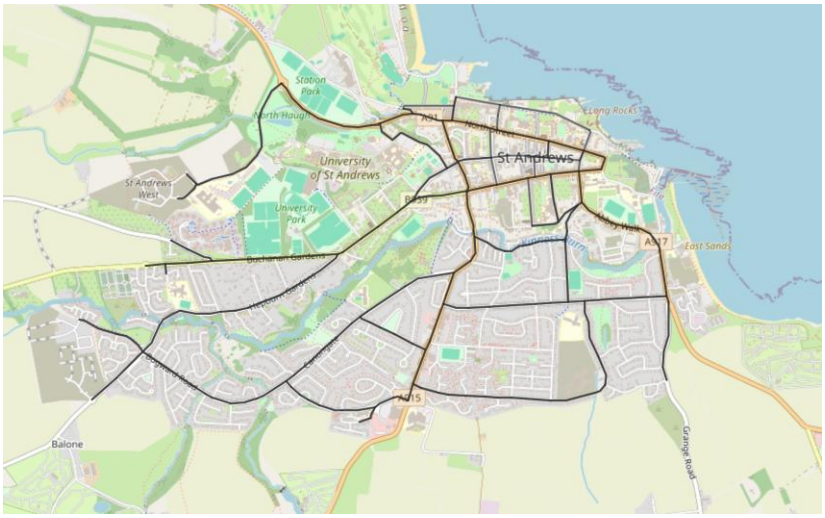



Figure 5.1: Simplified road network of St Andrews, extracted using OSMnx. We disregard residential streets as they do not matter for traffic flow across the network.

The graph is populated with speed limit values from OSMnx (serving as a proxy for free-flow speed), and there are 11 regions for the OD matrix, which are attached to nodes on the graph:




Figure 5.2: St. Andrews with OD regions

To assign traffic to the network use AON (All-or-Nothing) routing: for every pair of OD nodes, compute the shortest path in terms of travel time between the two nodes, and send all measured demand between these nodes down this route.

Using TomTom OD matrices computed during morning rush hour on 1 August 2024 we get the following:

**Algorithm 1** All-or-Nothing (AON) traffic assignment

**Require:** Input graph, edge travel time values, OD matrix $\mathbf{d}(i, j)$ with node assignment
  **for** each pair $(i, j)$ of OD nodes **do**
    shortestPath$_{i,j}$ ← shortest path between $i, j$
    **for** each edge $e$ in shortestPath$_{i,j}$ **do**
      $e$.flow ← $e$.flow + $\mathbf{d}(i, j)$
    **end for**
  **end for**

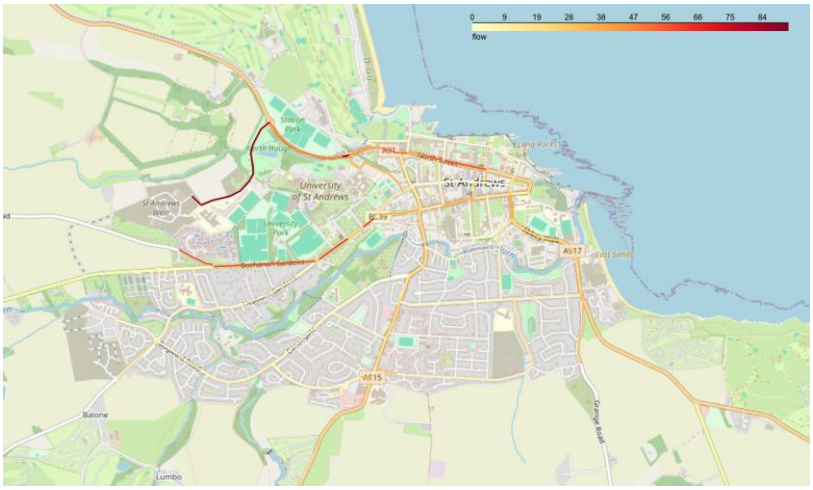



Figure 5.3: St Andrews network, colour-coded by demand

After performing this assignment, let $T_0$ be the total time taken by every road user to get from their origin to their destination. A simple measure of criticality of a road can be computed as follows:

- Remove the segment corresponding to a road $R$ from the graph,
- Recompute the shortest path between every OD pair,
- Compute $T_R$, the total time taken by road users after removing road $R$,
- The difference $T_R - T_0$ or the percentage increase $T_R/T_0 .100$ from the baseline total travel time serve as a measure of criticality.

Roads can be ranked by their criticality by computing $T_R$ for every road in the network. This rudimentary approach has some limitations:

- It is unrealistic to assume that all road users will take the shortest path, as this would cause massive congestion.
- If a road is not initially on a shortest path between two OD nodes, this measure of criticality is meaningless, since removing the road does not change the traffic assignment.

There are smarter traffic assignment methods available which invariably require more input data.

## 5.2 Traffic Assignment with the Frank-Wolfe Algorithm

Frank-Wolfe is a refined algorithm for traffic assignment. It is based on the fact that travel time along a road increases with congestion, and that this affects route choice. In addition to the input data for All-or-Nothing (it uses AON as a subroutine) it also requires:

- Road capacity data, i.e. how many cars per hour a road can support before congestion forms,
- A volume-delay function, which models how travel time increases as traffic volume increases.

If there are multiple routes between nodes $i$ and $j$, all users by default take the shortest route (in length). But as volume increases, so does travel time. This is modelled by the volume-delay function, the simplest being (for a road $a$)

$$T_{\text{congested},a} = T_{\text{free},a} \cdot \left(1 + \alpha \left(\frac{V_a}{C_a}\right)^{\beta}\right)$$

where $T_{\text{free},a}$ is the free flow travel time on road $a$, $V_a$ is the traffic volume on road $a$, $C_a$ is road $a$'s capacity (both in vehicles per hour) and $\alpha, \beta$ are model parameters.

As traffic volume along the shortest route increases it is no longer the most efficient route and part of the users switch to an alternative route in order to reduce their travel time. But then this means that travel time along the shortest route decreases (as volume reduces) and travel time along the alternative route increases (as volume increases). This leads to a complex relationship between travel time, optimal route and user choice.

We use the following implementation of Frank-Wolfe:

**Algorithm 2** Frank-Wolfe traffic assignment

**Require:** Graph $G$, OD matrix $\mathbf{d}$, volume-delay parameters $\alpha$=0.15, $\beta$=4, target gap $\varepsilon$
**Ensure:** Equilibrium flows $\mathbf{x}$

Compute free-flow travel times $t_a^0$ for all edges $a$
$\mathbf{x} \leftarrow$ All-or-Nothing($G, \mathbf{d}, \mathbf{t}^0$)
$k \leftarrow 1$
**repeat**
  **for** each edge $a$ **do**
    $t_a(x_a) \leftarrow t_a^0 \left(1 + 0.15 \left(\frac{x_a}{c_a}\right)^4\right)$
  **end for**
  $\mathbf{y} \leftarrow$ AON($G, \mathbf{d}, \mathbf{t}(\mathbf{x})$)
  $\text{gap}_k \leftarrow \frac{\sum_a t_a(x_a) \cdot y_a}{\sum_a t_a(x_a) \cdot x_a} - 1$
  $\mathbf{x} \leftarrow \mathbf{x} + 1/k \cdot (\mathbf{y} - \mathbf{x})$
  $k \leftarrow k+1$
**until** $\frac{1}{5} \sum_{j=k-4}^{k} \text{gap}_j < \varepsilon$
**return** $\mathbf{x}, k, \text{gap}_k$

First compute the all-or-nothing assignment: send everybody down the shortest path. Then re-compute travel times using the volume delay function with standard parameters $\alpha = 0.15$, $\beta = 4$, and recompute the shortest path with these travel times. Send half the people via the original shortest path, and send half on the new shortest path. The gap measures how far the current total network travel time is from the shortest travel time across the network (if everybody went down the fastest route). Through many iterations, the steps become smaller ($\frac{1}{k}$ at iteration $k$) towards the new shortest

solution. Run this until the mean of the last 5 gaps drops below a tolerance, or until a maximum number of iterations is reached.

Frank-Wolfe always converges to an equilibrium solution for the traffic flow across the network [8]. An network equilibrium is characterised by two properties:

- If more than one route between points $i$ and $j$ is assigned, then their travel time have to be equal, and
- There does not exist an other route between $i$ and $j$ that has lower travel time.

These conditions ensure that no user can gain anything by switching route and that the overall travel time is minimized. This is based on Wardrop's principles [8]:

1. Road users make choices to minimise their travel time (user-optimal)
2. Flows distribute across the network in such a way that the total travel time is minimised (system-optimal)

The latter is based on the observation that traffic participants seek to avoid congestion even if it leads to longer travel times.

#### Criticality scores

As before, one measure of criticality of a link is running Frank-Wolfe on the graph with that link removed and comparing the total network travel time to the baseline travel time. This is the measure used in our simulations. See [1] for a comparison of 5 different criticality measures, and [3] for further discussion.

## 5.3 Case Study: The road network in Glasgow

Due to its size and complexity, the Glasgow road network is an ideal testing ground for this method. The implementation involved many small design choices which we describe in the following.

#### Graph extraction

The graph is obtained from OSMnx. It is very complex and contains information on number of lanes, road names, OSM IDs for each segment, geometry in coordinates, road directionality, speed limit etc.

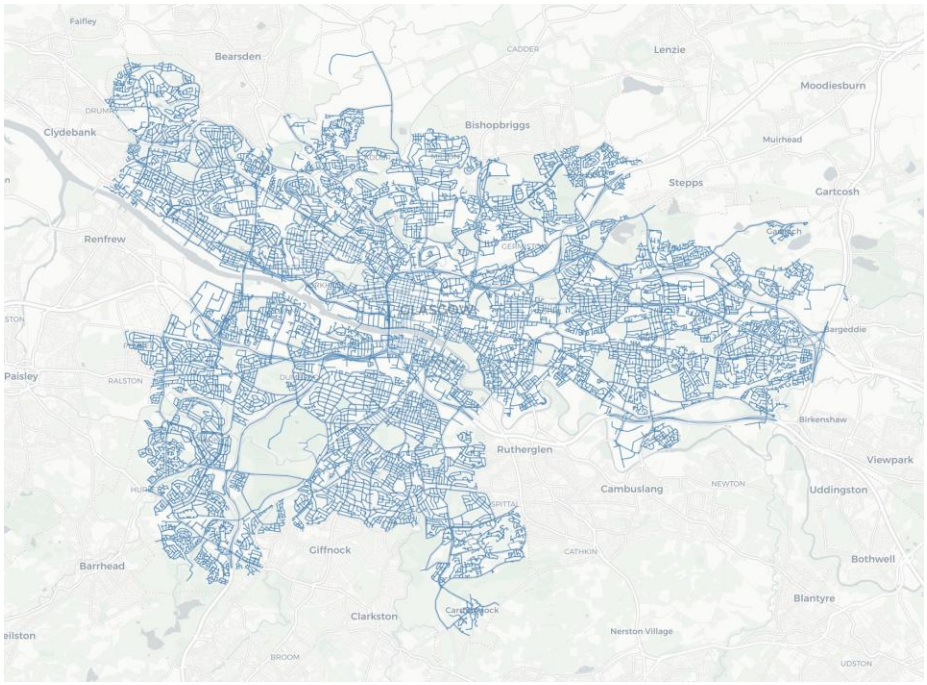



Figure 5.4: Full graph of Glasgow. Source: OpenStreetMap

### Obtaining OD matrices

The regions for the OD matrices are based on Louvain communities within the road network. The Louvain algorithm is a way of finding subregions within a graph based on internal connectivity. Its Python implementation is not entirely deterministic but different seeds yielded similar numbers of communities (about 100 communities) which suggests that whichever one we choose would be representative of the inherent structure within the road network. Figure 5.5 is a plot of the communities that we based our analysis on.

Take the convex hulls of the Louvain communities to create regions within the graph, see Figure 5.6. Some of them overlap and require manual cleaning.

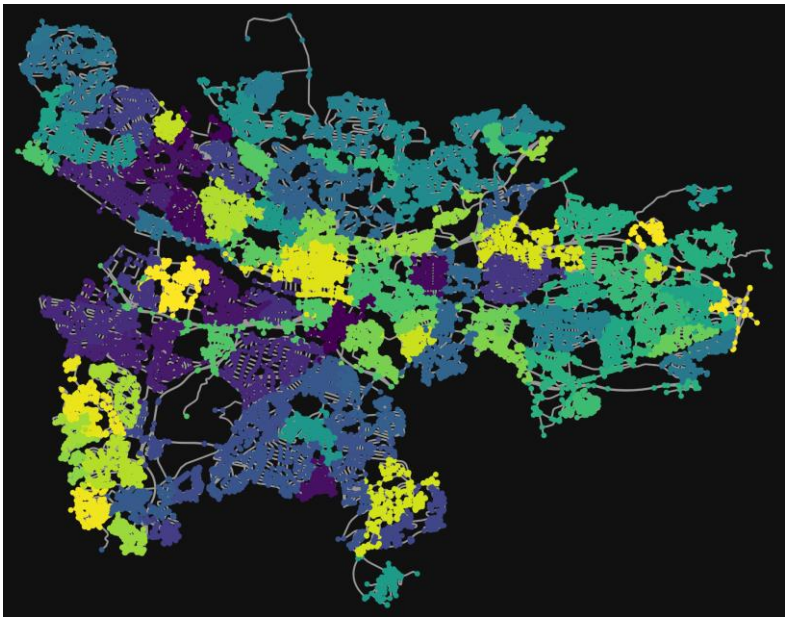


Figure 5.5: Louvain communities within the road network of Glasgow. There are 101 in total.

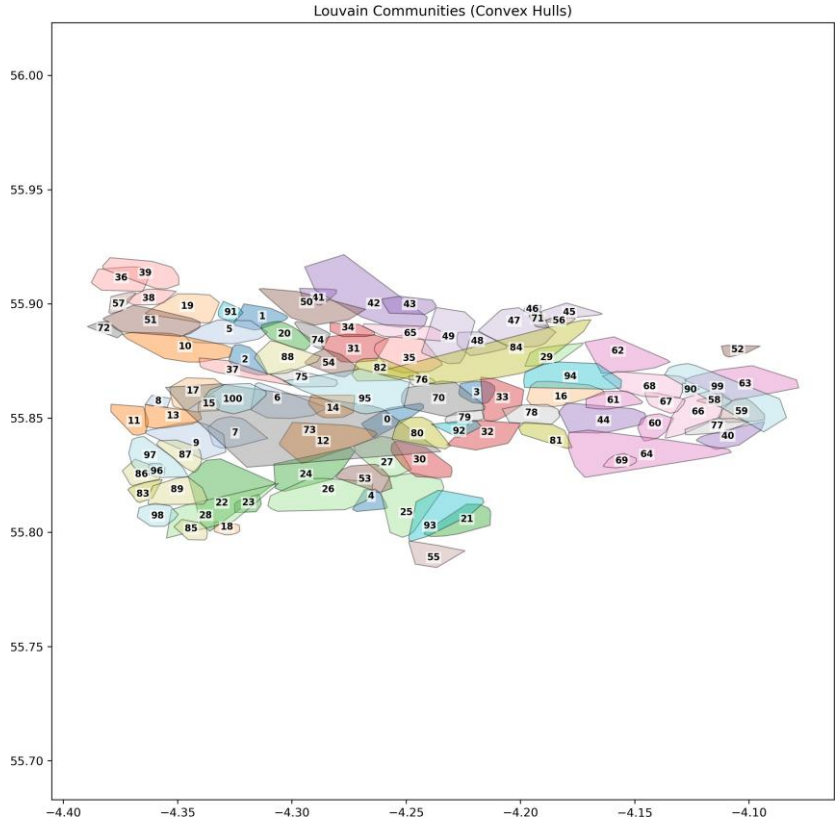



Figure 5.6: Convex hulls of the Louvain communities. Some of them overlap and require manual adjustment.

There are 94 non-overlapping regions left that we put into TomTom to obtain the OD matrices. This gave cumulative trip numbers from 7-10 am on 23 non-weekend days in July 2026. Assuming that TomTom captures about 10% of vehicles yields an average number of trips per hour between each pair of regions.

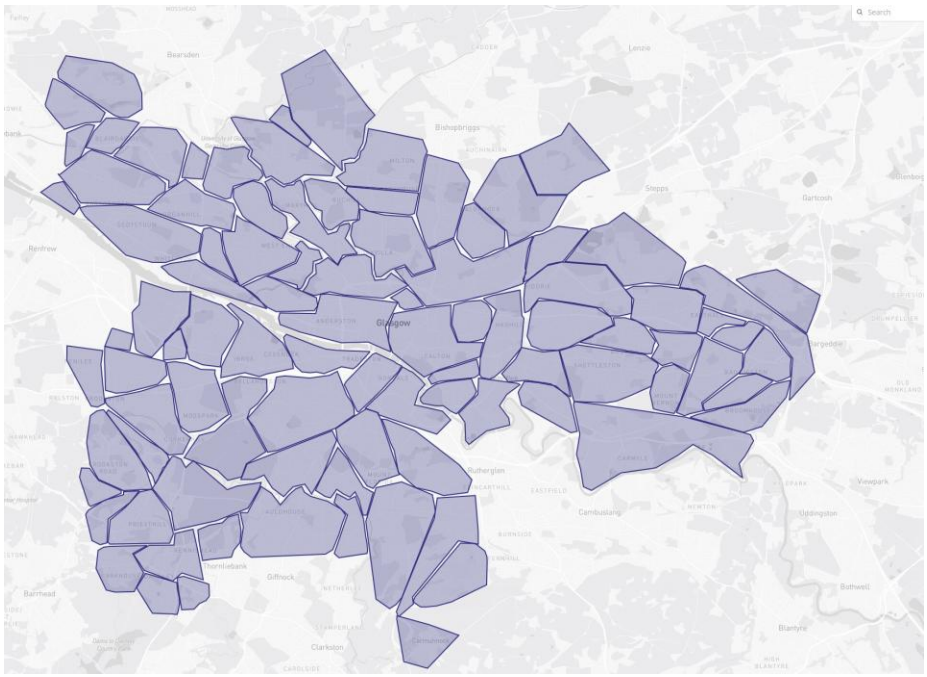


Figure 5.7: Louvain regions after manual cleanup

### Graph cleaning

Graph cleaning is in 2 steps: road removal and passthrough node simplification. We remove all roads labelled "Residential" by OSMnx assuming that they do not matter for traffic in between regions.

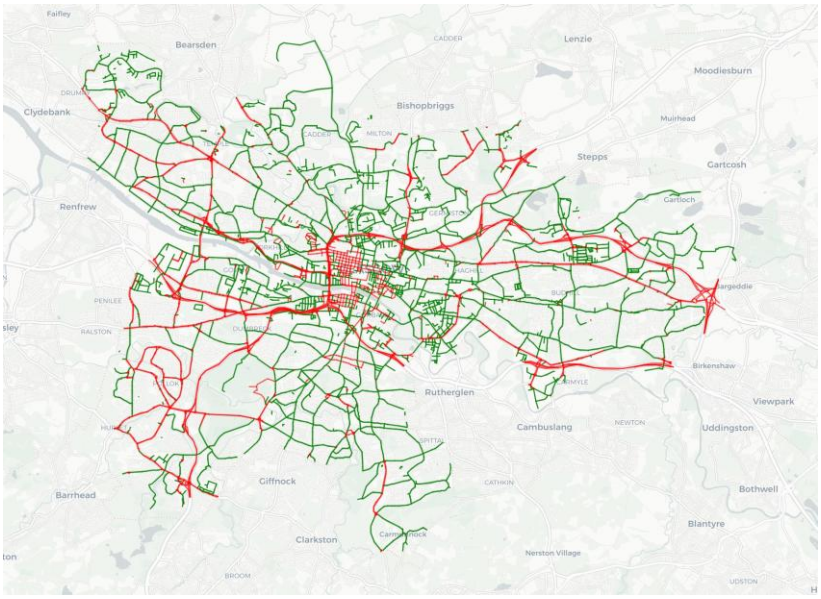



Figure 5.8: Glasgow with residential roads removed. Red edges are one-way, green are two-way.

The remaining graph contains nodes which split long roads into a sequence of short segments. Experience showed that this affects the traffic distribution algorithm. There are two types of nodes to remove (manually):

- Degree 2 nodes on one-way roads.
- Degree 4 nodes, splitting a bidirectional road.

Then take the largest strongly connected component: the largest component of the graph in which any two nodes are connected by directed paths. The simplification process results in the following graph:

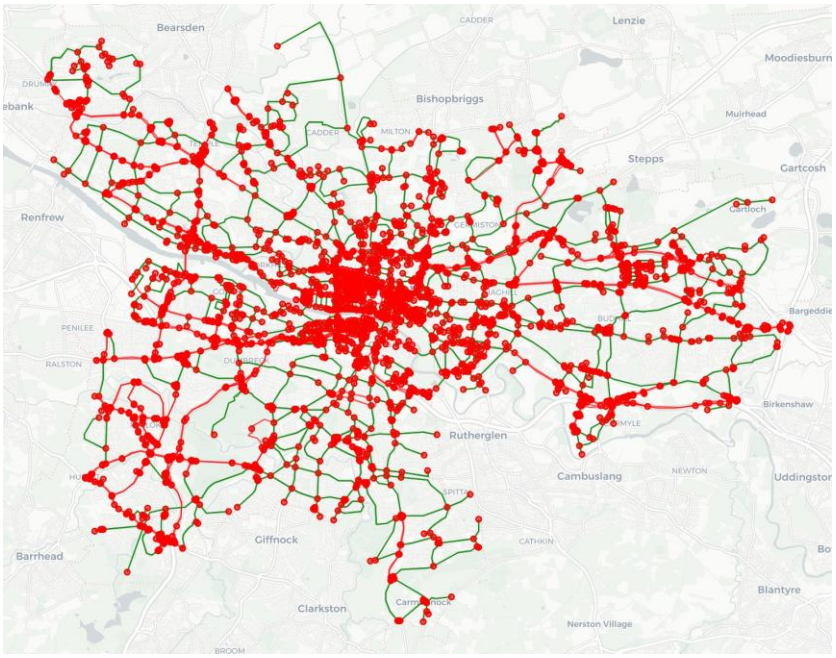



Figure 5.9: Fully cleaned up graph, with nodes showing

### Graph labelling

The graph is labelled with capacity and reference speed data. Capacity is assigned based on road type following [2] with 5% HGVs.

| Road type | Capacity (per lane per hour) |
|---|---|
| Motorway | 1900 |
| Primary | 1710 |
| Secondary | 1330 |
| Tertiary | 1330 |
| Unclassified | 1330 |
| Motorway link | 1900 |
| Primary link | 1710 |
| Secondary link | 1330 |
| Tertiary link | 1330 |

The type of road and the number of lanes were provided by OSMnx. Where the number of lanes is not present, we default to a single lane.

The reference speeds were taken from INRIX. Assign each edge in the graph its corresponding INRIX edge IDs, and where more than one INRIX ID was assigned to an edge, the reference speed is calculated as the floor of the mean of their reference speeds. Edges that did not have an INRIX ID assigned to them are assigned the mean of the reference speeds of their adjacent edges. If they do not neighbour any edges that have a reference speed they are removed from the graph.

### Adding in OD nodes

Next add the OD demand data to the network:

- Compute the centroid of each Louvain region (after cleanup) and add a node at the centroid location,

- Connect each centroid node to 2-3 neighbouring roads via edges whose travel times are set to zero. Try to connect to different roads in the region, and if a road is bidirectional connect to both directions.

### Assigning traffic

With the OD data loaded to the graph, run the Frank-Wolfe traffic assignment algorithm on the Glasgow network, obtaining flow values for every road in the network:

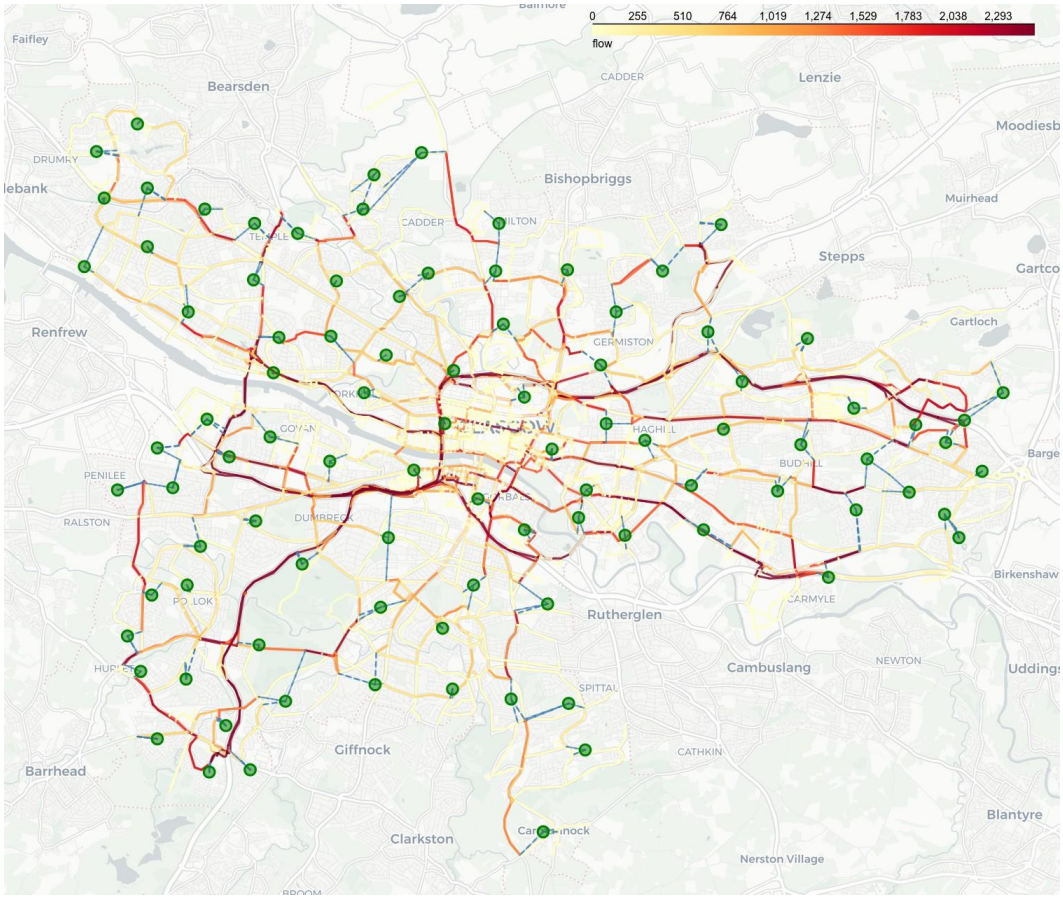



Figure 5.10: Glasgow graph after traffic assignment. OD nodes and connectors are also visible.

The above assignment consists of 134782 trips. We rank Frank-Wolfe for 95 iterations, converging to a solution within a tolerance of $5_{\times}\,10^{-4}$. The traffic assignment looks rather natural: motorways and the main express way have high traffic load assigned. Generally, arterial routes carry more flow than smaller roads contained within a neighbourhood.

### Criticality

Take the top 100 most loaded segments in the network and re-run the traffic assignment after removing them one at a time from the graph. Rank them by percentage increase in total travel time:

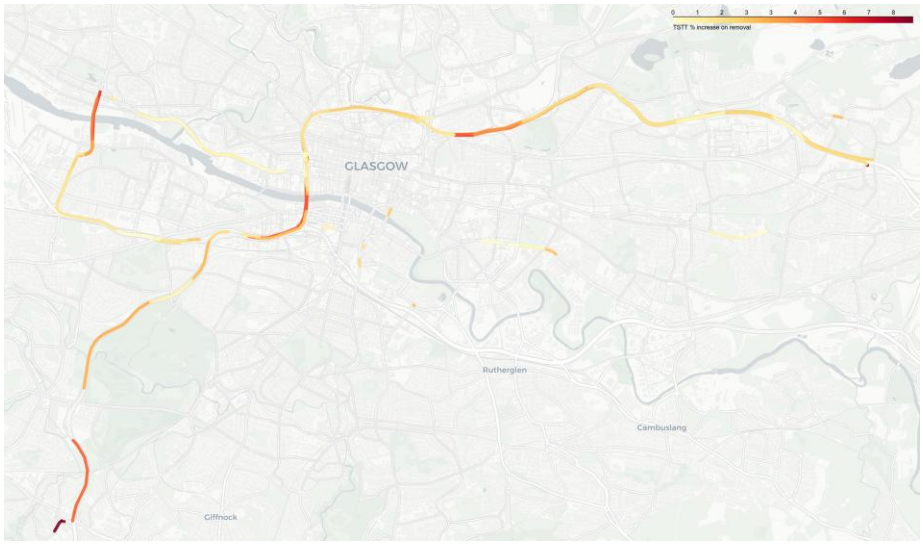



Figure 5.11: Top 100 segments, ranked by increase in travel time

The segments concentrate around the motorway, the expressway, Clyde Tunnel and Kingston bridge. A further analysis of travel time increase on more roads, chosen on the basis of a graph-theoretic index called betweenness yields:

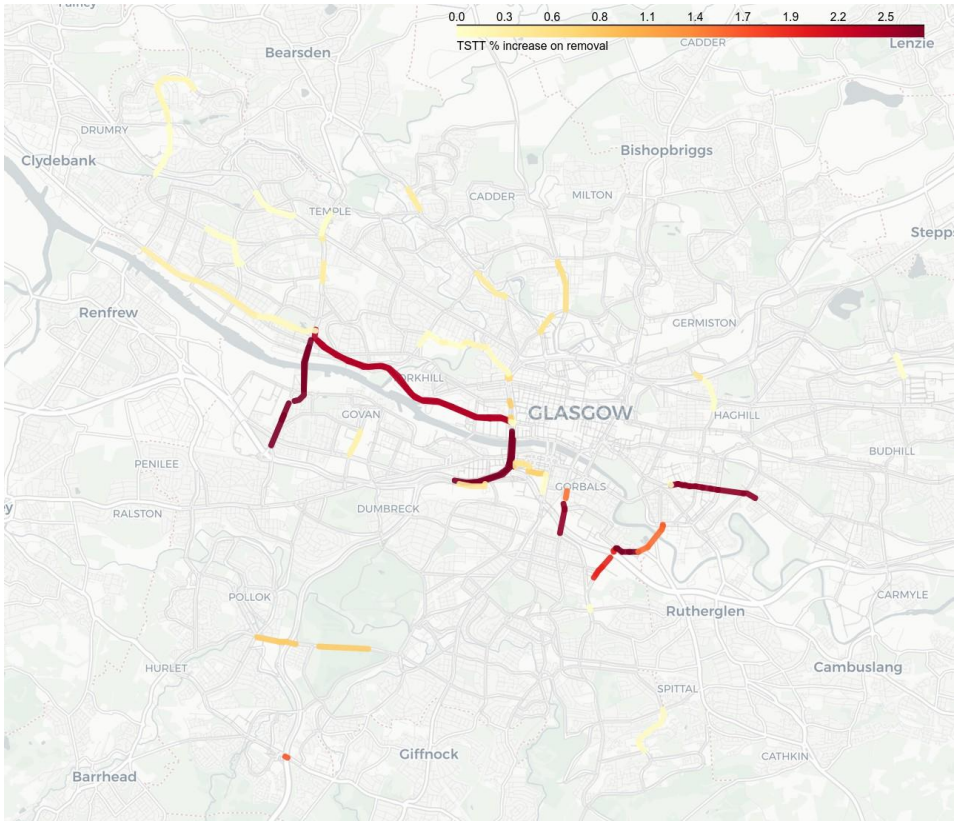



Figure 5.12: Criticality of various roads in Glasgow

Some examples of percent total system travel time (TSTT) increases (after total removal):

| Name | %TSTT increase |
|---|---|
| Kingston Bridge | 11.7 |
| Clyde Tunnel | 9.52 |
| Clydeside Expressway | 2.41 |
| London Road | 2.62 |
| Rutherglen Bridge | 1.52 |
| Woodlands Road | 0.06 |

The higher the TSTT increase the more critical a road is for traffic across the network.

# Chapter 6

# Conclusion and recommendations

## 6.1 Discussion of experiments

### 6.1.1 Traffic data analysis

Our experiments with statistical and machine learning methods aimed to establish a relationship between INRIX traffic data and roadworks. The results suggest that INRIX traffic data patterns have a mild predictive power for roadworks. This predictive relationship serves as a proxy for roadworks impact.

- Our attempts at correlating anomalous intervals in the INRIX traffic data with roadwork intervals were inconclusive. It seems that some roadworks have quite a subtle impact on traffic (or none at all), which is not easily picked up by these methods. They are very effective for detecting major incidents and structural differences in the data.
- The random forest models, by virtue of their interpretability, confirmed several intuitions: modifications in travel time seem to be the best predictor for roadworks, followed by modifications in traffic counts and speed data. The most relevant times of day for classification concentrate around the evening and afternoon rush hours, which is when we would expect roadworks to have the most impact on traffic.
- Out of all the models tested the transfomer-based model achieved the best ranking performance (in terms of assigning higher probabilities to actual roadwork periods). This indicates that the sequential nature of the traffic measurements contains meaningful information about roadwork impact, and should be taken into account.
- In the end, all models raised a large number of false positives which suggests that the patterns the models learned might not be exclusive to roadworks.

Based on these results, we do not think that historic INRIX data provides a viable way of understanding the impact of roadworks on traffic. Through our experiments we encountered several issues with the INRIX data. Firstly, it is very averaged and uniform. Secondly, a lot of data is discarded (due to the system assuming that stationary cars are parked). Roadworks that had a significant impact i.e. queues and delays would not appear in the data at all.

We have been made aware that INRIX will be changing their data collection mechanisms - for example, parked cars will be handled differently in the future. It would be interesting to see how this impacts the experiments we ran.

**Limitations**

The experiments have some limitations. First, we restricted ourselves to looking at data at the level of a single long road. It would be interesting to run the same experiments for a larger area. Second, we had to make choices and encountered some challenges in creating our training data. For example:

- Some of the models involve assigning roadworks to segments, based on geography. We considered this necessary since the INRIX data comes per-segment, but the segments themselves are arbitrary. We either assigned a roadwork to the nearest segment or to the two nearest segments - but it is not clear if this disregards possible upstream effects. Furthermore, it is not clear to what extent the impact of a roadwork correlates with its physical location. The radius of impact between different works is highly variable but we did not take that into account.
- When generating our training data, one repeated issue was having different roadwork-positive sample prevalence among the training, validation and test datasets. This is partially a consequence of the fact that INRIX data before August 2023 is much sparser, and we decided to not use it, affording us less data to work with.
- The impact of works varies widely. Amongst the many happening only very few will cause significant disruption. Having a binary roadworks variable does not take differing levels of impact into account.
- It is unrealistic to assume that a roadwork would be having an impact continuously the entire time that it is underway. The impact of a work will vary as it progresses. On the one hand differing traffic conditions throughout a day will be impacted in different ways, on the other hand works (and traffic management) change over time leading to different impact.
- All of our models rely on actual start and end date information for labelling our training samples. Discussions with the compliance officers revealed that these are not always accurate, and can be off by as much as a week (although accuracy has improved significantly in the last 2 years).
- Overall the information on the register does not provide enough detail to get an accurate picture of what was happening on the road at any given point in time.

Our experiments suggest that a roadwork happening by itself does not lead to traffic patterns visible in the data that are significantly different enough from traffic patterns without roadworks for a machine learning model to distinguish them. The conclusion is that a machine learning approach using the currently available traffic data is not the most fruitful way forward.

### 6.1.2 Network criticality analysis

Considering the limitations in the data and the simplifying assumptions made, the output for Glasgow looks promising. The model correctly identifies the main traffic links in town, and allows for traffic criticality computations. Repeating this calculation with better quality data will most likely yield even better results. It would be interesting to compare the outputs of the algorithm with the reinstatement categories of roads. Moreover it might be interesting to experiment with other measures of criticality in the network (see [1]).

The traffic criticality analysis could be developed in conversation with local authorities. It is crucial to check whether the results agree with local knowledge. This could be used to build a new scale of traffic criticality, combining model outputs with expert intuition.

This could be used for a new measure of disruption according to the (symbolic) formula:

$$\underset{\text{(i.e. in terms of TM type)}}{\text{Roadwork severity}} + \text{Road traffic criticality} \rightarrow \text{Disruption index}$$

The graph approach has potential for conflict checking too. An example of that would be checking for simultaneous works on North Bridge and South Bridge in Edinburgh. We would expect these to constitute a minimal cut or near-cut of the graph of arterial routes in the city. Our graph-based approach could be used to find roads or sets of roads which disconnect the city, but also near-cuts of the traffic-loaded graph, i.e. subsets of roads that when affected simultaneously increase the total travel time of the network significantly. The register could then check for conflicts along these subsets.

Furthermore, the model provides extra functionalities, such as computing detours in the case of a road closure, see B.

This method could easily be adapted to Edinburgh (and it would be interesting to see the outcome of better quality OD matrices, such as those from UBDC). For medium-sized towns it might be a good idea to also take into account the surrounding motorways and some neighbourhood geometry. In rural areas this could be repeated on a regional scale. These methods could potentially be used for an analysis of the whole road network of Scotland with an appropriate level of granularity (e.g. motorway network, taking cities as nodes in a graph).

**Limitations**

The efficacy of this method is determined by data quality and availability. This is because it requires a multitude of data sources: traffic data, OD matrix data, graph data, capacity data and free-flow speeds. We highlight some limitations in the data that was used:

- The number of lanes and road types were obtained from OpenStreetMap. We do not know how accurate the data is and had to use default metrics for lane numbers and road types for a significant proportion of edges due to missing data. The OpenStreetMap wiki highlights that volunteers often confuse whether lane count is per direction or not. Errors in lane counts (especially for large roads) impacts the capacity computations, which spills down through the entire pipeline.
- We do not know what fraction of traffic volume TomTom measures. We made an assumption that it covers about 10% of traffic and scaled the OD matrices accordingly (fine-tuning this scaling by hand to 7% by checking the resulting flow assignments - we tried to limit the number of roads whose assigned flows had more than triple the capacity). Moreover coverage might vary across different regions meaning that we cannot simply scale the whole matrix.
- The number of and the positioning of OD regions (as well as how they attach to the graph) is important: we made educated guesses, but choosing zones based on knowledge of neighbourhood structure could increase the accuracy of the model.
- Capacities were assigned based on standard values but in an urban environment many factors change the capacity of a road of certain type. Doing some work to figure out capacities in more detail would improve accuracy.

## 6.2 Practical suggestions for the Scottish Roadworks Register

From our discussions with local authorities we have found that there are many simple, practical measures that could be implemented on the Register to improve user experience. They touch on data visualisation and filtering. Operators find the visual aspect of the register very useful, and having the ability to filter which works and conflicts appear on screen would be appreciated.

- Adding functionality for restricting the types of roadworks shown on the map, e.g. only roadworks that are currently in progress, or planned during a period in the future.

- Adding a focused view of main roads, ignoring residential areas.
- Adding an option to filter by geographical location, either by specifying a custom radius, or more generally being able to draw a specific region on the map.
- The register currently shows conflicts between works that are planned, but have not started yet. This is especially problematic for works that have a very long advance notice period. It would be useful to add more granular time-based filtering. Users could filter for works scheduled within a certain time period.
- Adding an option to filter for works happening along the same road.
- Adding proximity to schools as an optional conflict check.
- Adding the possibility of disabling proximity flags between main road and side street. These kinds of conflicts happen frequently, but do not have a significant traffic impact.
- Adding a filter for traffic management when flagging conflict, e.g. removing conflicts with sidewalk-only works.
- When registering a road closure and a diversion route is proposed, the diversion route is plotted on the register. Including works/conflicts along the diversion route in the conflict report for the road closure.
- Conversely, during the period of road closure (for significant works) alternative routes could be marked as 'sensitive' (adjacent to 'restricted') in order to indicate increased traffic demand which might increase the impact of works along these routes.
- The ordering in which conflicts are displayed matters. Road restrictions/embargoes should be the first item the conflict report shows.
- Order conflicts shown by traffic criticality.
- Works that need to be amended/rescheduled could show as 'provisional' and not be included in the conflict check until dates/location has been finalised.
- An option to amend the traffic management of a set of works on the register as the works progress so that it accurately reflects the evolution of works. Currently changing traffic management on a notice is recorded as a noticing failure and requires an error correction agreement which might disincentivise people from changing it.

## 6.3 Data-related suggestions

Having access to quality data is essential for any kind of quantitative solutions. Below are some data-related suggestions that would be useful for future developments.

- While we have identified a number of traffic data sources, it is possible that many local councils collect their own traffic data, just like Glasgow City Council. It would be useful to have an overview of all traffic data collected across Scotland.
- We found it very informative to watch traffic coordination happening in real time. If the register could track how conflicts are currently handled thereby creating a data set of roadworks conflicts alongside with resolutions, this could potentially be used in optimizing conflict filtering (see 7.4).

- Traffic data providers such as INRIX have opaque data processing practices, which makes it hard to interpret what the data actually measures. It would be very useful to obtain a de-tailed, technical description of how the data is collected and processed. We tried to obtain this information from INRIX, but were ultimately not successful.
- OD matrices are vital for any graph/traffic-assignment based approach. While services like TomTom provide OD matrix data, it is unclear how to relate these to real-life conditions. Having access to accurate OD matrices would significantly improve the accuracy and relevance of traffic assignment models.
- We obtained bus delay data from Lothian buses for a single route over a 4 months. It would be very interesting to obtain bus delay data spanning several years and several routes for further analysis. This could provide a beneficial avenue of research for subsequent AGQ CDT cohorts.
- It would be useful if the Gazetteer contained more information about roads, such as speed limit, number of lanes (in each direction). When implementing our Glasgow graph experiment, we had to rely on numbers of unknown accuracy provided by OSMnx.

# Chapter 7

# Further work

There are many further avenues of exploration that we could not investigate due to time constraints. Here are some that we believe might be fruitful for future years.

## 7.1 Refining and building on the traffic sensitivity analysis

Rather than relying on the data from Open Street map it might be possible to determine the number of lanes and hence the capacity of a road from satellite images. This is process that could be automated using pre-trained image segmentation and recognition models.

This might be particularly effective if the traffic sensitivity analysis is to be extended to a larger part of Scotland.

Running the analysis with better OD data and experimenting with different measures of criticality.

Working out the capacities of roads in an urban environment, taking into account junctions, traffic lights, parking etc.

Running Frank-Wolfe with a more sophisticated volume-delay function and trying different measures of criticality.

Combining the traffic criticality analysis with roadwork type yields a possible impact score which might allow administrators to prioritise their workload.

## 7.2 Revisiting the traffic data

As INRIX has changed (as of July 2026) the filtering in order to discard less data, it might be worth checking whether data from 07/2026 could serve to establish a more meaningful predictive relationship between traffic behaviour and roadworks. Moreover it would be good to experiment with training on several roads simultaneously to see if that improves predictions or whether structures in traffic behaviour are highly local.

## 7.3 Simulating traffic in a small neighbourhood

There are many deterministic traffic models. This might be especially helpful in order to determine the cumulative effect of several roadworks in close proximity. Most of these models require traffic

counts as well as origin destination matrices. Simulating traffic in a local neighbourhood and adding in roadworks might provide insight into how traffic disturbances arise and dissolve.

## 7.4 Learning conflict resolution

If a data set of roadwork conflicts and their resolutions was available, this could be used in order to train a model to learn which combinations of roadworks in which locations are problematic and how they are resolved. The downside of this approach is that it would be entirely reflective of and determined by current practices.

## 7.5 Leveraging the bus delay data

We found the bus delays to be a valuable proxy for the real traffic conditions on the road. If more bus data was available (several years' worth) it might be possible to establish a predictive relationship between roadworks and bus delays which would be an other way of quantifying the impact of works.

## 7.6 Bayesian methods

Due to the uncertainty involved in the roadworks data it might be helpful to use Bayesian methods to model the distribution of roadworks across the network in time. From there it might be possible to predict locations of conflicts.

# Acknowledgements

We would like to express our sincere gratitude to the Scottish Roadworks Commissioner's office for their invaluable support throughout. This research would not have been possible without their generous explanations.

We thank Tudor Dimofte and Darrick Lee for many helpful discussions.

We thank the local authorities for giving us an understanding how the roadworks register is used in practice.

We thank Lothian buses for providing the bus delay dataset.

We acknowledge Dr Alan Logan, Assistant Professor, Centre for Sustainable Road Freight and the Centre for Logistics and Sustainability, Edinburgh Business School, Heriot-Watt University.

We acknowledge Daniel Mitchell and the TransiT team and specifically: Louise MacMaster Project Officer - TransiT, College of Science and Engineering, University of Glasgow, Jie Zhao, Project and Industry Engagement Manager (TransiT) Professor David Flynn, Head of Research Division, Autonomous Systems and Connectivity (ASC) James Watt School of Engineering College of Science and Engineering, University of Glasgow , Co-Principal Investigator Transit.

Finally, we would like to thank the AGQ CDT team for their support. This project has been very much a team effort.

# Appendices

# Appendix A

# A note on model evaluation

## A.1 Metrics

The point of this section is to detail the meaning of some of the metrics reported after training various models. Consider a generic binary classification problem: given input data, answer a yes/no question (in our case, "yes roadworks" or "no roadworks"). Say we have trained a model $M$ on a classification question. Given input data (such as traffic data in a day), the model outputs a numerical value $M$(input data). In order to obtain a prediction, we pick a threshold $\tau$ and define

- Model says "yes" if $M$(input data) > $\tau$
- Model says "no" if $M$(input data) < $\tau$

There are various ways of picking the threshold, based on classification goals.

**Precision and Recall** Let $M$ be a trained classifier, and pick a threshold $\tau$. We run the classifier on testing data, and see how the predictions match with what we know is true. From this, we compute the following two values:

$$\text{Precision} = \frac{\#\text{ of true positives}}{\#\text{ of predicted positives}}$$

$$\text{Recall} = \frac{\#\text{ of true positives}}{\#\text{ of actual positives}}$$

In other words, precision measures how accurate the the model is when saying yes (in our case, how often there actually was a roadwork when the model says there was a roadwork). Recall measures how good the model is at saying "yes" when something happened (in our case, how often the model said there was a roadwork when there actually was a roadwork happening).

**PR AUC** PR AUC (Precision-Recall Area Under Curve) does not involve picking a threshold; it is a measure of how well the model is able to separate the positive class from the negative class via its predictions.

For every possible threshold $\tau$ we can compute the precision $P_\tau$ and recall $R_\tau$ given by using $\tau$ for decision-making. As we vary $\tau$, the points $(R_\tau, P_\tau)$ draw a curve in the unit square, known as the Precision-Recall curve. The PR AUC score is given by

$$\text{PR AUC} = \text{Area under the Precision-Recall curve}$$

The higher the PR AUC score, the better the model is at separating the positive and the negative class. For example, a perfect classifier would rank every truly positive test higher than every truly negative test. Generally, the higher the PR AUC score is, the better the model is at remaining precise in its predictions while catching a good chunk of actual positives.

- A perfect classifier would have a PR AUC score of 1.
- If the proportion of positive test cases in all test cases is $a$, then a totally random classifier would have a PR AUC score of $a$. As a result, if the test data is unbalanced, PR AUC should always be considered with respect to positive case prevalence.

**ROC AUC** ROC AUC is a similar metric to PR AUC, but the curve is now traced out by plotting the false positive rate against recall. A random model always has a ROC AUC of 0.5 (independent of the class prevalence), but for imbalanced datasets such as ours, PR AUC is a better performance metric.

**$F_\beta$ scoring** For a fixed $\beta$, the $F_\beta$ score at a threshold $\tau$ is given by

$$F_\beta = \frac{1+\beta^2}{\frac{1}{\text{Precision}_\tau} + \frac{\beta^2}{\text{Recall}_\tau}}$$

The $F_\beta$ score can be used to decide on a threshold for making decisions, by picking the threshold that maximizes the $F_\beta$-score. The parameter $\beta$ controls the relative weighting of recall vs precision: if $\beta = 1$ they are given equal weight, if $\beta > 1$ having high recall is favoured, and if $\beta < 1$ having high precision is favoured. The threshold should always be computed on a validation dataset, and not on a test dataset.

**Confusion matrix** For a model with a fixed threshold, tested on a testing dataset, the confusion matrix is the matrix containing the values

| | Predicted no | Predicted yes |
|---|---|---|
| Actual no | # True negatives | # False positives |
| Actual yes | # False negatives | # True positives |

which gives a concrete assessment of how the model performed.

# Appendix B

# Importance of OD attachment nodes

We want to highlight the importance choosing the attachment points for the origin-destination zones. We encountered the following situation in our setup. There is an OD node corresponding to the Shieldhall industrial area, right below Clyde Tunnel, that was attached as follows:

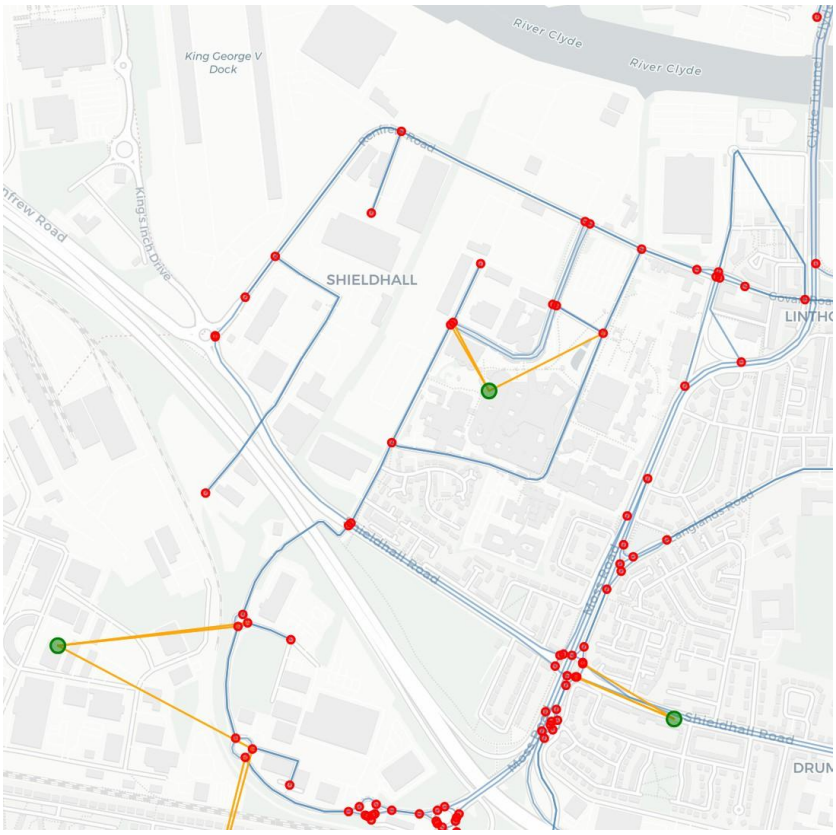



Figure B.1: Node placement around Shieldhall

This rendered the top 2 most critical segments in the entire network to be the following (blue was the most critical):

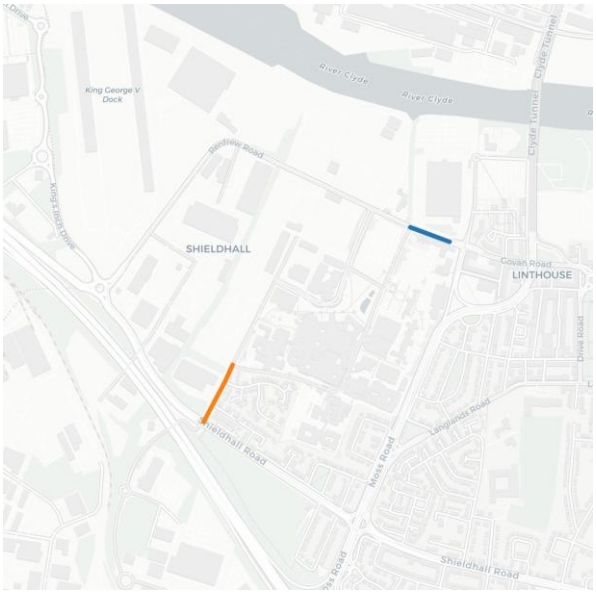



Figure B.2: Top 2 most critical segments in Glasgow

In particular, the travel time of users using these segments turned out to correspond to 40% of the total travel time across the entire network. This is due to the fact that these segments form a bottleneck around the Shieldhall node. We computed the shortest detour path around the blue segment:

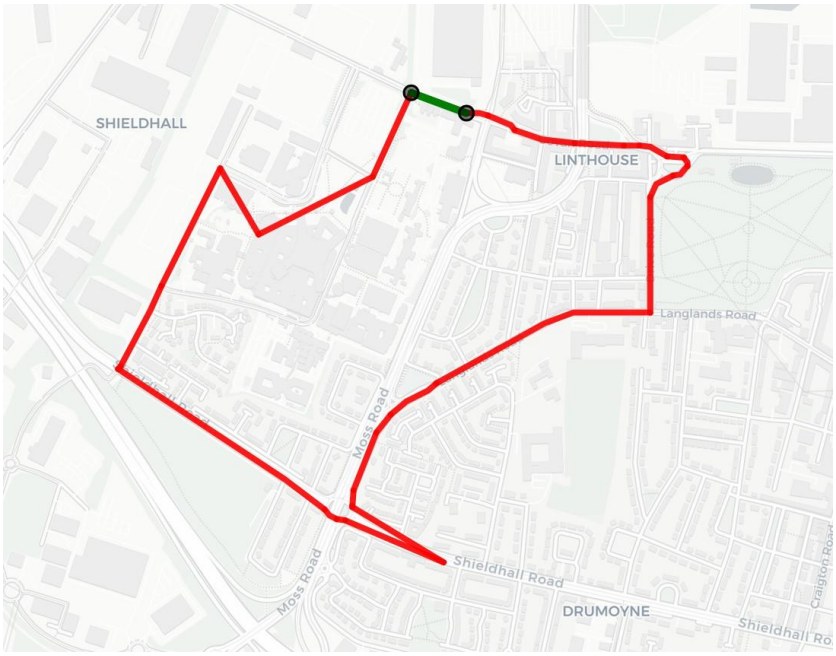



Figure B.3: Blue segment detour path

The detour path is around 3km long. These artefacts disappeared when we changed the attaching points of the Shieldhall node:

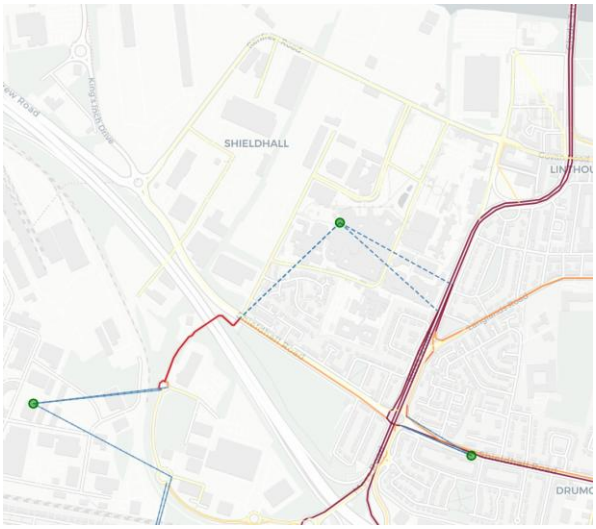



Figure B.4: Corrected Shieldhall node attachment

# Bibliography


[1] Chandra Balijepalli and Olivia Oppong. Measuring vulnerability of road network considering the extent of serviceability of critical road links in urban areas. *Journal of Transport Geography*, 39:145–155, 2014.

[2] Department for Transport et al. Traffic signs manual, chapter 8: Traffic safety measures and signs for road works and temporary situations. Technical report, Department for Transport, 2009.

[3] Feiyan Li, Hongfei Jia, Qingyu Luo, Yongxing Li, and Lili Yang. Identification of critical links in a large-scale road network considering the traffic flow betweenness index. *PloS one*, 15(4):e0227474, 2020.

[4] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on data mining*, pages 413–422. IEEE, 2008.

[5] Minghao Liu, Shengqi Ren, Siyuan Ma, Jiahui Jiao, Yizhou Chen, Zhiguang Wang, and Wei Song. Gated transformer networks for multivariate time series classification. *arXiv preprint arXiv:2103.14438*, 2021.

[6] Peter J Rousseeuw and Christophe Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical association*, 88(424):1273–1283, 1993.

[7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[8] John Glen Wardrop and James Ivor Whitehead. Correspondence. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(5):767–768, 1952.