Horsham Transport Study
Appendix B

Horsham Transport Model Data Report

On behalf of **Horsham District Council**

Project Ref: 45539| Rev: A | Date: November 2019

Registered Office: Buckingham Court Kingsmead Business Park, London Road, High Wycombe, Buckinghamshire, HP11 1JU
Office Address: Caversham Bridge House, Waterman Place, Reading, Berkshire RG1 8DN
T: +44 (0)118 950 0761 F: +44 (0)118 959 7498 E: reading@peterbrett.com

Document Control Sheet

Project Name: **Horsham Transport Study**

Project Ref: **45539**

Report Title: **Horsham Highway Model Data Report**

Doc Ref: **HTM/R1**

Date: **1st November 2019**

	Name	Position	Signature	Date
Prepared by:	Robert Dziurla	Senior Transport Planner	RD	01/11/2019
Reviewed by:	Paul Gebbett	Senior Associate	PG	01/11/2019
Approved by:	Kevin Lumsden	Director	KL	01/11/2019
For and on behalf of Stantec UK Limited				

Revision	Date	Description	Prepared	Reviewed	Approved

Contents

Figures

Tables

Appendices

This page is intentionally blank

1 Introduction

1.1 Overview

1.1.1 Peter Brett Associates LLP (PBA) was commissioned by Horsham District Council (HDC) to create a Strategic Transport Model of the area, the Horsham Transport Model (HTM). The model is to be developed to support HDC in production of a transport evidence base for the emerging Local Plan for the District up to 2036.

1.1.2 This Data Report will be the first in a series of reports which will be produced to inform the transport evidence base and in production of the model.

1.1.3 The remit of the HTM required that as much existing data as possible, this is consistent with good practice guidance contained in the Department for Transport's (DfT) internet-based transport appraisal guidance known as TAG[1]. This guidance recognises that data collection can consume significant resources in a modelling project. Therefore, wherever possible, new data collection must be kept to a minimum while making efforts not to compromise the robustness of the model and its potential to conform to TAG criteria for calibration and validation.

1.1.4 This report summarises the data that will be used to create the HTM and includes both existing data and newly collected data.

1.1.5 The types of existing and newly collected data comprise of the following and the method for collecting this data is discussed later in this report:

- Automatic Traffic Counts (ATC)
- Manual Classified Counts (MCC)
- Pedestrian and cycle counts
- Journey Time data
- Mobile Phone data for matrix building
- Traffic Signal Data

1.2 Study Area and Model Overview

Study Area

1.2.1 The detailed study area is shown in Figure 1-1 and will include the urban area of Horsham town and the entire district of Horsham. It will also include some roads outside of the Horsham District, where the Horsham District Local Plan strategic housing allocations may have an impact and would need to be reported. These links include parts of the SRN, as well as links within neighbouring district authorities within West Sussex. and within Surrey.

[1] https://www.gov.uk/guidance/transport-analysis-guidance-TAG

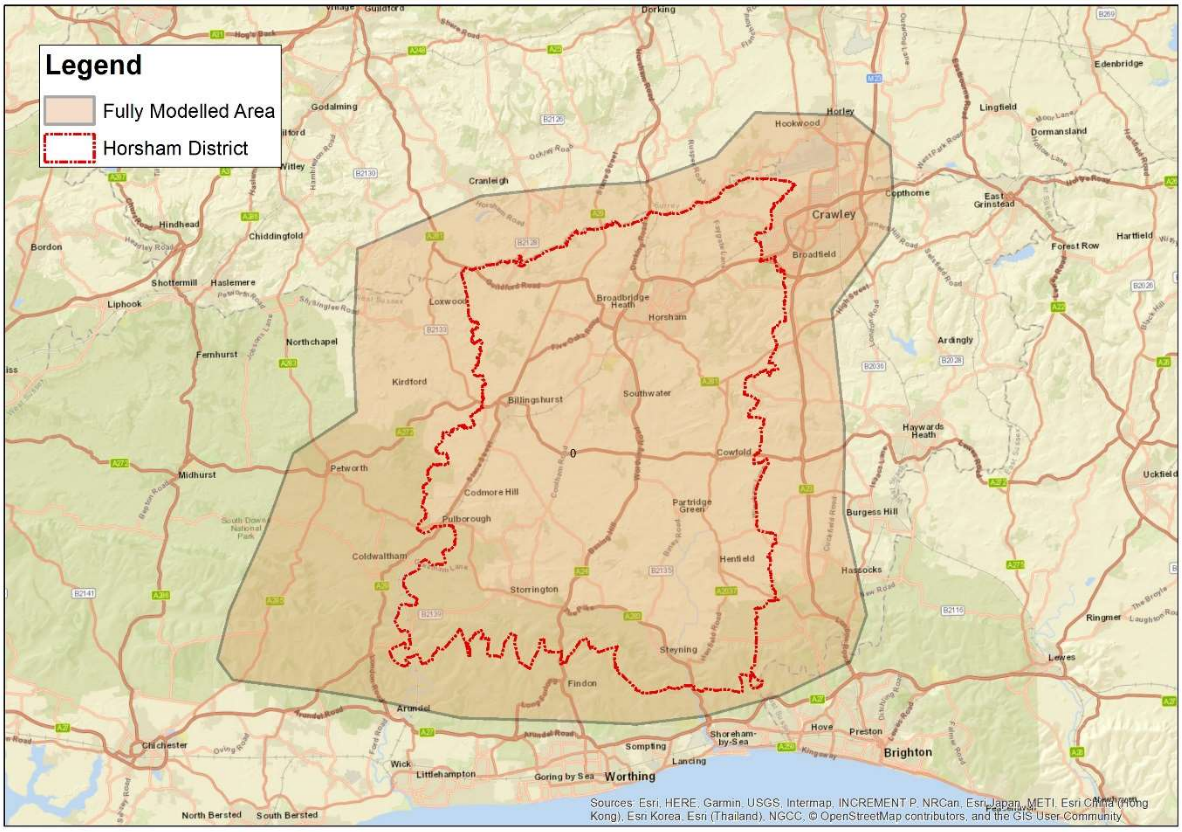


Figure 1-1: Horsham Transport Model – Study Area

Model Overview

1.2.2 The aim of the overall project is to develop a traffic model for the year 2019. This is known as the ‘Base Year model’ and will be developed with the aim to replicate the current conditions across the model area, thus meeting the validation criteria set out in TAG as best as possible. Validation is the process of checking the modelled outputs against observed traffic data. This process utilises traffic link and turning count data, along with journey time data. It should be noted that the aim is not to validate every link and turning count within the model area, but to provide a representation of the picture as best as possible, Further information on the development of the ‘Base Year’ model will be provided within a Local Model Validation Report, on completion of this stage.

1.2.3 Once a ‘Base Year’ model is produced, this then be used to create what is known as ‘Reference Case’ future forecast models. These models will be produced to represent possible future traffic conditions at the end point of the Local Plan period, which in this case is 2036. Further information on the development of the ‘Reference Case’ models will be provided within a Model Forecast Report on completion.

1.2.4 The ‘Reference Case’ models will then form the basis on which the assessment of the impact of proposed Local Plan development within Horsham District, as well as potential impacts within neighbouring authorities will be measured.

1.2.5 During the course of the development of the model, West Sussex County Council, the highway authority for the road network within Horsham District, will be consulted and will provide support in the development of the model.

1.2.6 Whilst none of the Strategic Road Network (SRN) is within Horsham District, Highways England, who manage the SRN, will also be consulted through the development of the model.

1.2.7 The HTM is a highway network model being developed using the established SATURN software. The model will be produced to represent an AM peak hour model (08:00 to 09:00), an average IP hour model (10:00 to 16:00) and a PM peak hour model (17:00 to 18:00). The model will consist of three vehicle types, comprising car, Light Goods Vehicles (LGV) and Heavy Goods Vehicles. The car element will then be broken down further to represent three trip purposes, namely commute, employer's business and other.

1.2.8 Following this introduction, this report is presented with the following structure:

- Section 2 Outlines ATC Surveys
- Section 3 Outlines MCC Survey data
- Section 4 Outlines Traffic Master Journey Time Data
- Section 5 Outlines the mobile phone data
- Section 6 Outlines Signal Junction Data

2 Automatic Traffic Count (ATC) Surveys

2.1 Overview

2.1.1 Automatic traffic count (ATC) data collection is undertaken using pneumatic rubber tubes which are placed across a road. These count vehicles as they pass over the tube. The data is stored electronically and can differentiate vehicle types based on vehicle length. The ATC data has come from three separate sources;

- Newly collected data – collected by specialist data collection company, Streetwise who were commissioned by PBA
- Existing WSCC Data
- Highways England Data

2.2 New ATC Data

2.2.1 New ATC surveys were undertaken by Streetwise over a two-week period (14 days) from Friday 10th May 2019 to Thursday 23rd May 2019.

2.2.2 The data survey company was able to discern and classify the ATC data into cars, LGV, OGV1 and OGV2.

2.2.3 The data is available in excel format, broken down into 1-hour intervals. This data is provided in a digital format (DVD) accompanying this report.

2.2.4 The locations of the new ATC surveys data are set out in Table 2-1 and shown in Figure 2-1. The data was collected at nine locations.

Table 2-1: New ATC locations

ATC Reference	Location Description	OSGR
1	Haven Road (Bucks Green)	TQ 08492 31988
2	Rowhook Road (Rowhook)	TQ 12260 34327
3	Muggeridge's Hill (Rusper)	TQ 18119 37815
4	Rusper Road (Rusper)	TQ 20329 39458
5	A29 Stane Street near St Andrews Farm	TQ 07888 24180
6	B2139 Coolham Road (South of A272)	TQ 11897 22482
7	B2133 Lordings Road (Billingshurst)	TQ 07128 24607
8	Marringdean Road (Billingshurst)	TQ 09089 23506
9	West Chiltington Lane (Storrington)	TQ 09984 23088
10	Pound Lane (Green Street)	TQ 14975 22557
11	Littleworth Lane (Cowfold)	TQ 19159 22122
12	A283 Shoreham Road (South of Roundabout with A2037)	TQ 19774 09375

ATC Reference	Location Description	OSGR
13	Annington Road (Botolphs)	TQ 18612 09568
14	Bostal Road (Steyning)	TQ 16443 09929
15	A281 Guildford Road near Weyhurst Farm	TQ 07451 33268
16	A272 Newbridge Road near River Arun	TQ 06919 25942
17	A29 London Road near Ingrams Farm	TQ 04083 17715
18	Loxwood Road (Bucks Green)	TQ 07426 32655
19	Brook Lane (Coldwaltham)	TQ 03251 16205
20	B2110 Handcross Road near Howards Nursery	TQ 22531 27655
21	B2116 Wheatsheaf Road near Gainsborough	TQ 21977 17539
22	Lambs Green Road (Lambs Green)	TQ 21697 36484
23	Tower Road (Faygate)	TQ 22087 33604
24	Forest Road (Colgate)	TQ 22156 32677
25	Grouse Road (Colgate)	TQ 22377 29712
26	Hammerpond Road (Horsham)	TQ 22449 28754
27	Horn Lane (Henfield)	TQ 21931 13870
28	Edburton Road (Henfield)	TQ 21294 11425
29	Blackbridge Lane N (Horsham)	TQ 16434 30152
30	Blackbridge Lane E (Horsham)	TQ 16395 30569
31	Wimblehurst Road (Horsham)	TQ 17672 31903
32	Foundary Lane (Horsham)	TQ 18205 31658
33	Depot Road (Horsham)	TQ 18789 30790
34	St Leonard's Road (Horsham)	TQ 18413 29984
35	Bashurst Hill (Bashurst Hill)	TQ 12227 28650
36	Fulfords Hill (Itchingfield)	TQ 13427 29851
37	Golding Lane (Mannings Heath)	TQ 20904 29279

2.2.5 The primary purpose of the ATC data is to provide independent data with which to inform the validation of the model as shown in Figure 2-1.

2.2.6 ATC data also provide evidence of day to day weekday flows to allow checks to be made, such that the model is representative of a weekday in a neutral month. The data will complement existing data from West Sussex County Council and Highways England ATC data collected as part of their permanent and periodic monitoring sites.

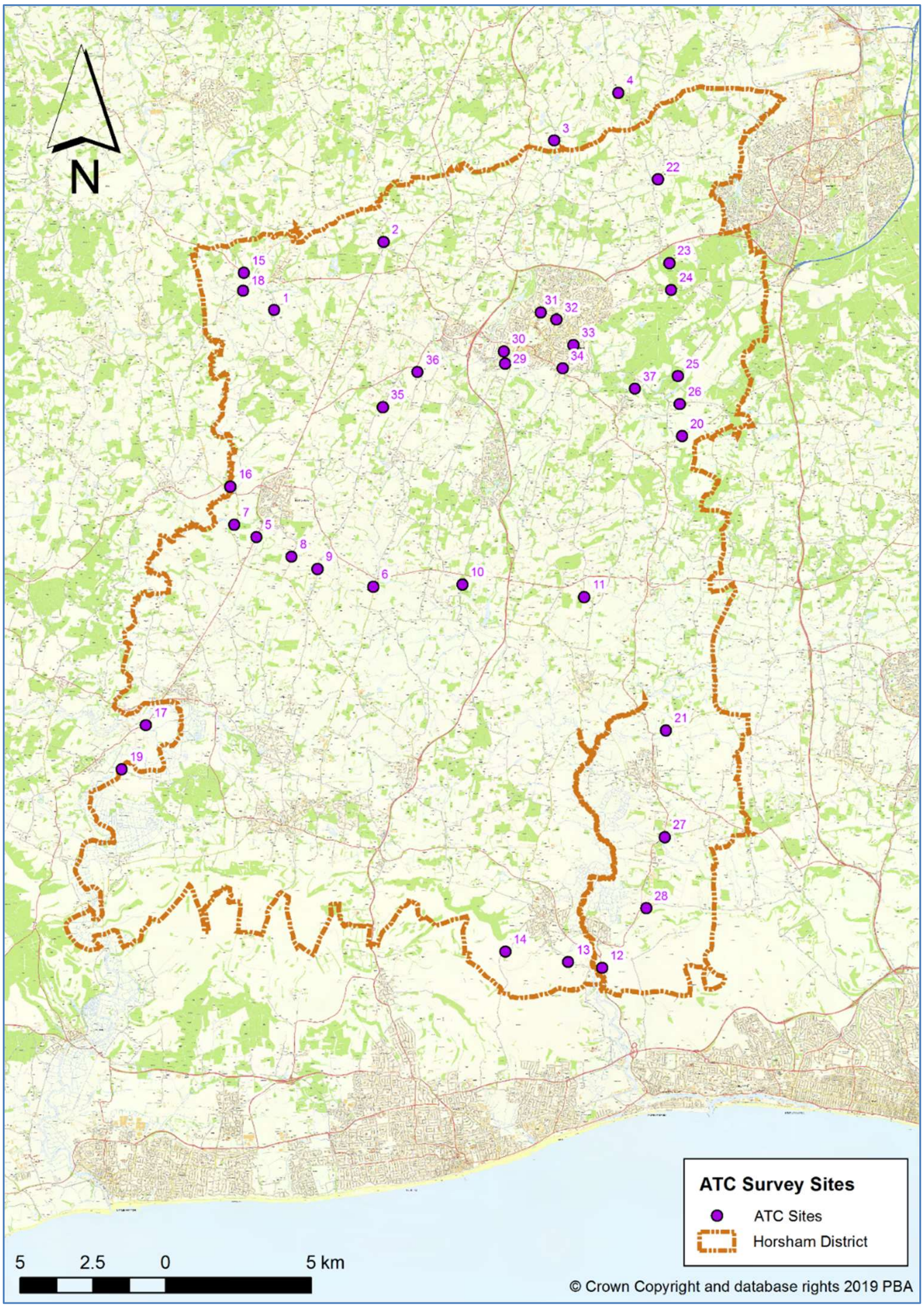


Figure 2-1: New ATC Survey data Locations

2.3 Issues

2.3.1 There have been several report issues with certain ATC sites, these issues have been identified as part of the quality checks during the ATC collection.

- Site 2: Tubes suffered damage on Friday May 17th. –fixed the day after, sufficient data available to acquire a statistically viable weekday average.
- Site 6: 1 week of data unusable due to ATC tube issues, sufficient data available to acquire a statistically viable weekday average.
- Site 15: Guildford Road – data is missing for extended period. To compensate for this, extra data was collected up until the 28th June 2019.
- Site 22: Tubes suffered damage from 16th-18th May. – Tubes fixed, enough data to provide weekday average.
- Site 33: Affected by vehicles parked on the tubes of the counter on the 14th and 15th May. Sufficient data available to acquire a statistically viable weekday average.
- Site 35: Issue with the tubes mid-way through week 2. Sufficient data available to acquire a statistically viable weekday average.
- Site 36: Tubes suffered damage from 15th-18th May. Sufficient data available to acquire a statistically viable weekday average.

2.4 ATC Data Analysis

2.4.1 Analysis of the ATC data has been undertaken to provide information on trends and peak flows. Furthermore, check have been done in order to remove any spurious data

2.4.2 Figures 2-2 to 2-38 provide 24-hour flow profile plots for each site by direction. The plots show how hourly vehicles flows vary across the day based on an average neutral day Tuesday to Thursday (i.e. all data from these days are combined and averaged).

2.4.3 In most cases it can be seen that the flows are tidal. The profiles also show that in most cases the flows peak at 08:00 to 09:00 in the AM peak period and at 17:00 to 18:00 in the PM peak period. This confirms the appropriateness of modelling these hours as the peak hours and modelling of an average IP hour.

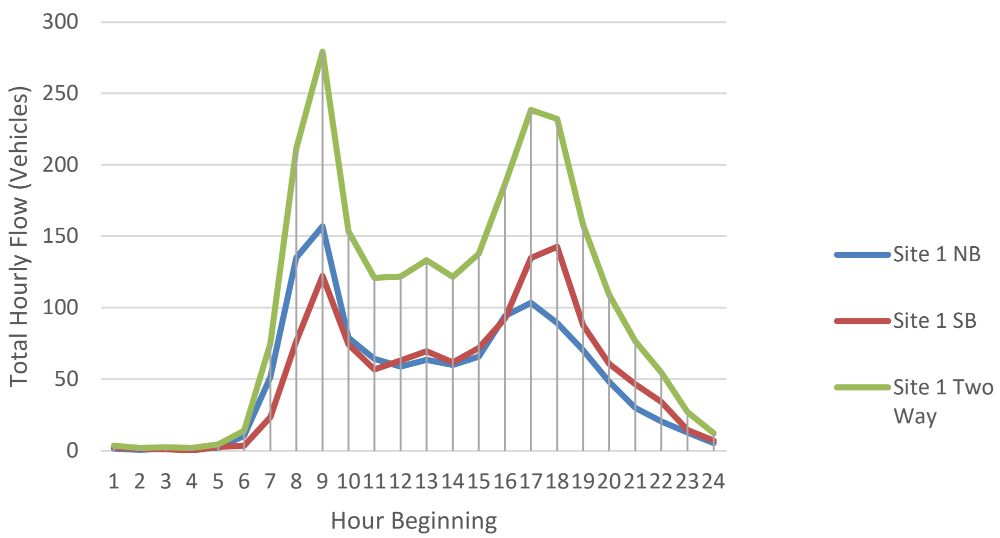


Figure 2-2 : ATC Site 1 Average Tues-Thurs Flow Profile

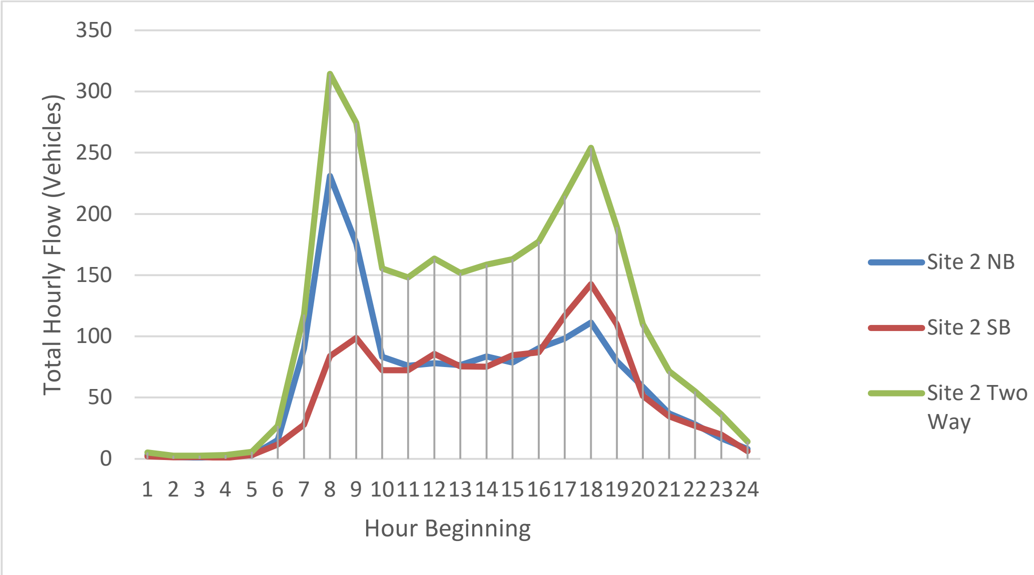


Figure 2-3: ATC Site 2 Average Tues-Thurs Flow Profile

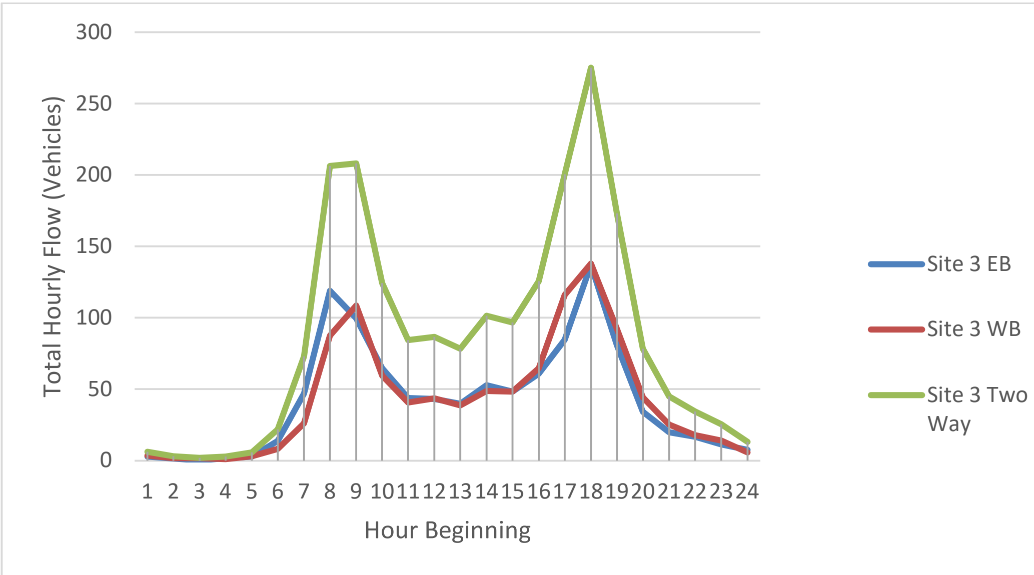


Figure 2-4: ATC Site 3 Average Tues-Thurs Flow Profile

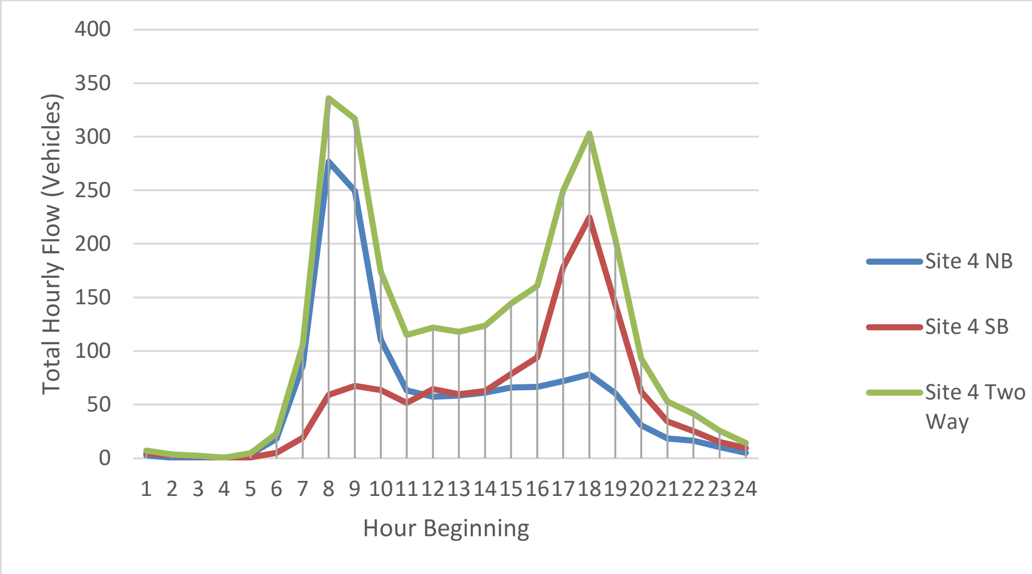


Figure 2-5: ATC Site 4 Average Tues-Thurs Flow Profile

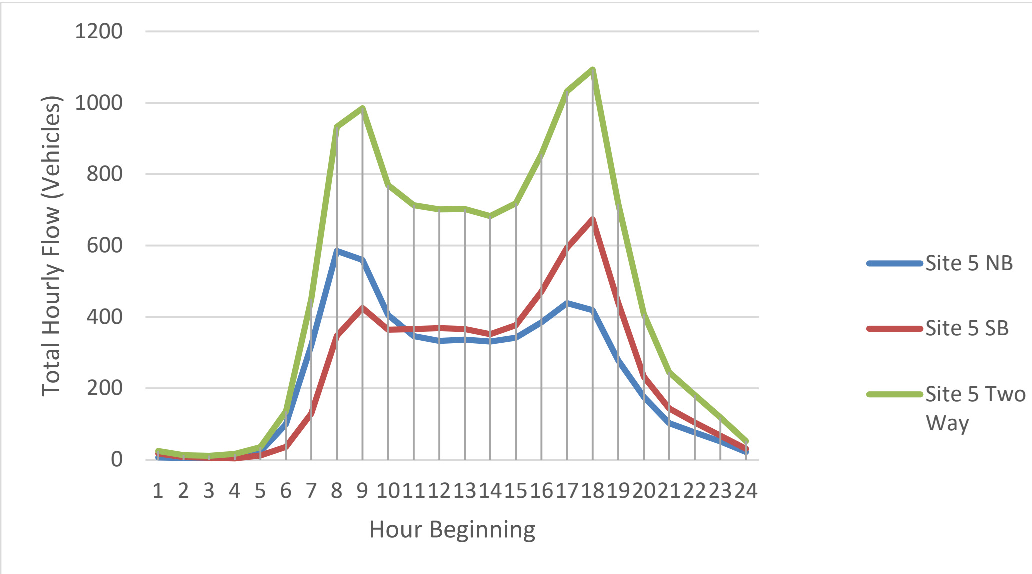


Figure 2-6: ATC Site 5 Average Tues-Thurs Flow Profile

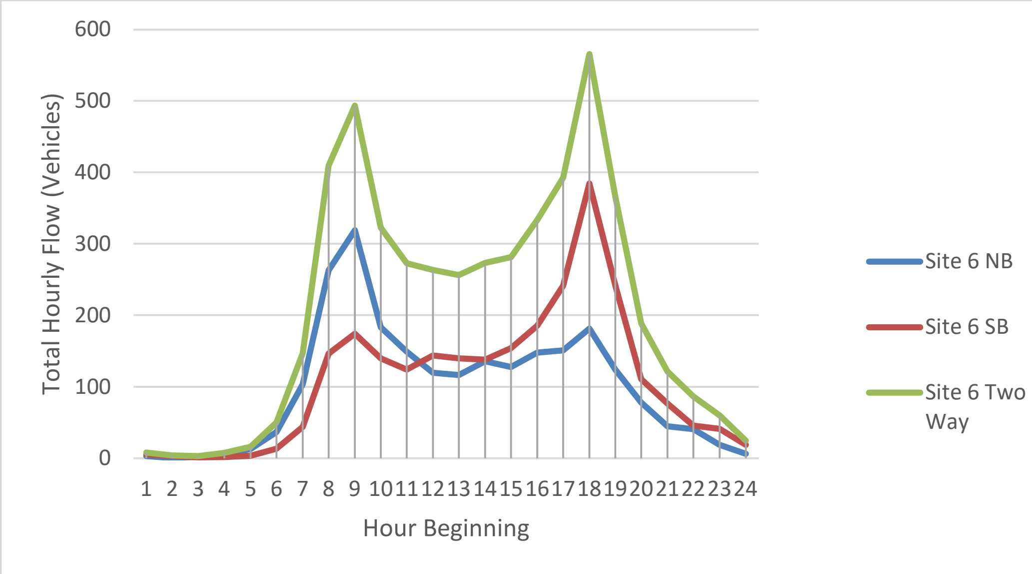


Figure 2-7: ATC Site 6 Average Tues-Thurs Flow Profile

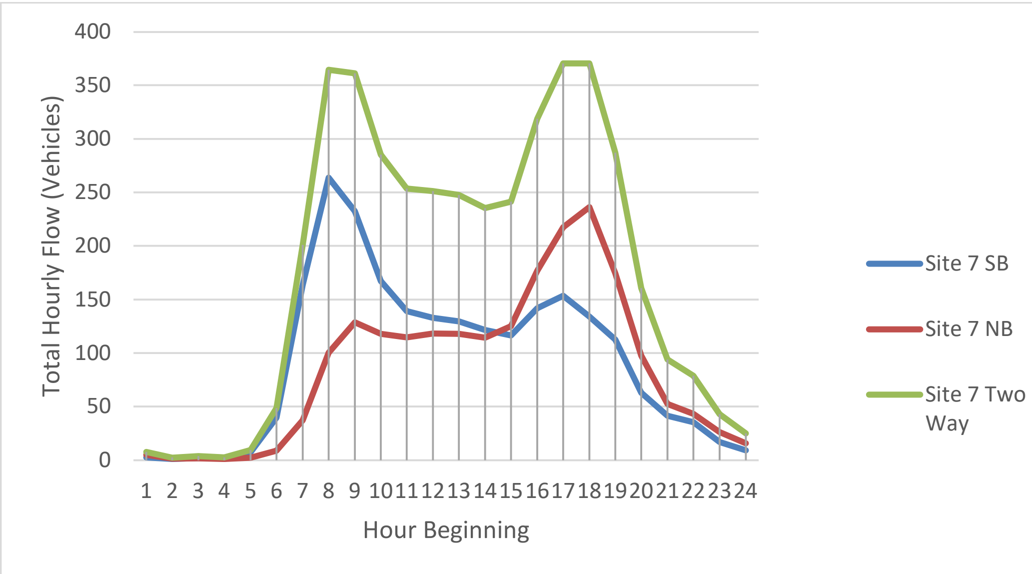


Figure 2-8: ATC Site 7 Average Tues-Thurs Flow Profile

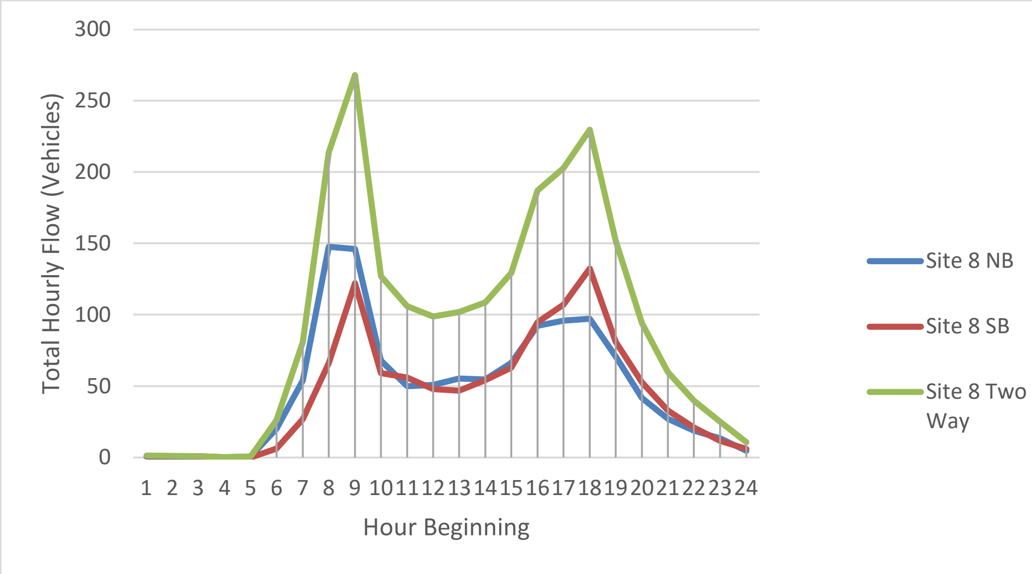


Figure 2-9: ATC Site 8 Average Tues-Thurs Flow Profile

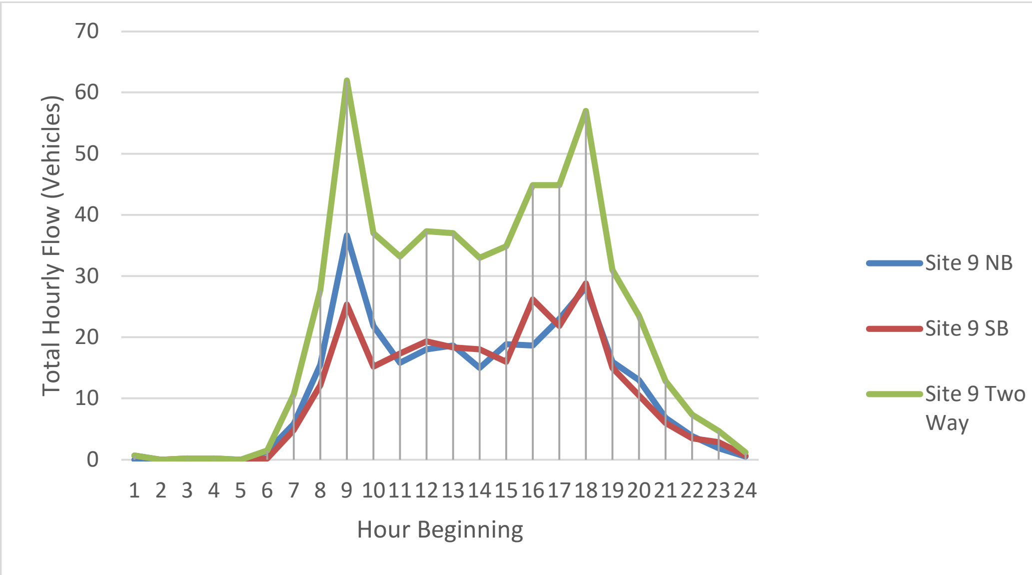


Figure 2-10: ATC Site 9 Average Tues-Thurs Flow Profile

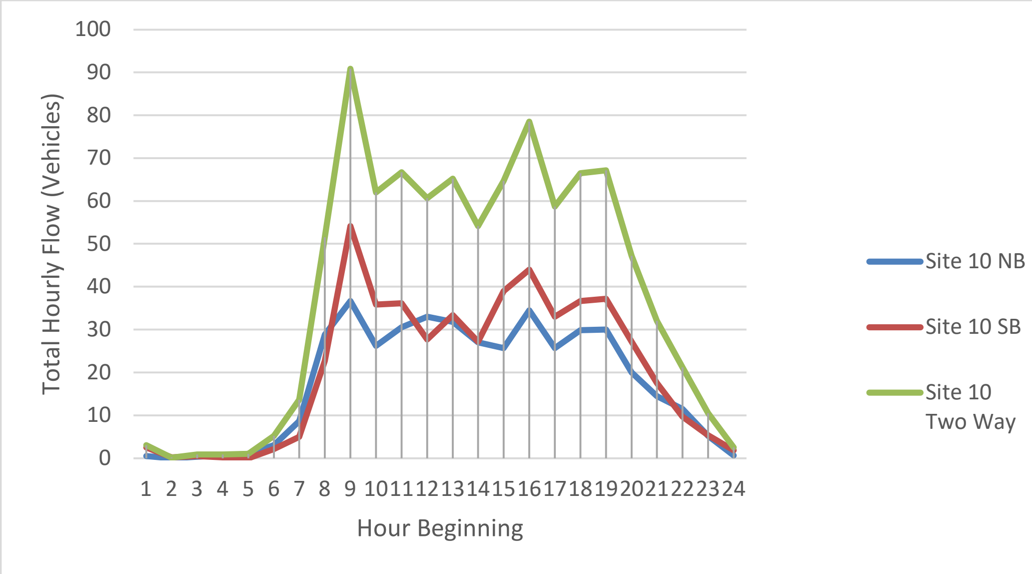


Figure 2-11: ATC Site 10 Average Tues-Thurs Flow Profile

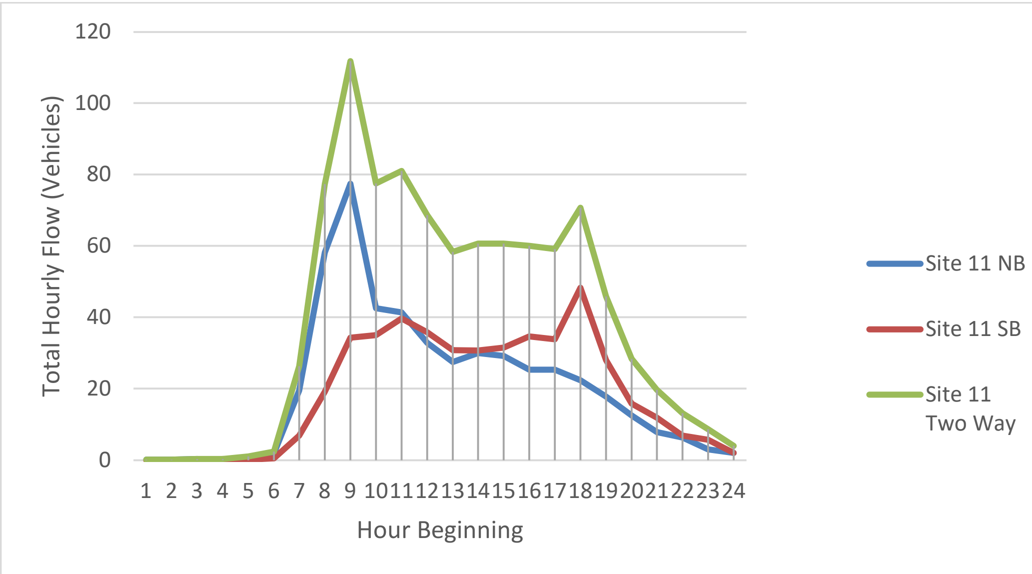


Figure 2-12: ATC Site 11 Average Tues-Thurs Flow Profile

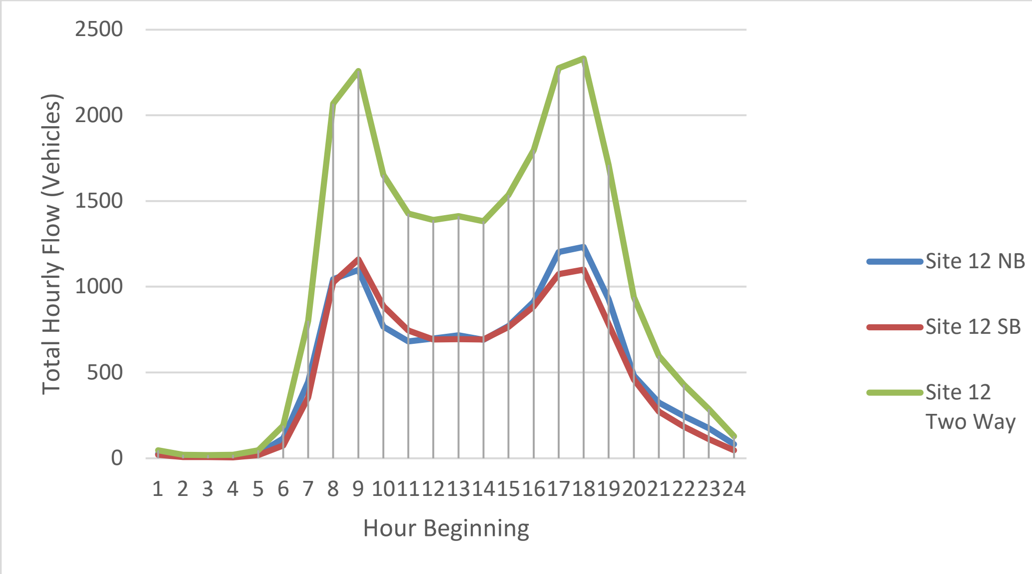


Figure 2-13: ATC Site 12 Average Tues-Thurs Flow Profile

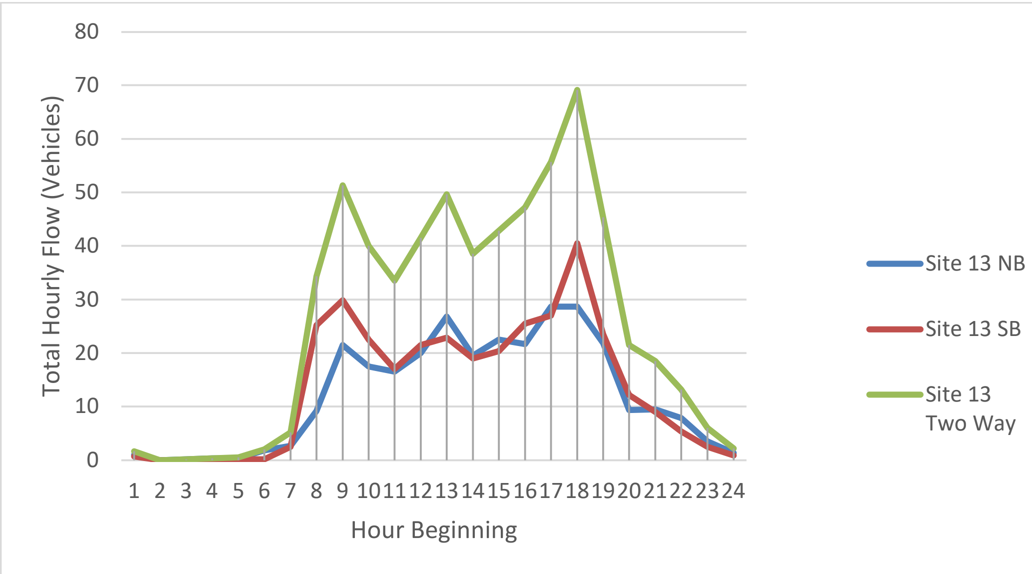


Figure 2-14: ATC Site 13 Average Tues-Thurs Flow Profile

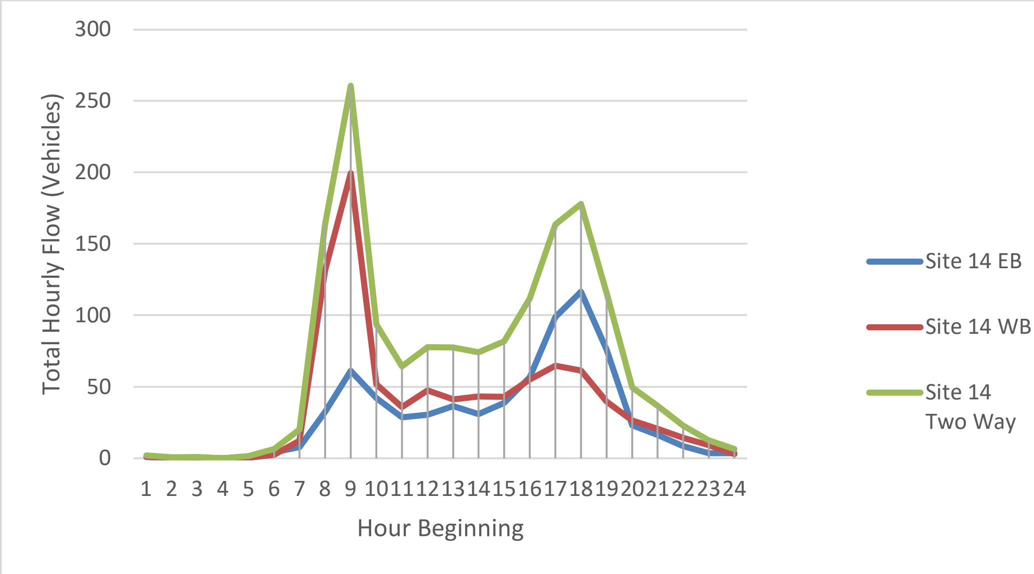


Figure 2-15: ATC Site 14 Average Tues-Thurs Flow Profile

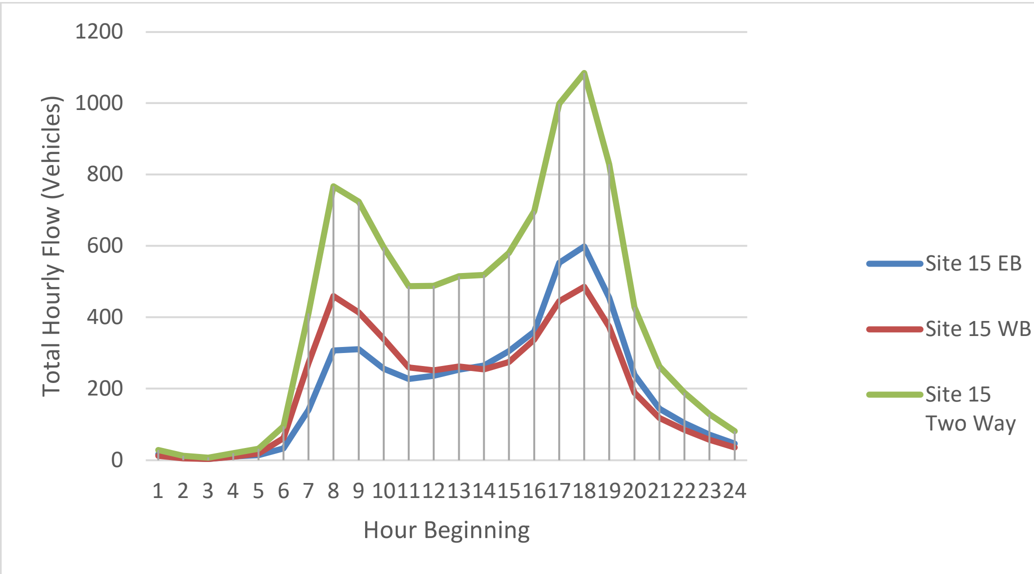


Figure 2-16: ATC Site 15 Average Tues-Thurs Flow Profile

Figure 2-17: ATC Site 16 Average Tues-Thurs Flow Profile

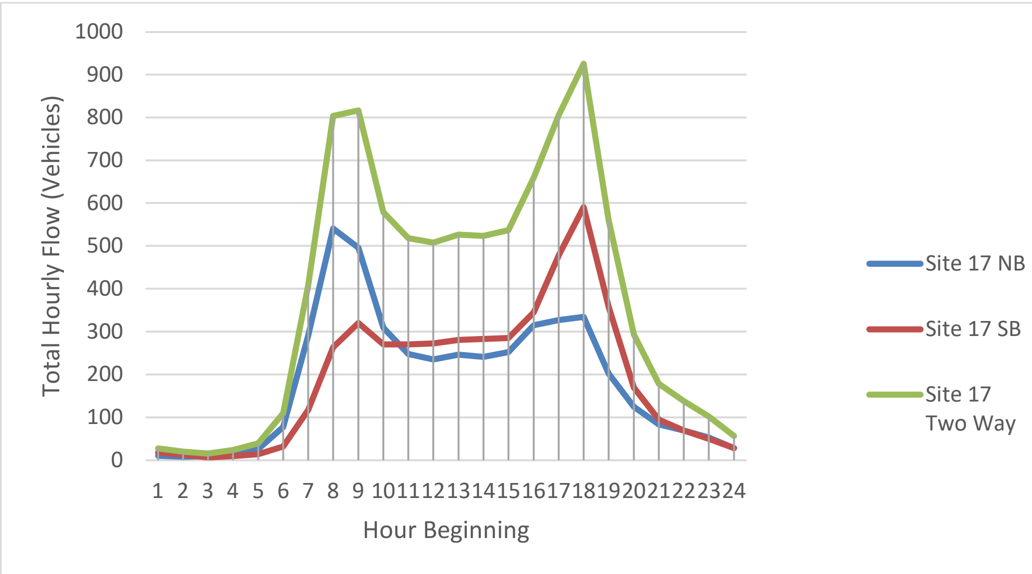


Figure 2-18: ATC Site 17 Average Tues-Thurs Flow Profile

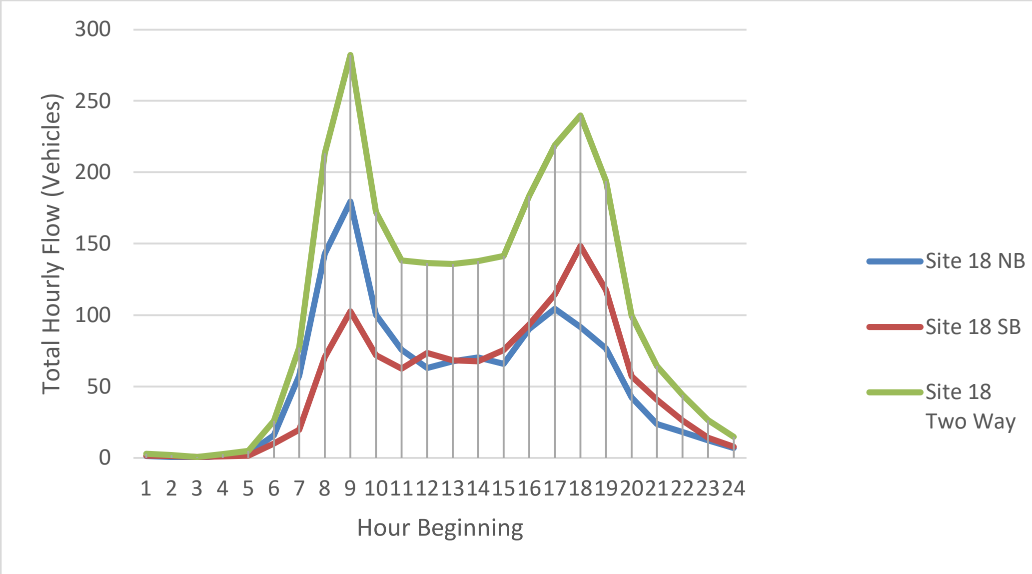


Figure 2-19: ATC Site 18 Average Tues-Thurs Flow Profile

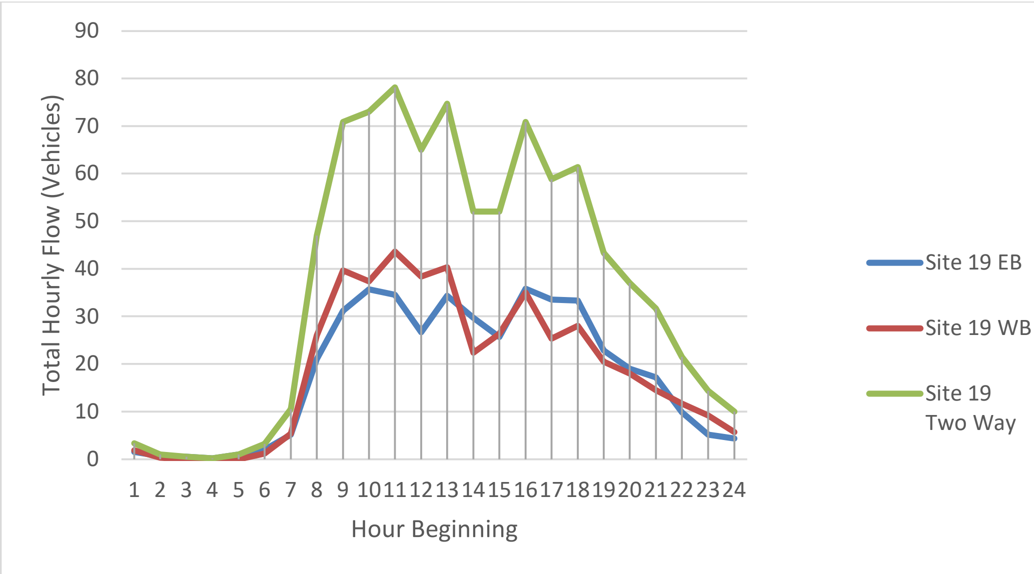


Figure 2-20: ATC Site 19 Average Tues-Thurs Flow Profile

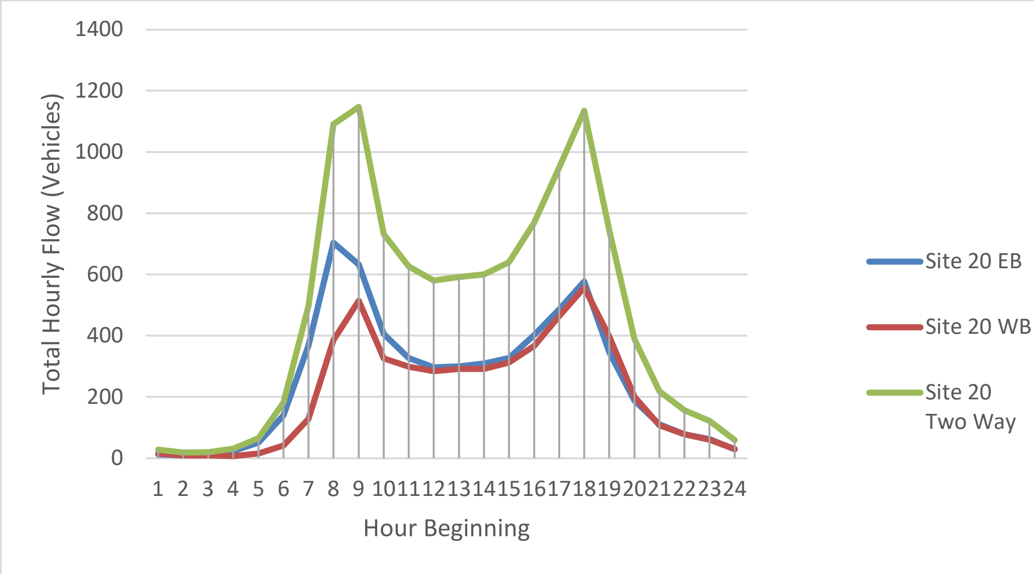


Figure 2-21: ATC Site 20 Average Tues-Thurs Flow Profile

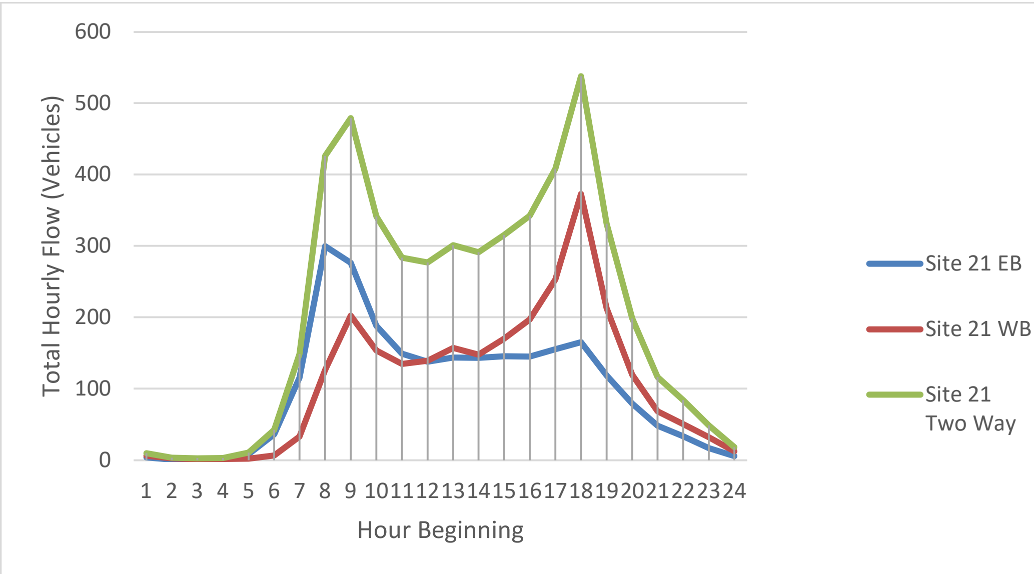


Figure 2-22: ATC Site 21 Average Tues-Thurs Flow Profile

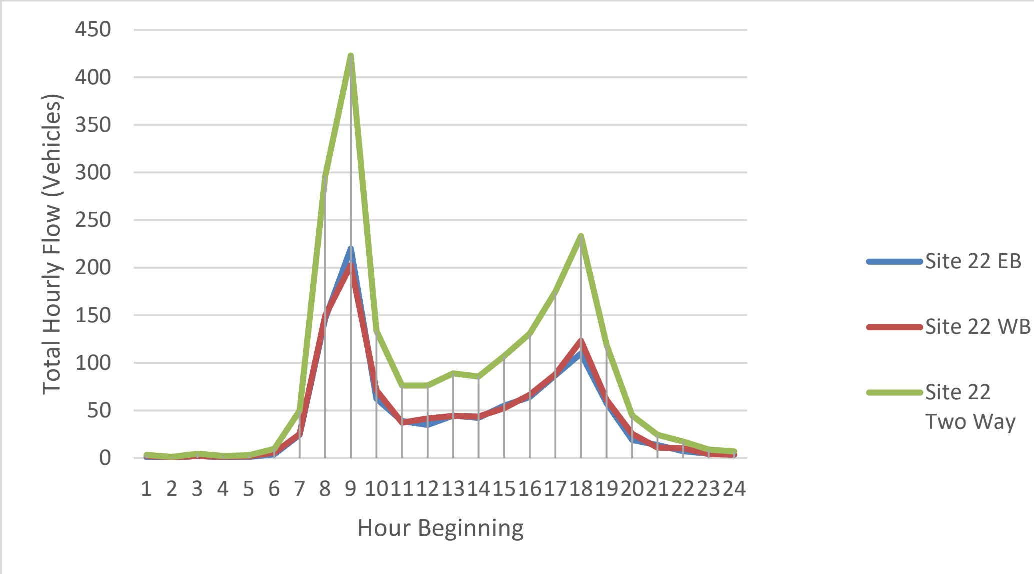


Figure 2-23: ATC Site 22 Average Tues-Thurs Flow Profile

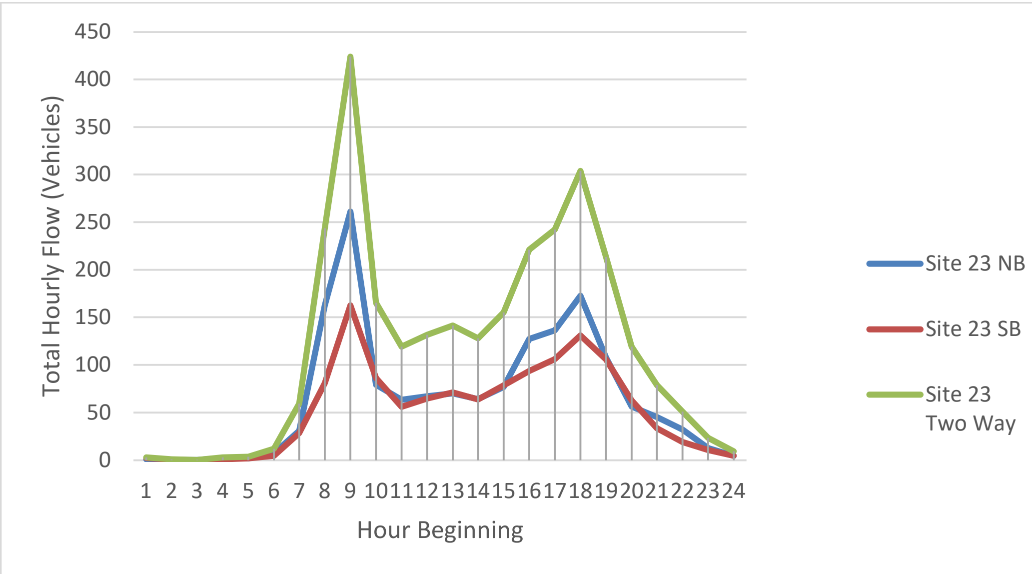


Figure 2-24: ATC Site 23 Average Tues-Thurs Flow Profile

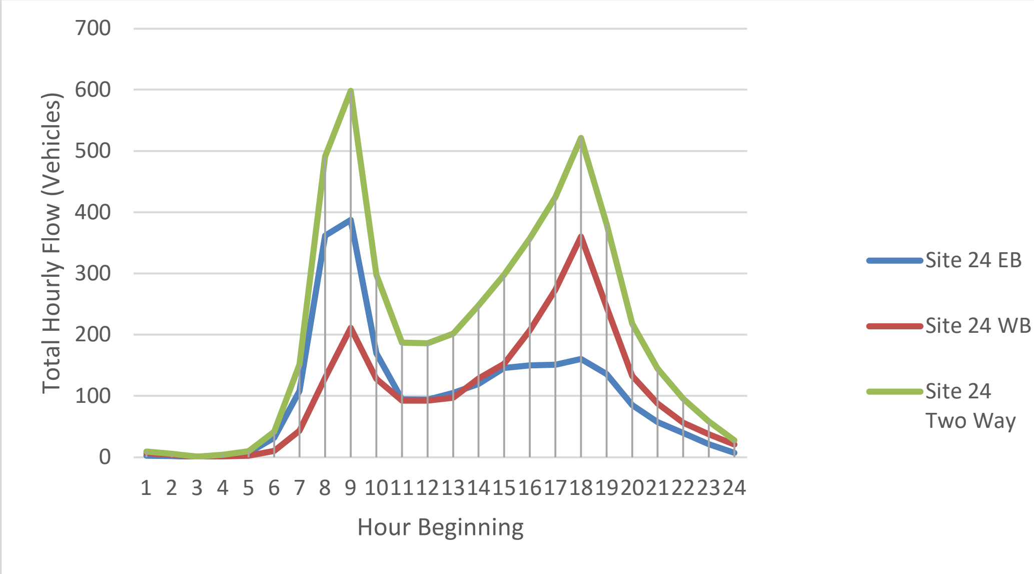


Figure 2-25: ATC Site 24 Average Tues-Thurs Flow Profile

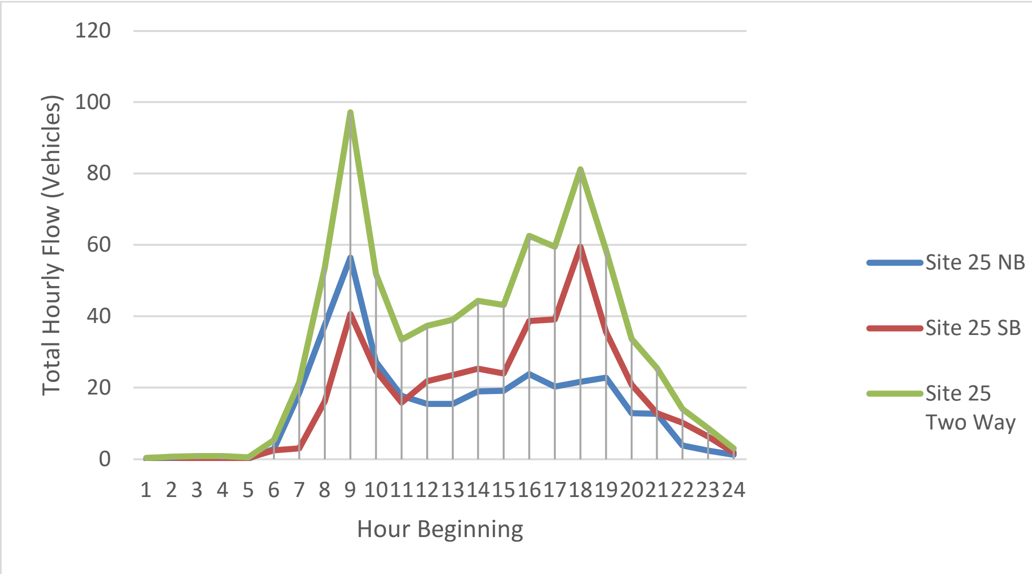


Figure 2-26: ATC Site 25 Average Tues-Thurs Flow Profile

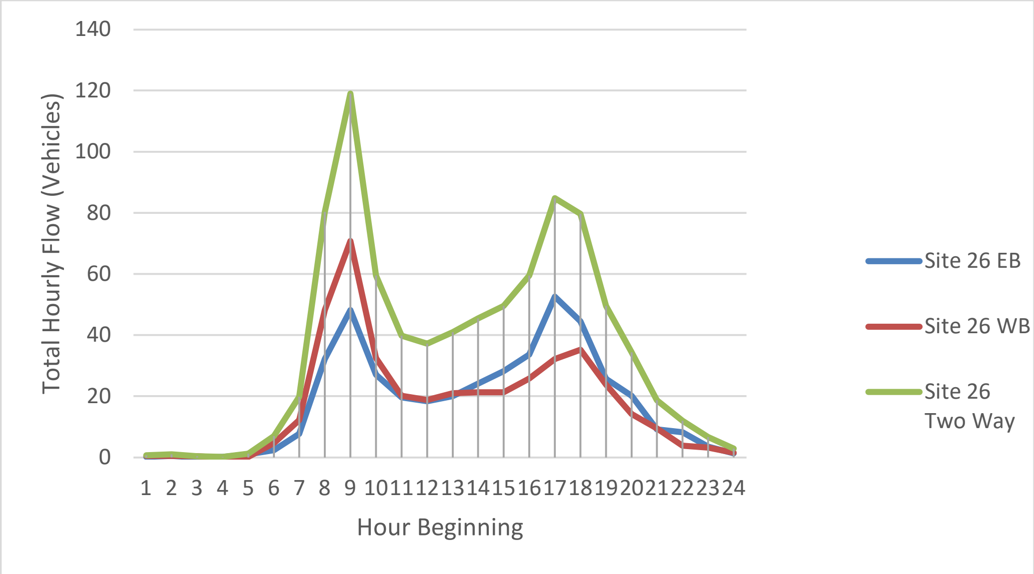


Figure 2-27: ATC Site 26 Average Tues-Thurs Flow Profile

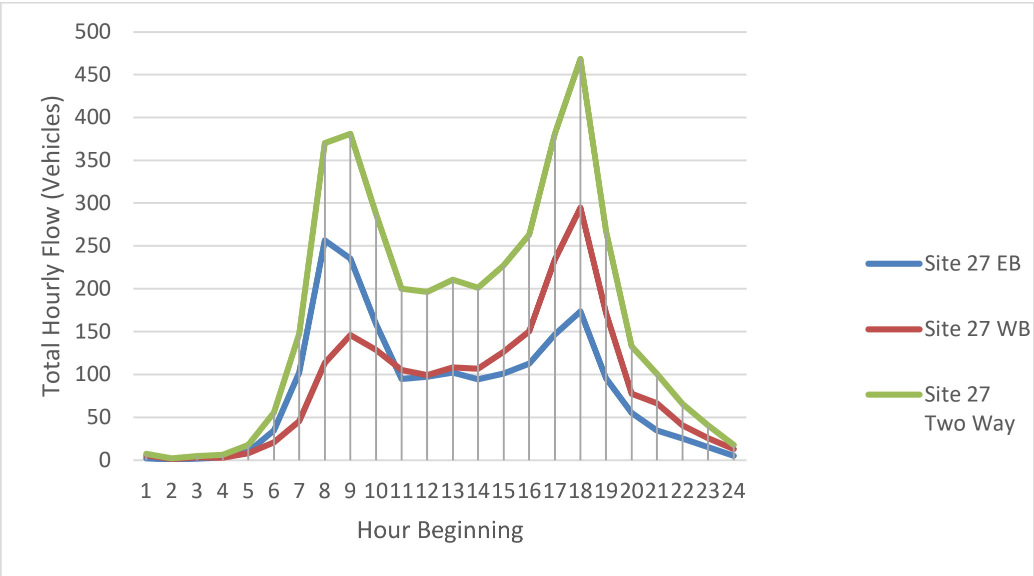


Figure 2-28: ATC Site 27 Average Tues-Thurs Flow Profile

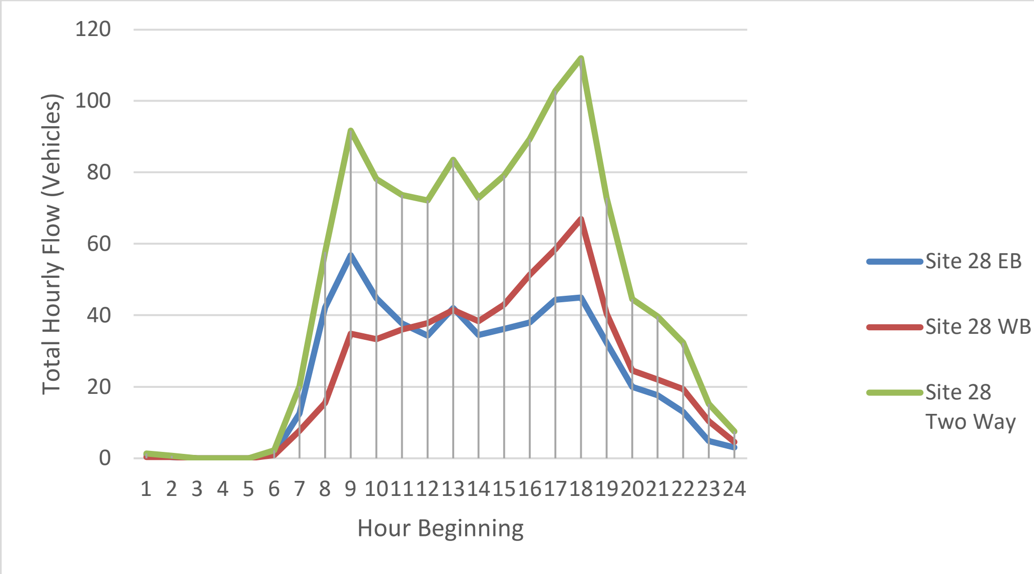


Figure 2-29: ATC Site 28 Average Tues-Thurs Flow Profile

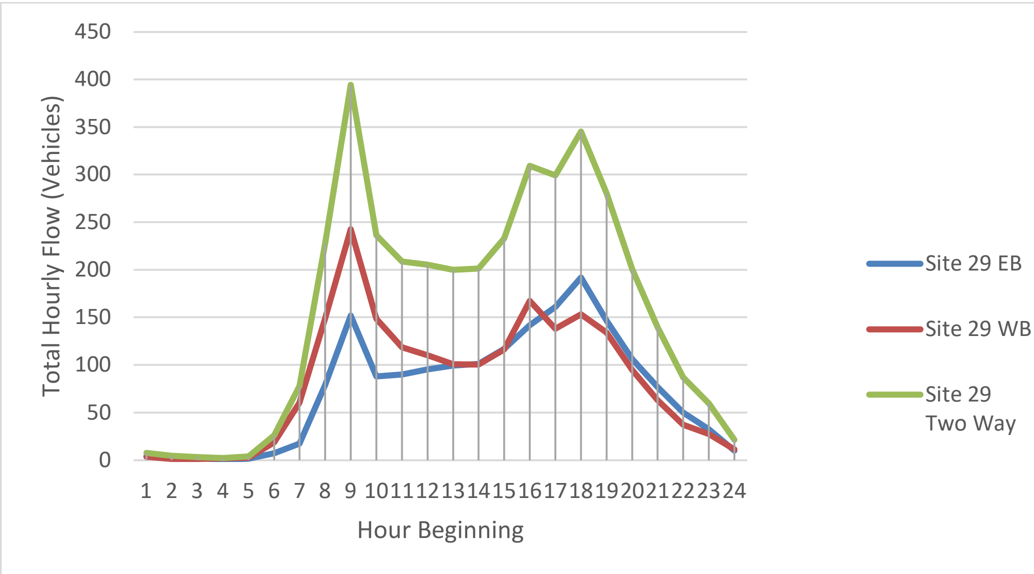


Figure 2-30: ATC Site 29 Average Tues-Thurs Flow Profile

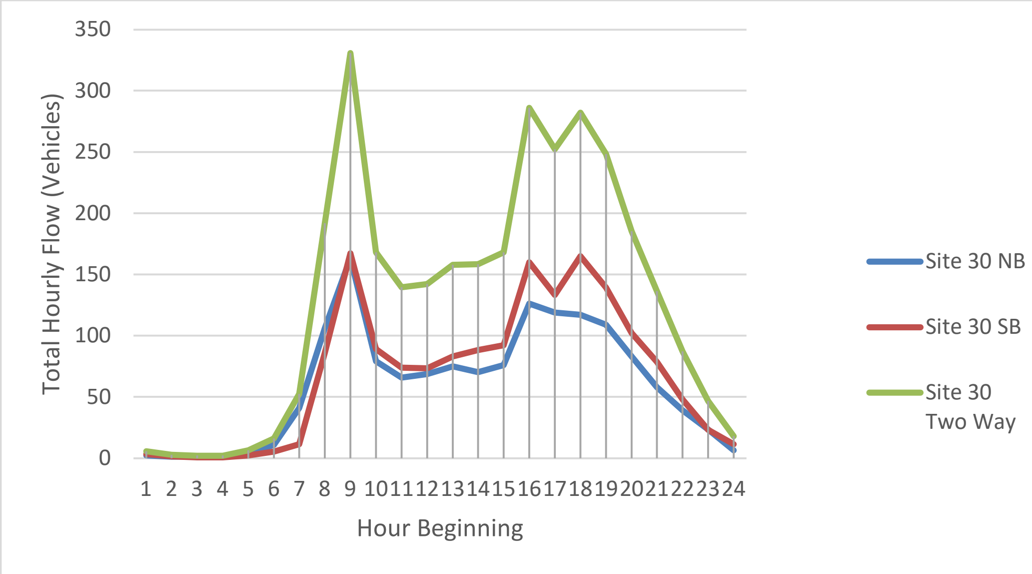


Figure 2-31: ATC Site 30 Average Tues-Thurs Flow Profile

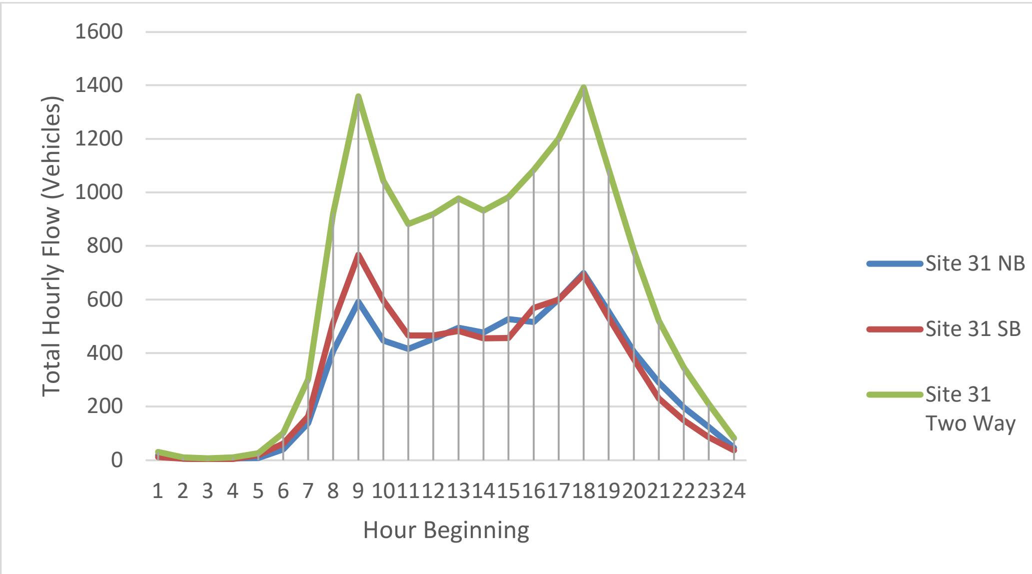


Figure 2-32: ATC Site 31 Average Tues-Thurs Flow Profile

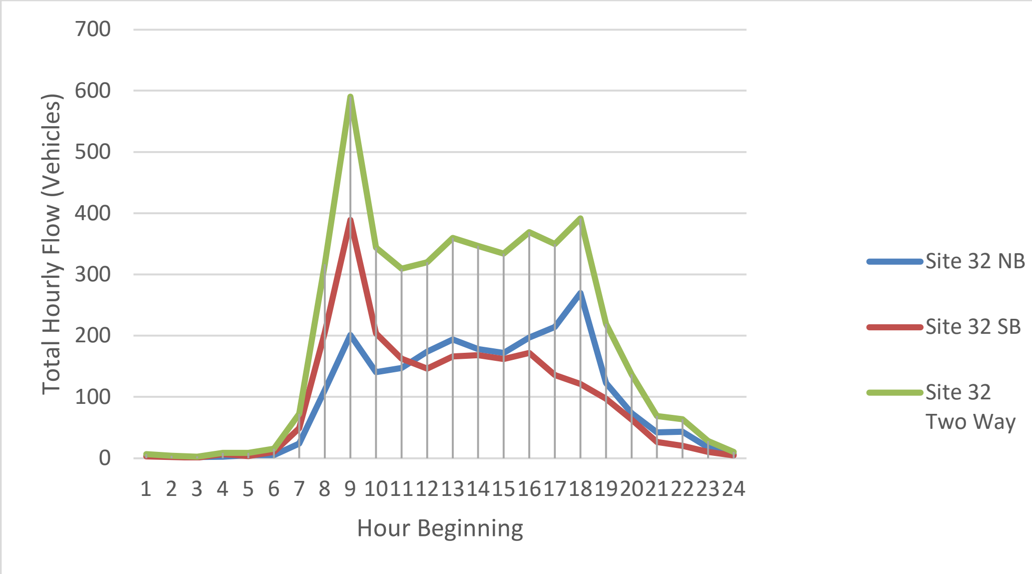


Figure 2-33: ATC Site 32 Average Tues-Thurs Flow Profile

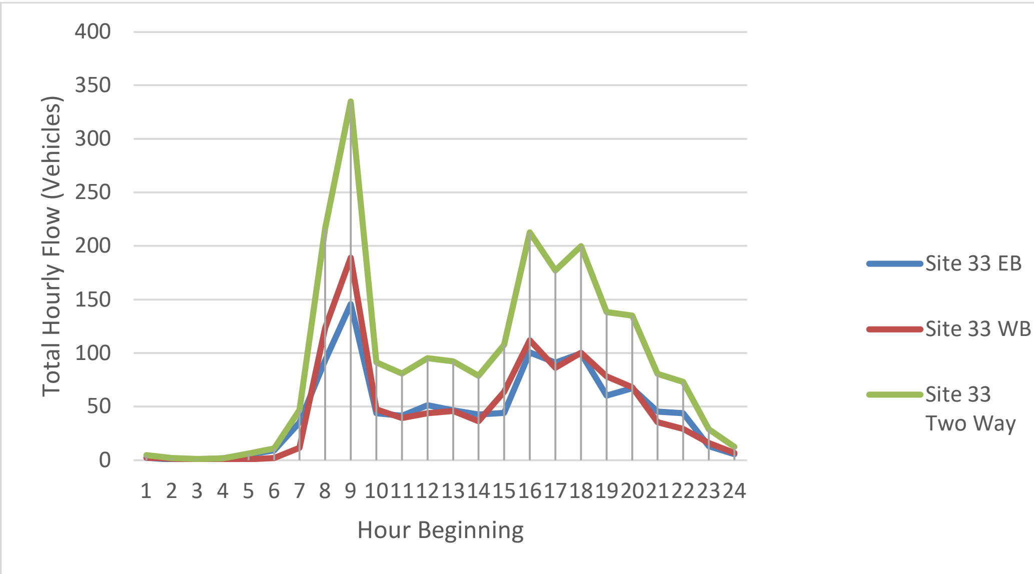


Figure 2-34: ATC Site 33 Average Tues-Thurs Flow Profile

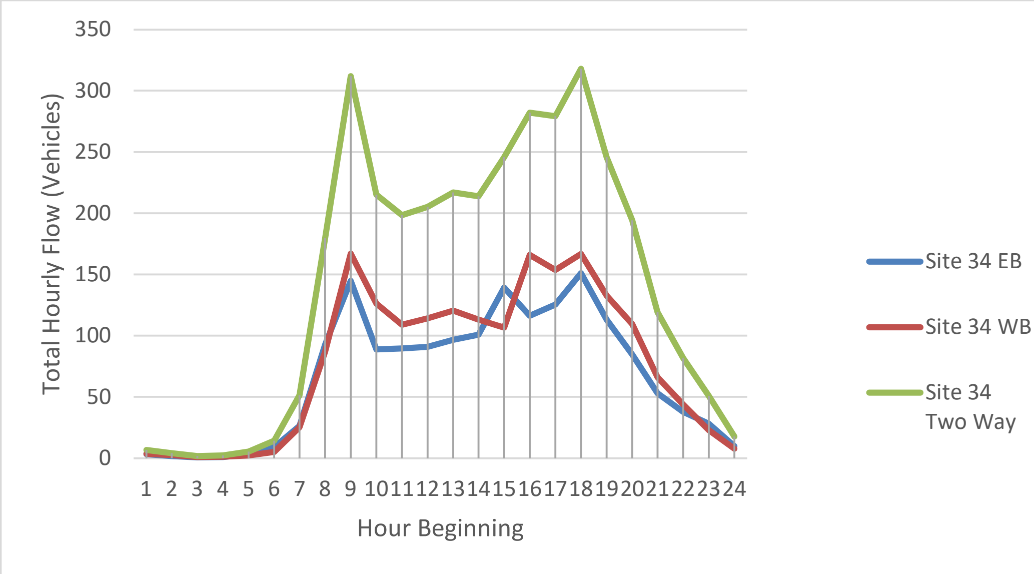


Figure 2-35: ATC Site 34 Average Tues-Thurs Flow Profile

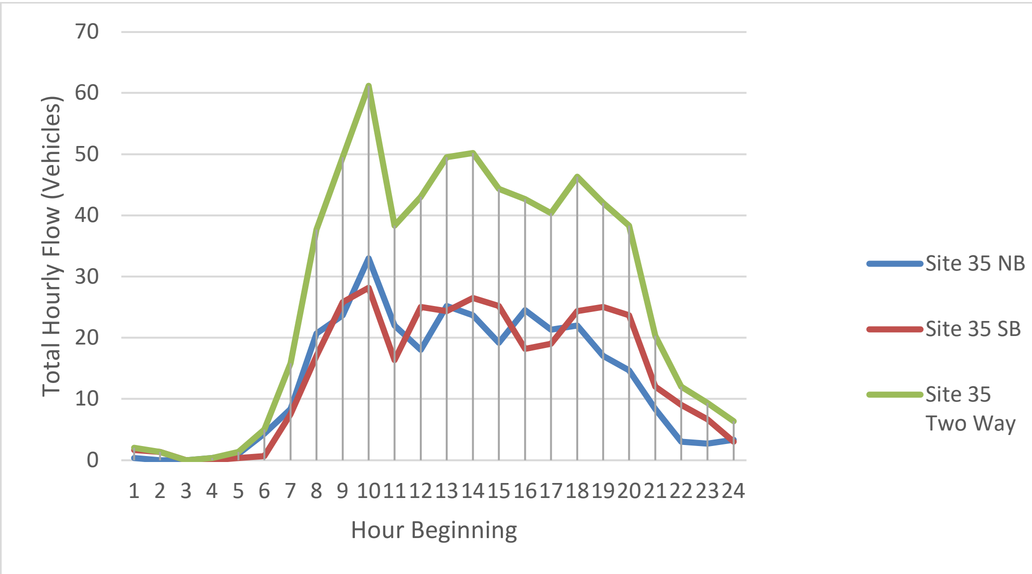


Figure 2-36: ATC Site 35 Average Tues-Thurs Flow Profile

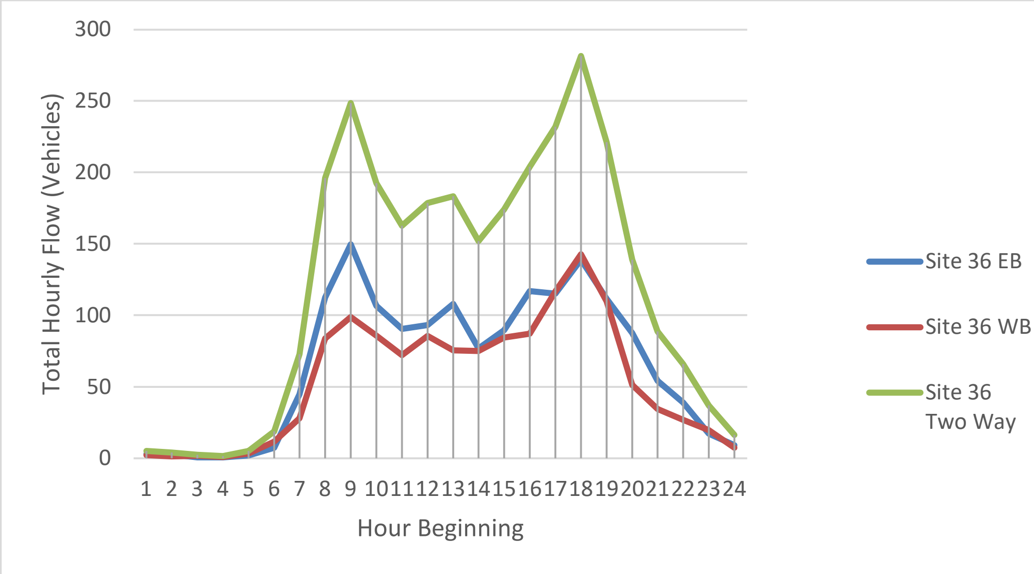


Figure 2-37: ATC Site 36 Average Tues-Thurs Flow Profile

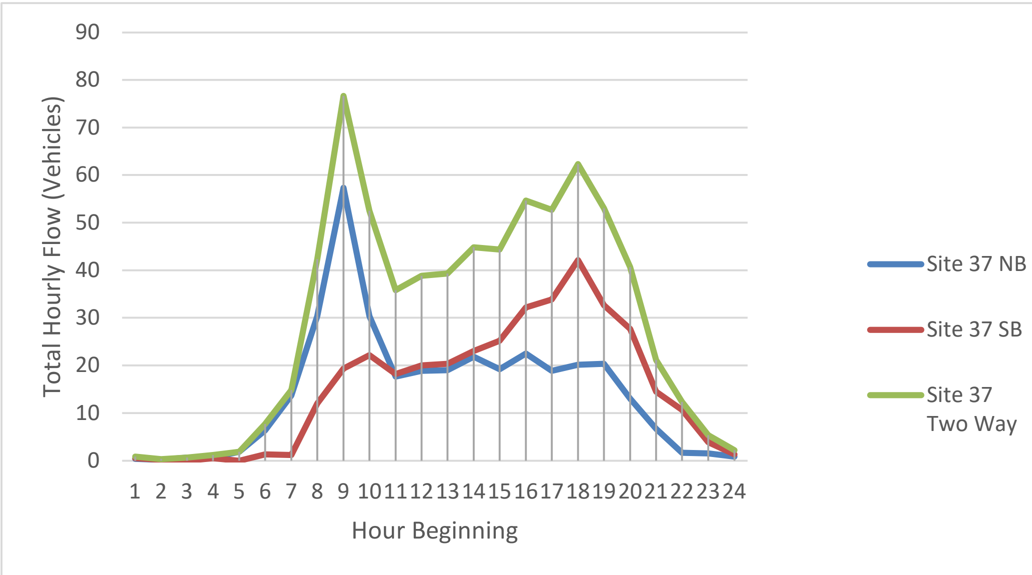


Figure 2-38: ATC Site 36 Average Tues-Thurs Flow Profile

2.5 Existing West Sussex County Council ATC Data

2.5.1 WSCC collect data across the county, including a number of sites within Horsham District. This data has been obtained from WSCC. These sites collect data permanently.

2.5.2 Existing ATC survey locations to be used in the model validation process are summarised in Table 2-2.

2.5.3 The flows at these sites are total flows and are not classified by vehicle type. The period of data collection aims to capture the latest existing data within a neutral month and over a two-week period, although this has not been used in the validation process of the model the counts will aid in the calibration process.

Table 2-2: Existing WSCC ATC locations

ATC Reference	Location
WSCC-1	Henfield A281, Brighton Rd. Just E. of Mill Dr
WSCC -7	A272 Cowfold, Cowfold Rd, Just E. of Fairfield Cot
WSCC -8	A281 Cowfold, Henfield Rd., By Singers Farm
WSCC -10	Pulborough, A283 Stopham Rd., W. of A29
WSCC -11	A24 Shipley, Worthing Road South of A24a272 Jct
WSCC -13	A24 Kingsfold, Layby 12 Mile S. of Surrey Border
WSCC -14	A29 Billingshurst, N. Of Town Just S. of New Rd
WSCC-16	A29 Bognor Road (Just S. of Surrey County Border)
WSCC -17	A281 Rudgwick, Guilford Rd., by House Called Hyes
WSCC -19	B2139 Amberley, New Barn Rd, E. of Railway Station
WSCC -23	A24, Washington, Horsham Rd
WSCC -27	A264 Faygate, Crawley Rd, by Park Road
WSCC -31	A24 Horsham, Broadbridge Heath S. of A281 Roundabout
WSCC -35	A281 Horsham, Brighton Rd S. of St Leonards Rd
WSCC -35	A281 Horsham, Brighton Rd S. of St Leonards Rd
WSCC -36	Horsham, Kings Rd S. of St Georges Gds
WSCC -37	B2237 Horsham, Warnham Rd N. of The Dog & Bacon Public House
WSCC -38	A281 Horsham, Guilford Rd East of Merryfield Drive
WSCC -39	B2237 Horsham, Worthing Rd N. of Tower Hill
WSCC -39	B2237 Horsham, Worthing Rd N. of Tower Hill
WSCC -40	B2195 Horsham, Harwood Rd Just E. of Elgin Close

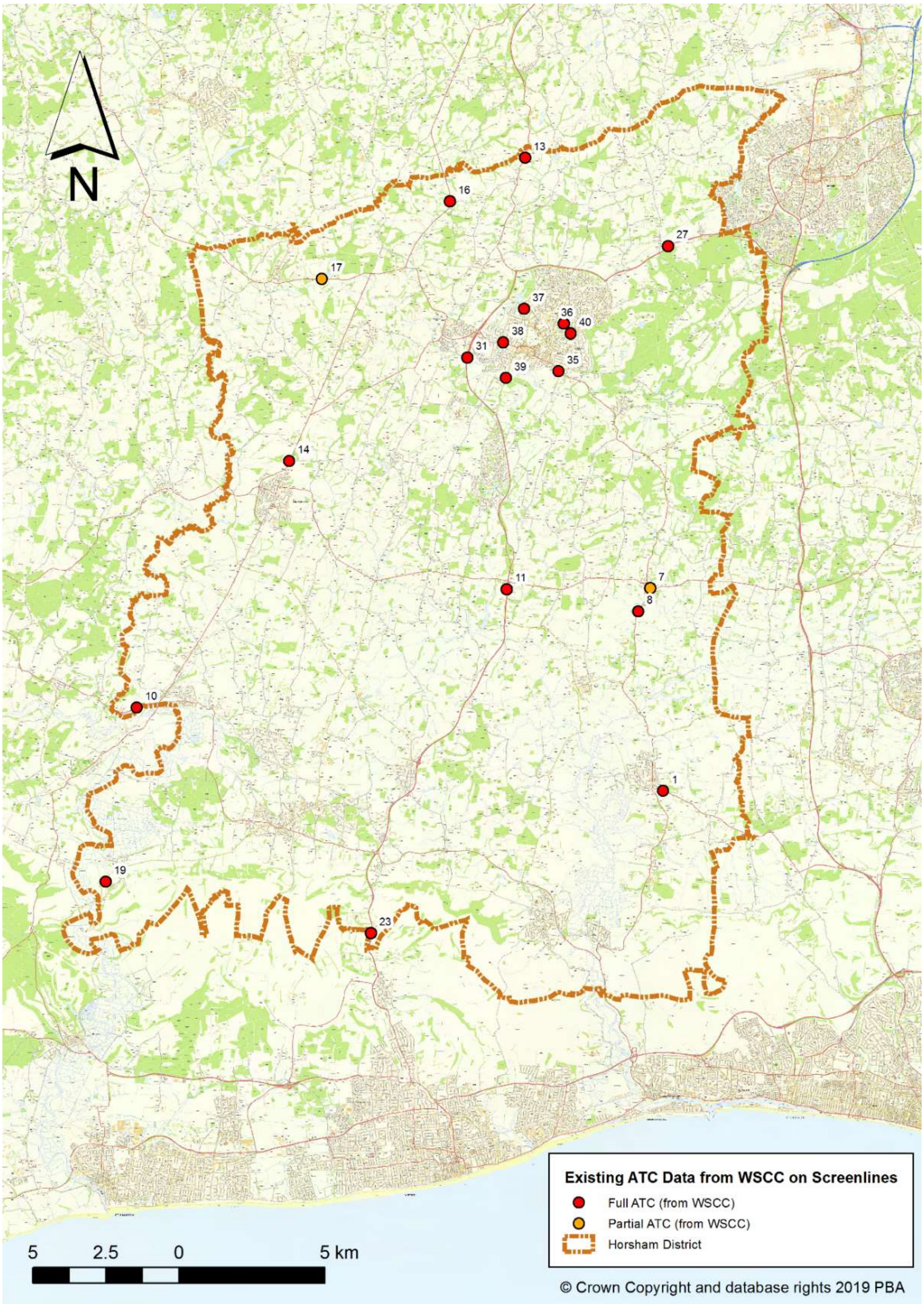


Figure 2-39 – Existing WSCC ATC Sites

2.6 Highways England ATC Data

2.6.1 To inform flow calibration and validation on the Highways England network within the Horsham model area, count data was obtained from Highways England's open data source website[2]. The data was downloaded for May 2019 and analysed for weekday (Tuesday to Thursday) AM and PM Peak period flows.

2.6.2 The data covers the M23 and A23 which is part of the SRN managed by Highways England. The locations of the data collected is listed in Table 2-3 and shown in Figure 2-40. The data is classified by vehicle length in metres making it possible to discern vehicle classes into car (<5.2m), LGV (5.21-6.6m), OGV1 (6.61-11.6m) and OGV2 (above 11.6m).

Table 2-2: Highways England Count Locations

ATC Reference.	Site Location
HE-1	M23 J10A NB On-slip
HE-2	M23 J10A SB Off-slip
HE-3	M23 between J10A and J11
HE-4	M23 J11 NB On-slip
HE-5	M23 J11 NB
HE-6	M23 J11 SB
HE-7	A23 SB between B2110 and B2114
HE-8	A23 SB B2114 Off-slip
HE-9	A23 NB between B2110 and M23/A264
HE-10	A23 SB between B2110 and B2115
HE-11	A23 NB at B2115 Junction
HE-12	A23 NB at B2115 Off-slip
HE-13	A23 NB at B2115 On-slip
HE-14	A23 NB at A272 Off-slip
HE-15	A23 NB at A272
HE-16	A23 SB at A272 Off-slip
HE-17	A23 SB at A272
HE-18	A23 SB at A2300
HE-19	A23 SB at A2300 Off-slip
HE-20	A23 NB at A2300 Off-slip
HE-21	A23 NB at A2300
HE-22	A23 SB at A281 On-Slip
HE-23	A23 SB at A281
HE-24	A23 NB at A281 Off-slip

[2] http://webtris.highwaysengland.co.uk/

ATC Reference.	Site Location
HE-25	A23 NB at A281

2.6.1 24-hour flow profiles are plotted for a number of selected Highways England (HE) maintained ATC sites for illustration. The flows were analysed for Tuesday to Thursday. The profiles can be seen for a selection of locations in Figures 2-41 to 2-51. The analysis suggests that the motorway links (M23) tend to peak earlier particularly in the AM peak between 07:00 to 08:00 compared to the more local network. The data has been collected over a one-month period between the 1st and the 31st of May 2018.

2.6.2 The data is classified by vehicle length in metres making it possible to discern vehicle classes into car (<5.2m), LGV (5.21-6.6m), OGV1 (6.61-11.6m) and OGV2 (above 11.6m).

2.6.3 Several HE ATC Sites listed above have been omitted due to data discrepancies and inconsistencies, these include HE Site 6,10,16,17,22 & 23

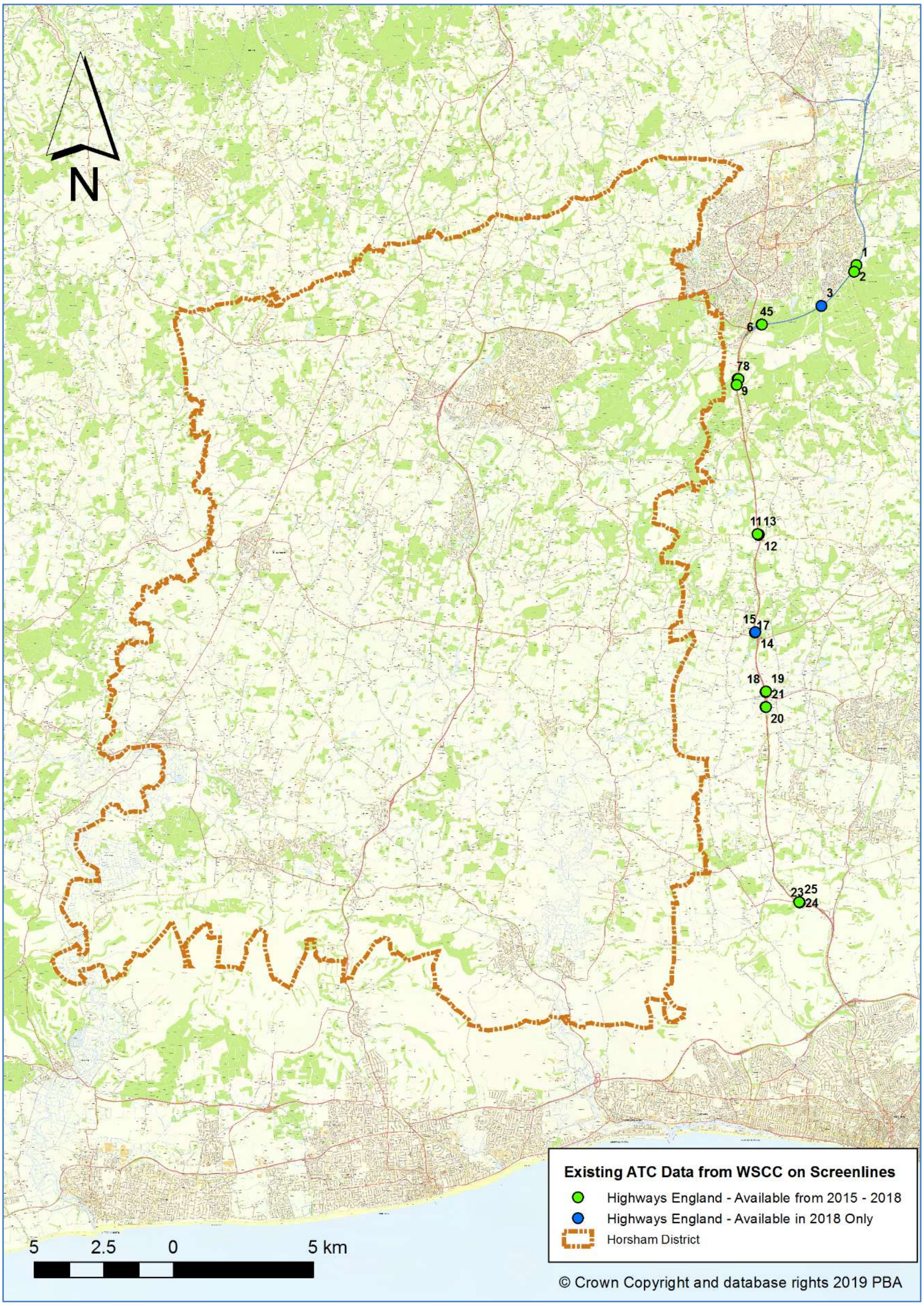


Figure 2-40: Highways England ATC data Locations

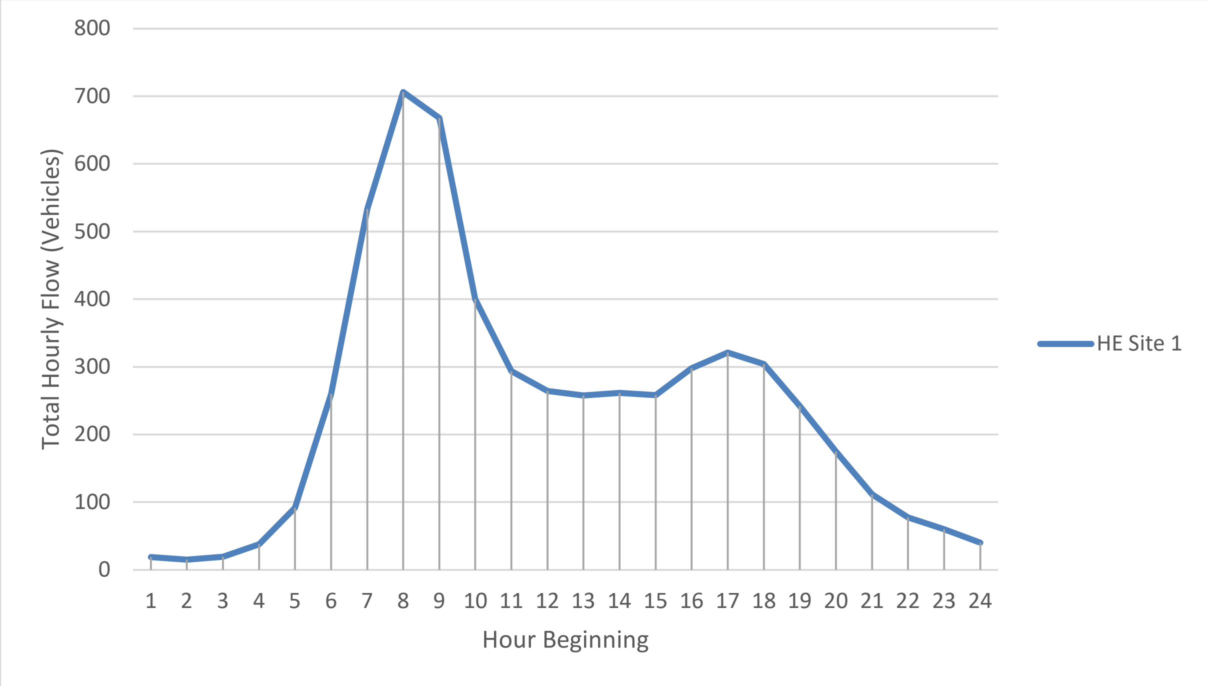


Figure 2-41: HE ATC Site 1

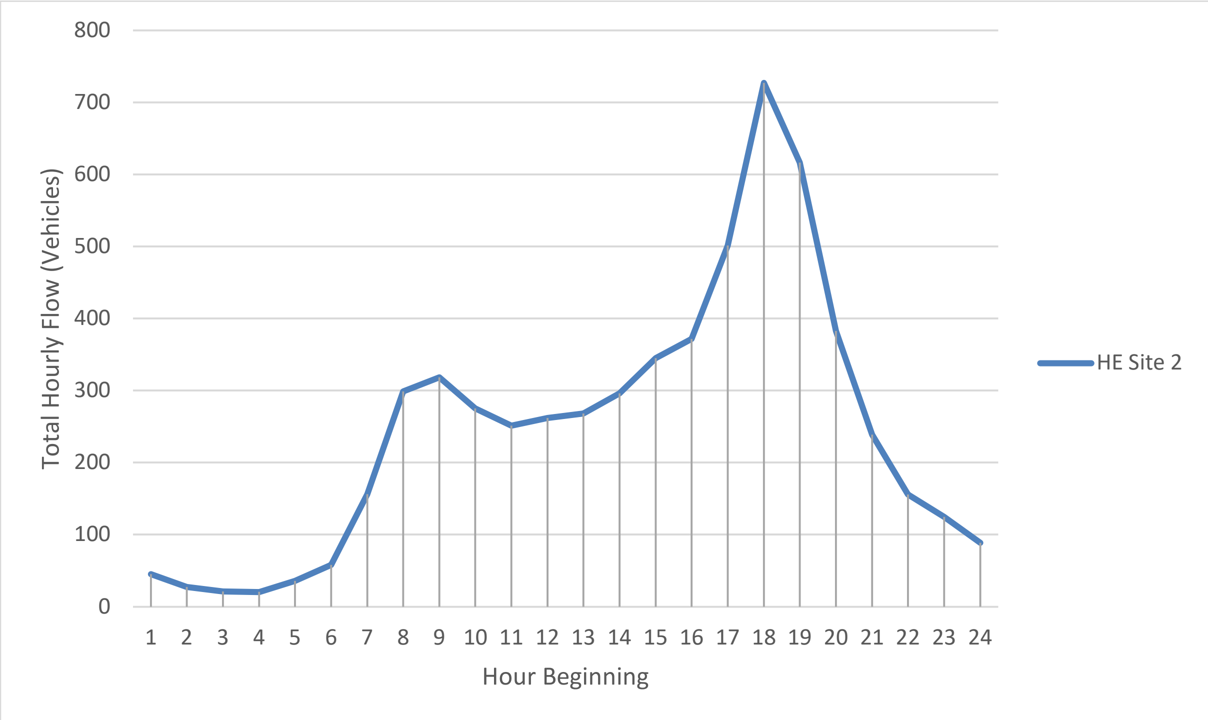


Figure 2-42: HE ATC Site 2

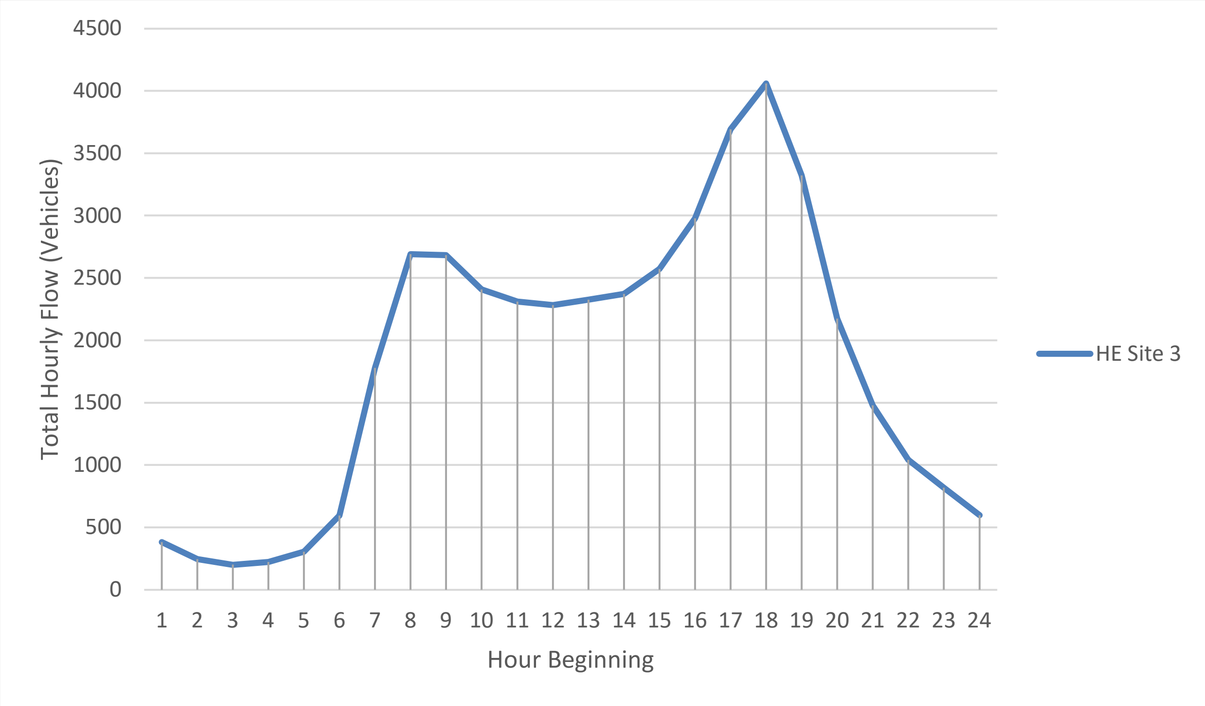


Figure 2-43: HE ATC Site 3

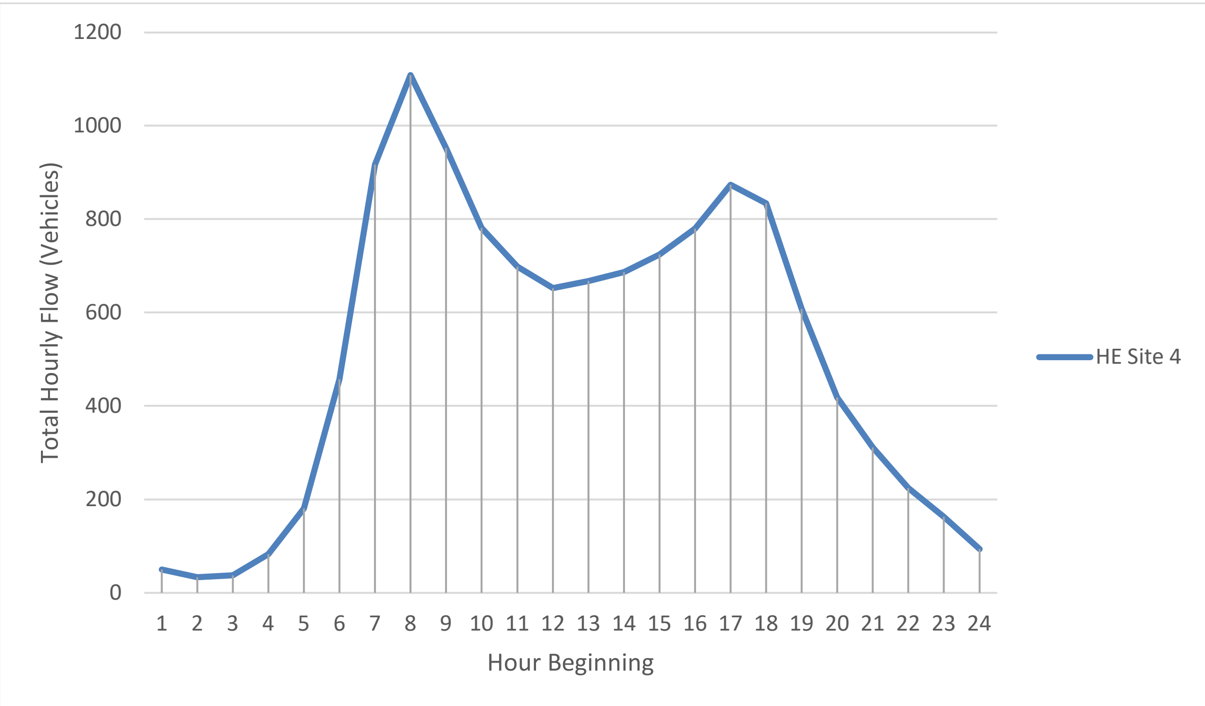


Figure 2-44: HE ATC Site 4

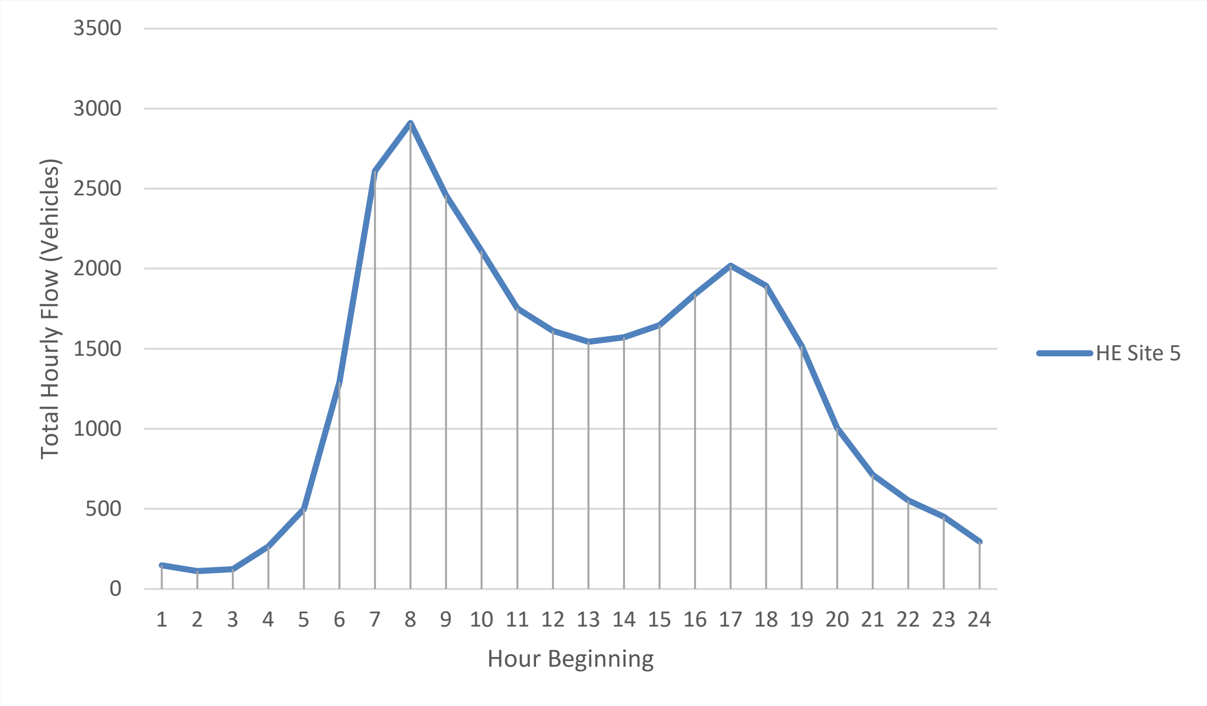


Figure 2-45: HE ATC Site 5

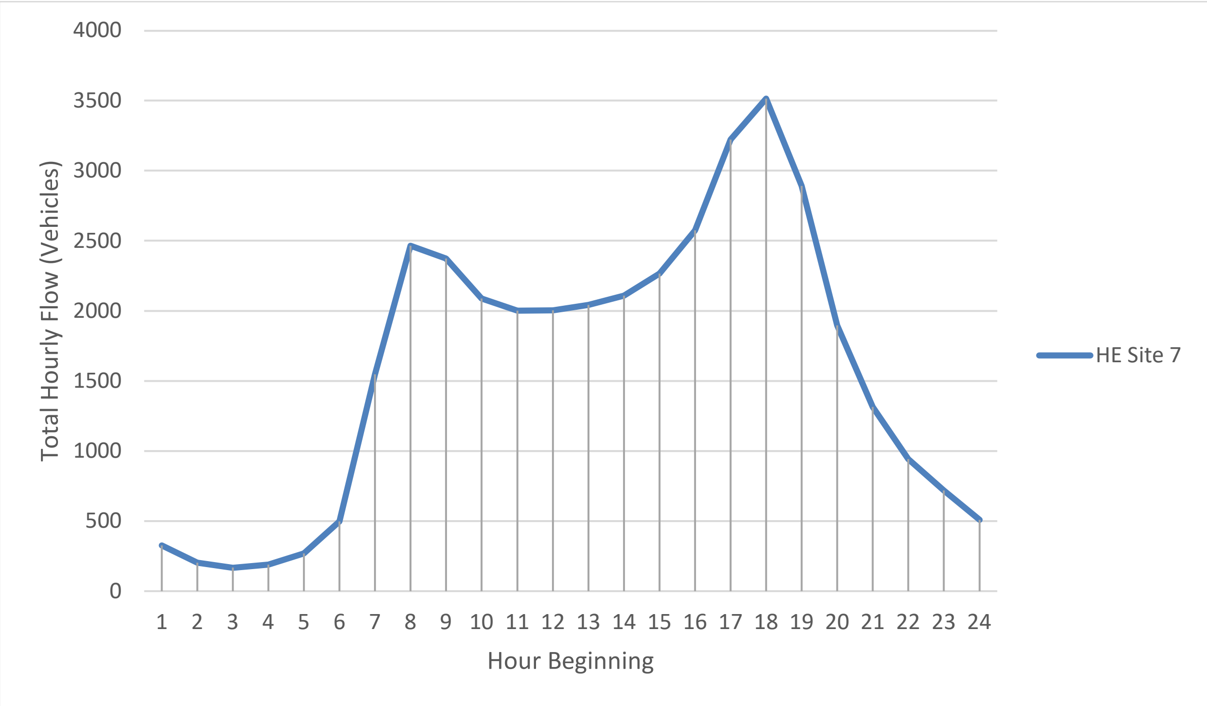


Figure 2-46: HE ATC Site 7

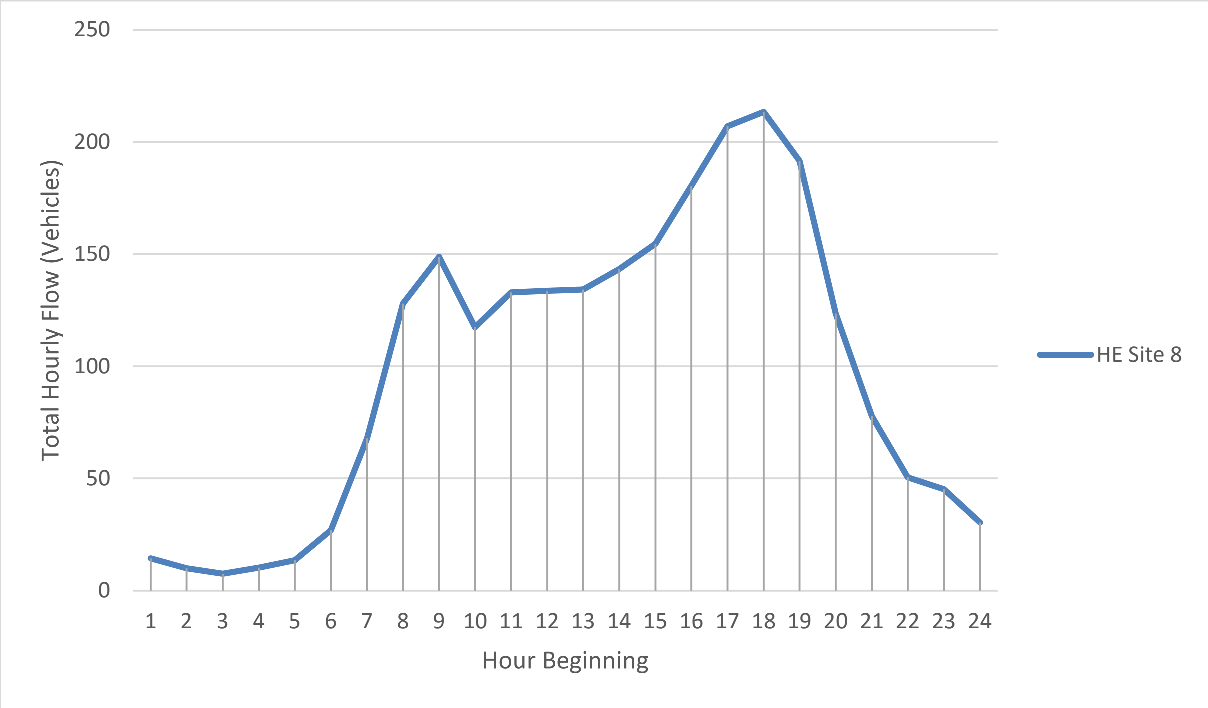


Figure 2-47: HE ATC Site 8

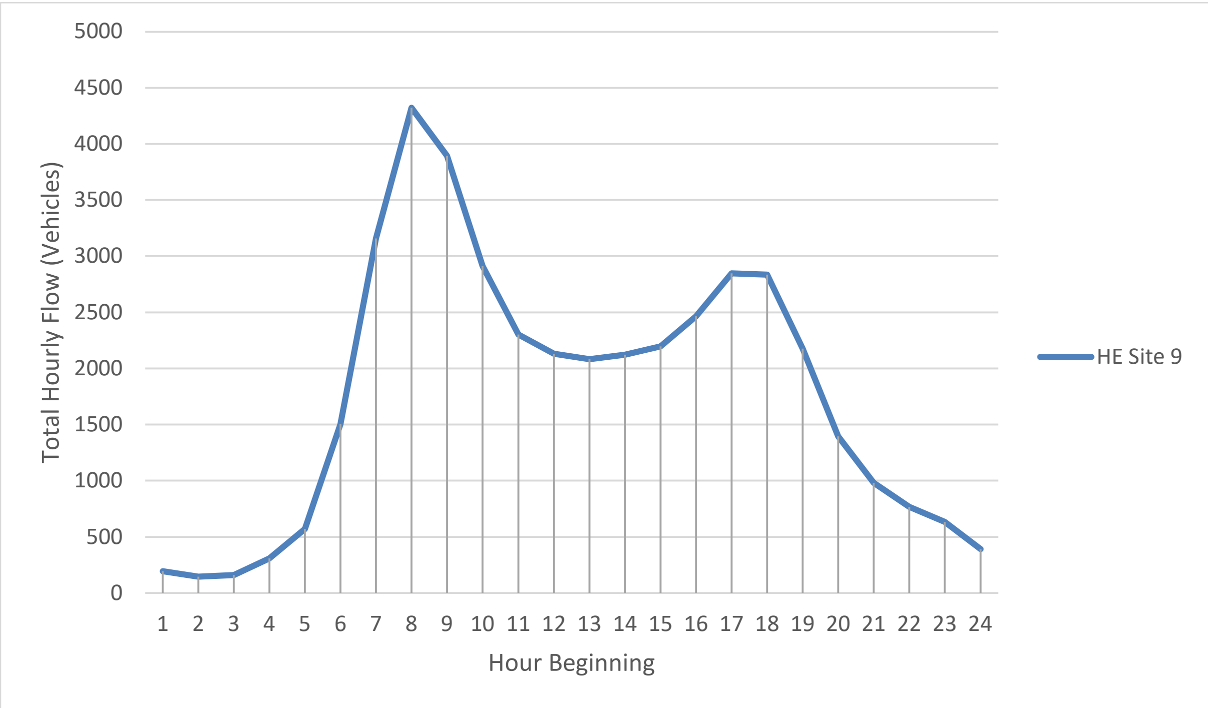


Figure 2-48: HE ATC Site 9

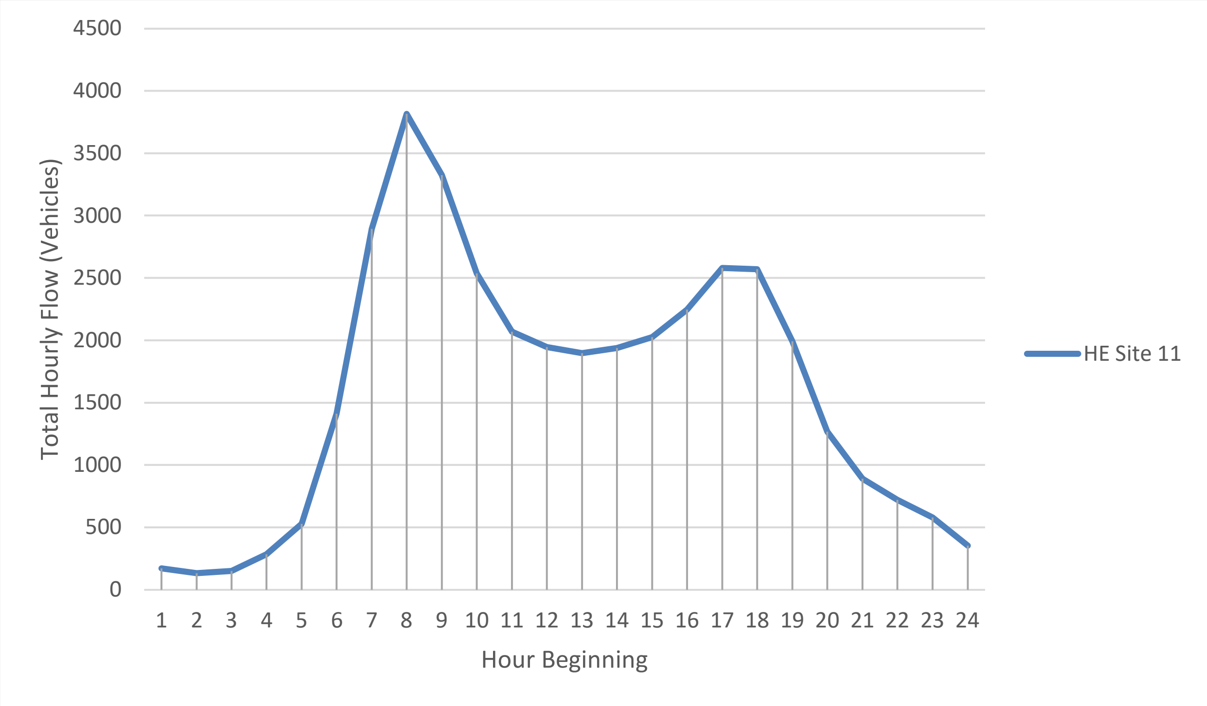


Figure 2-49: HE ATC Site 11

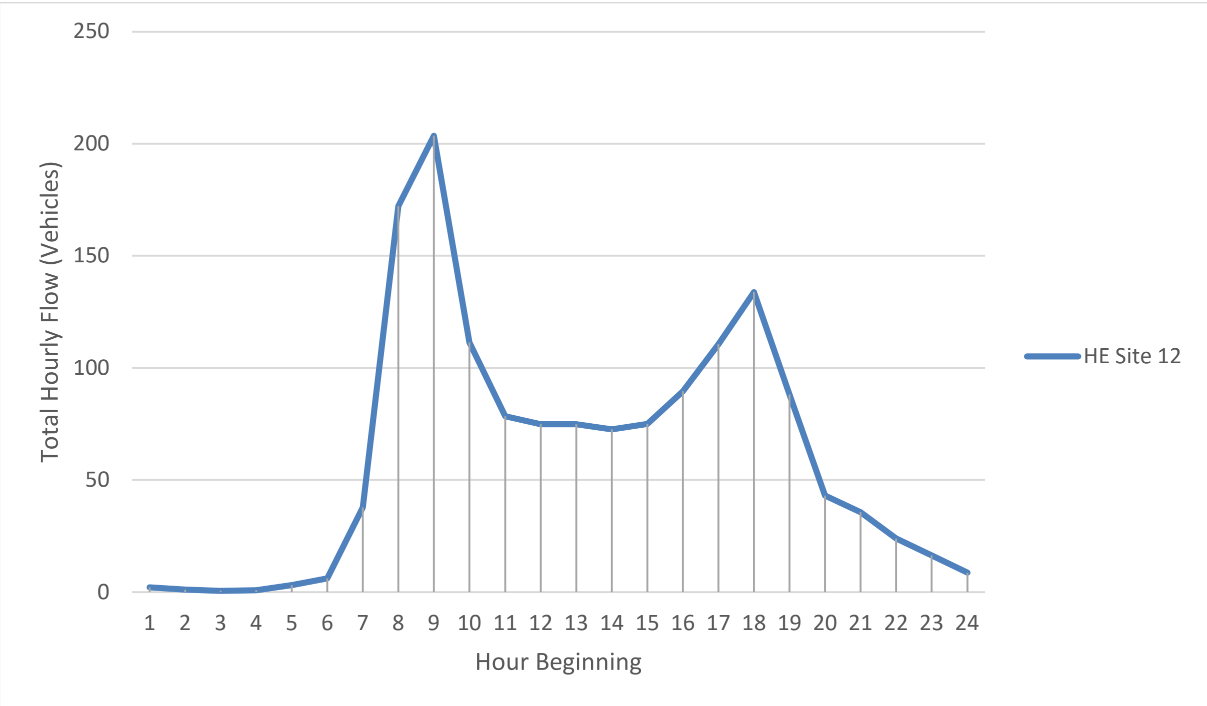


Figure 2-50: HE ATC Site 12

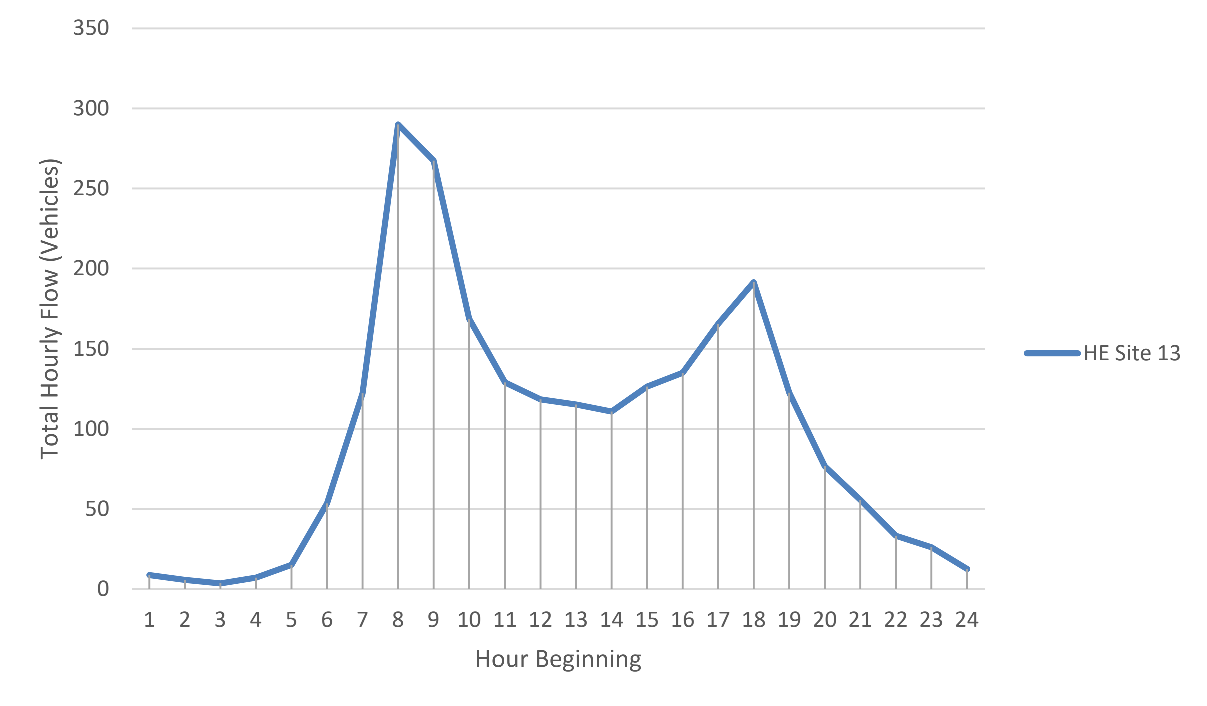


Figure 2-51: HE ATC Site 13

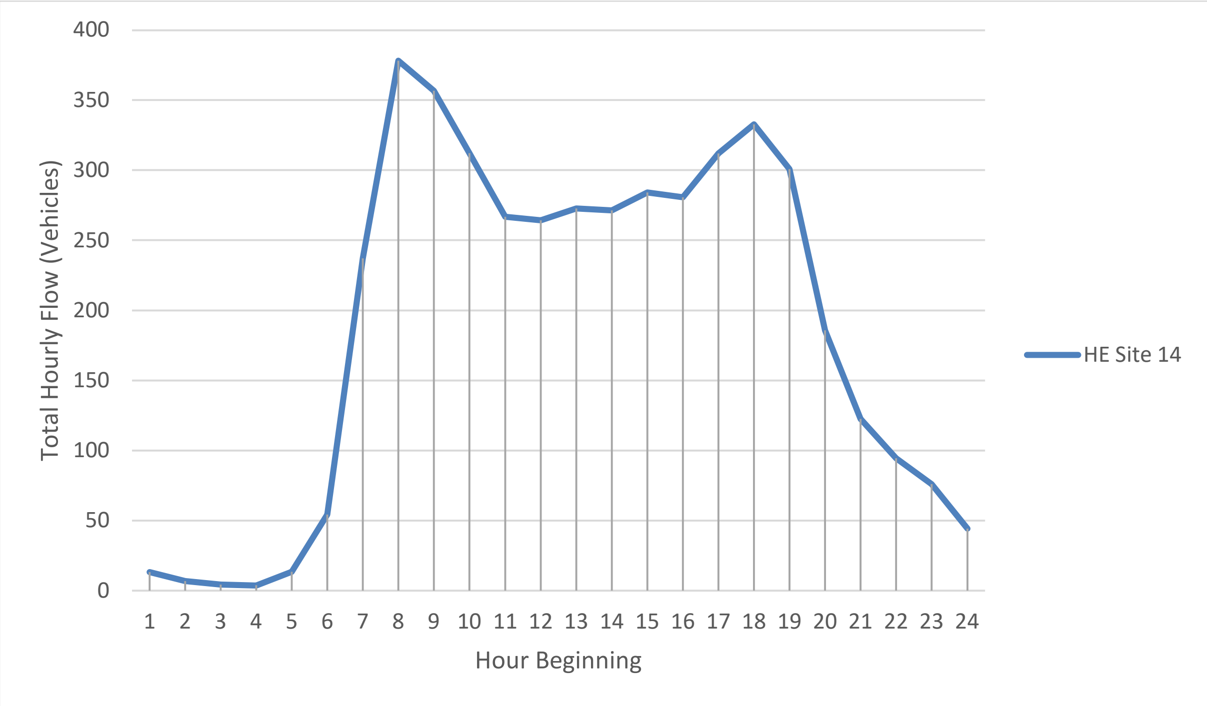


Figure 2-52: HE ATC Site 14

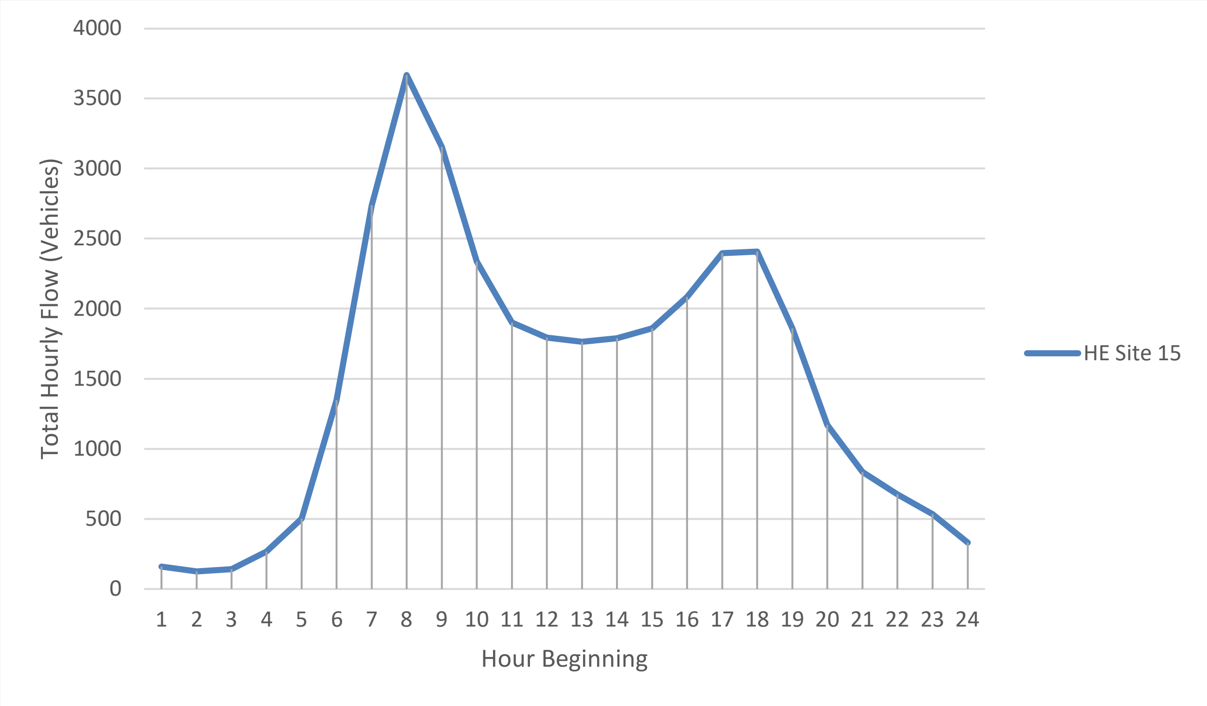


Figure 2-53: HE ATC Site 15

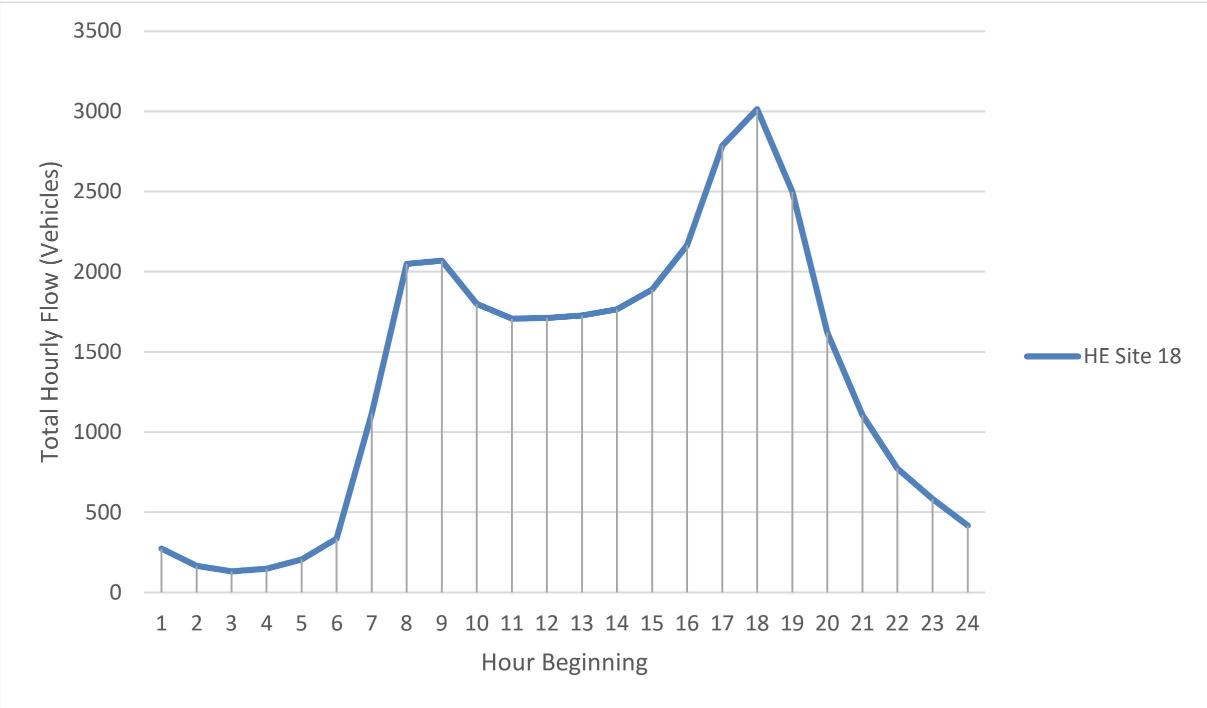


Figure 2-54: HE ATC Site 18

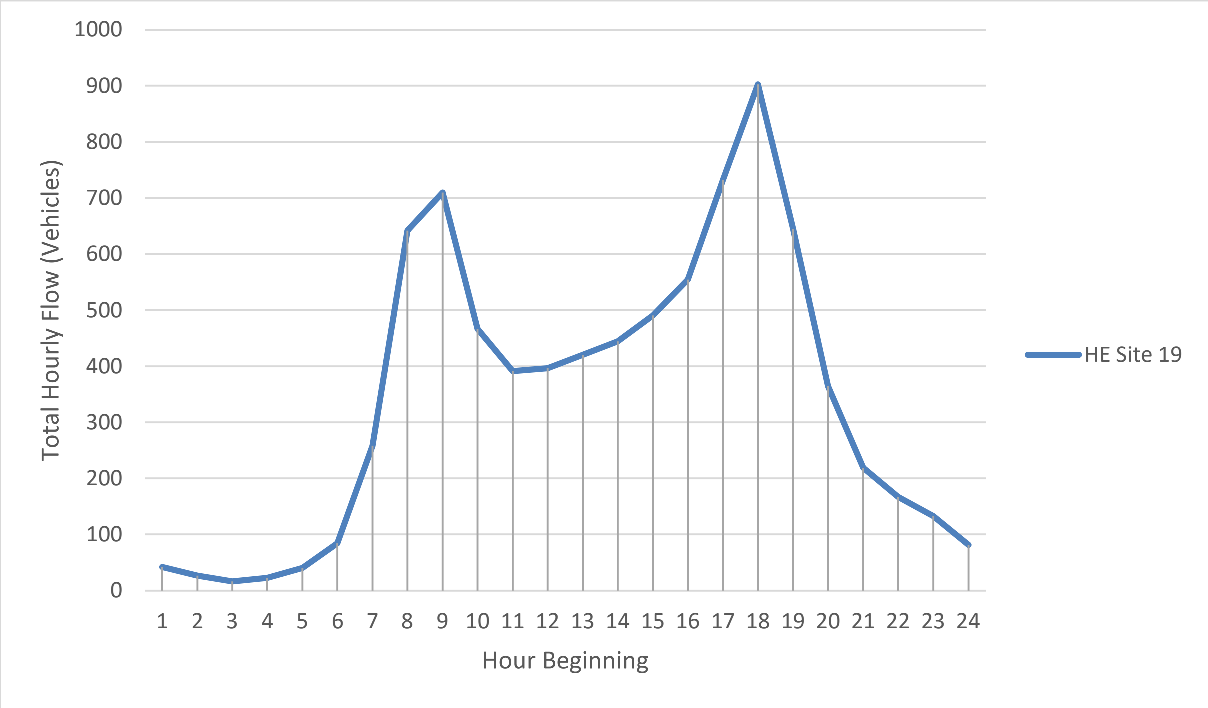


Figure 2-55: HE ATC Site 19

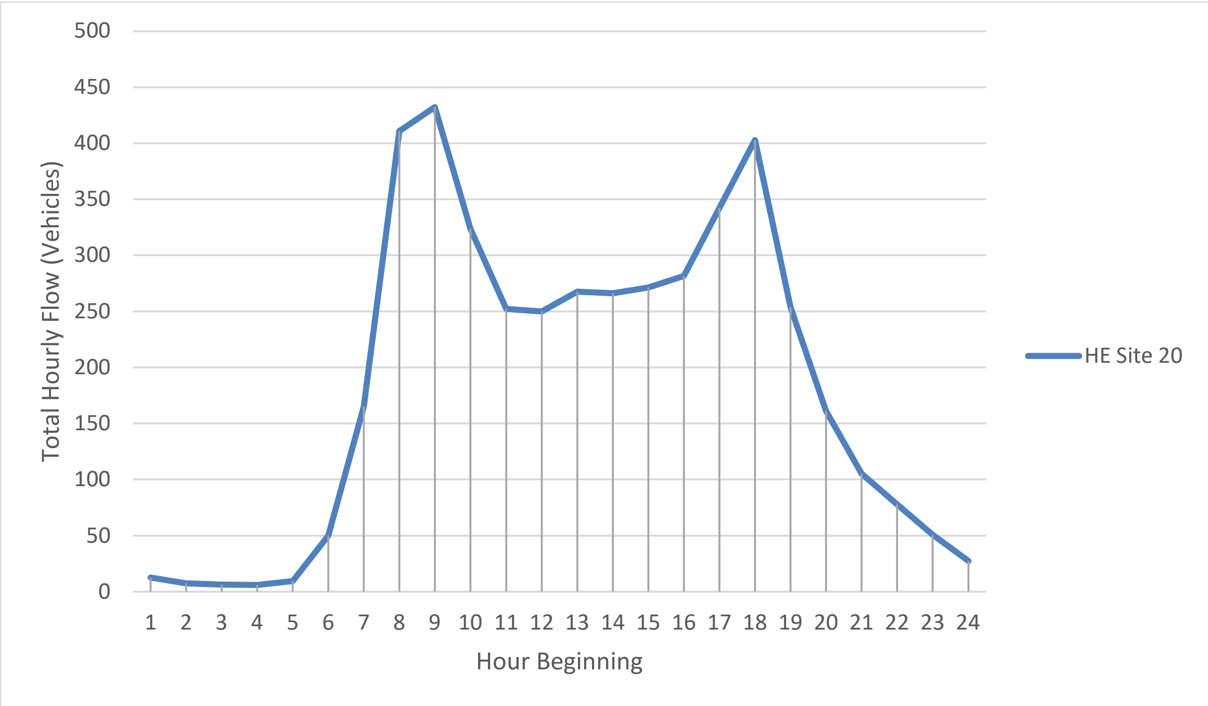


Figure 2-56 : HE ATC Site 20

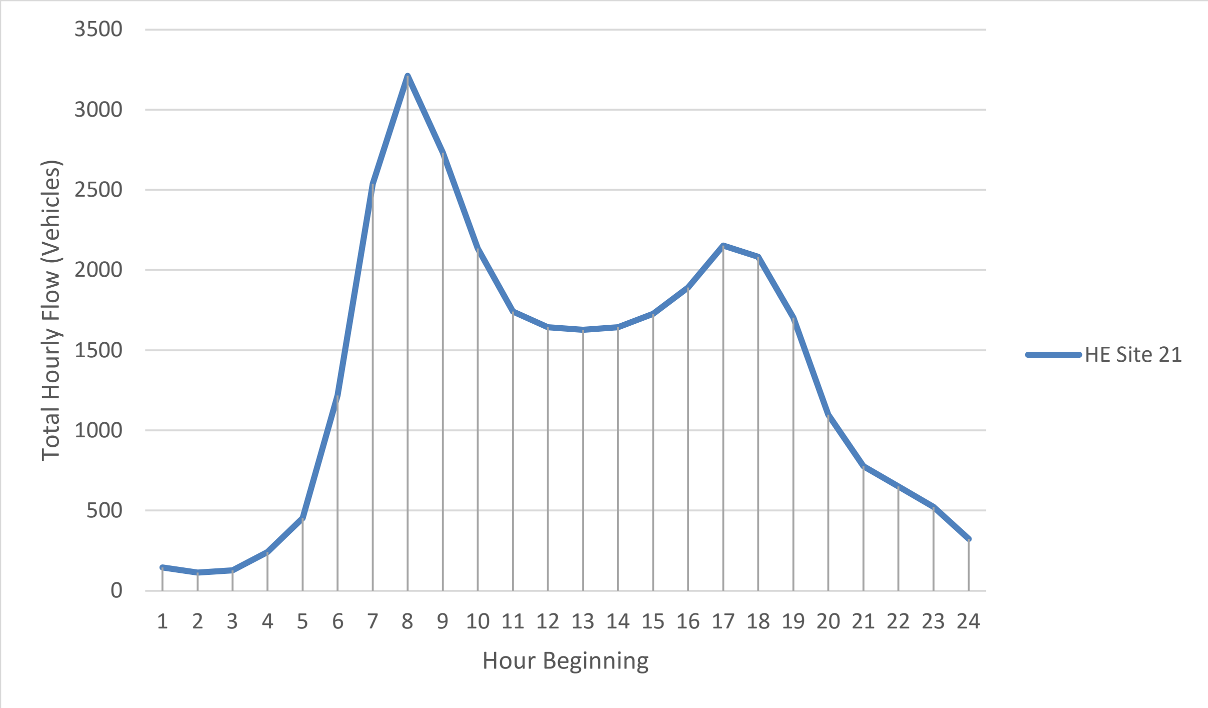


Figure 2-57: HE ATC Site 21

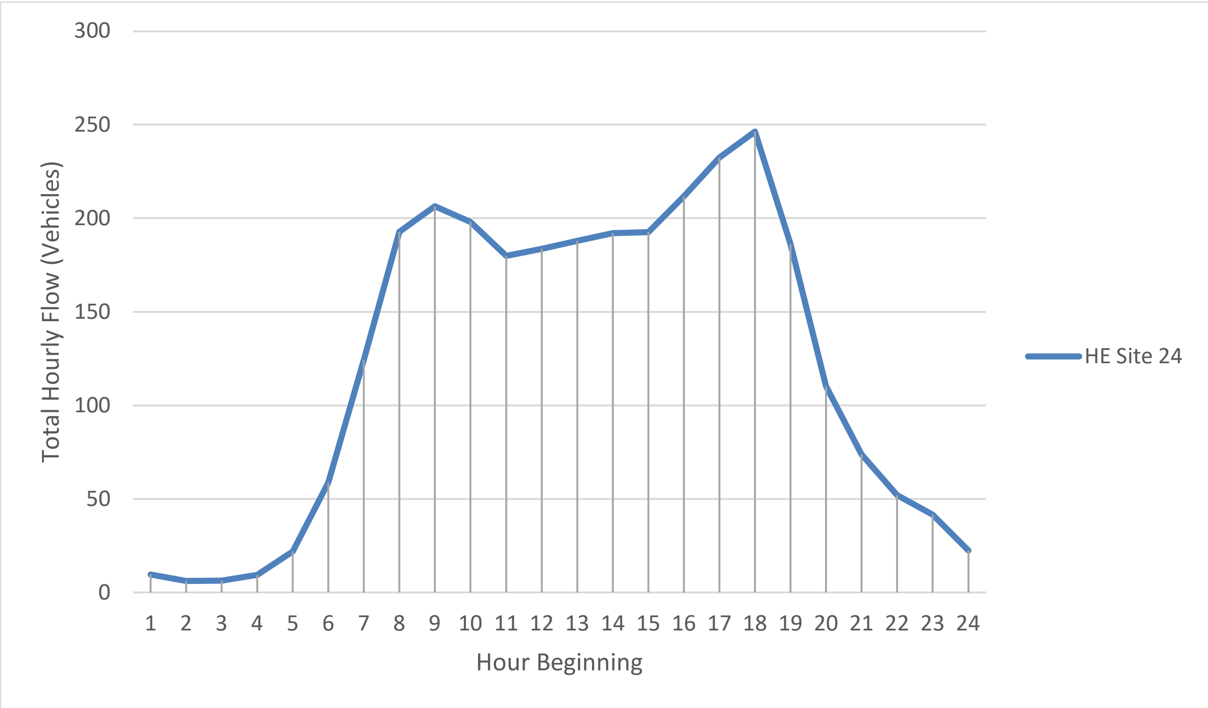


Figure 2-58: HE ATC Site 24

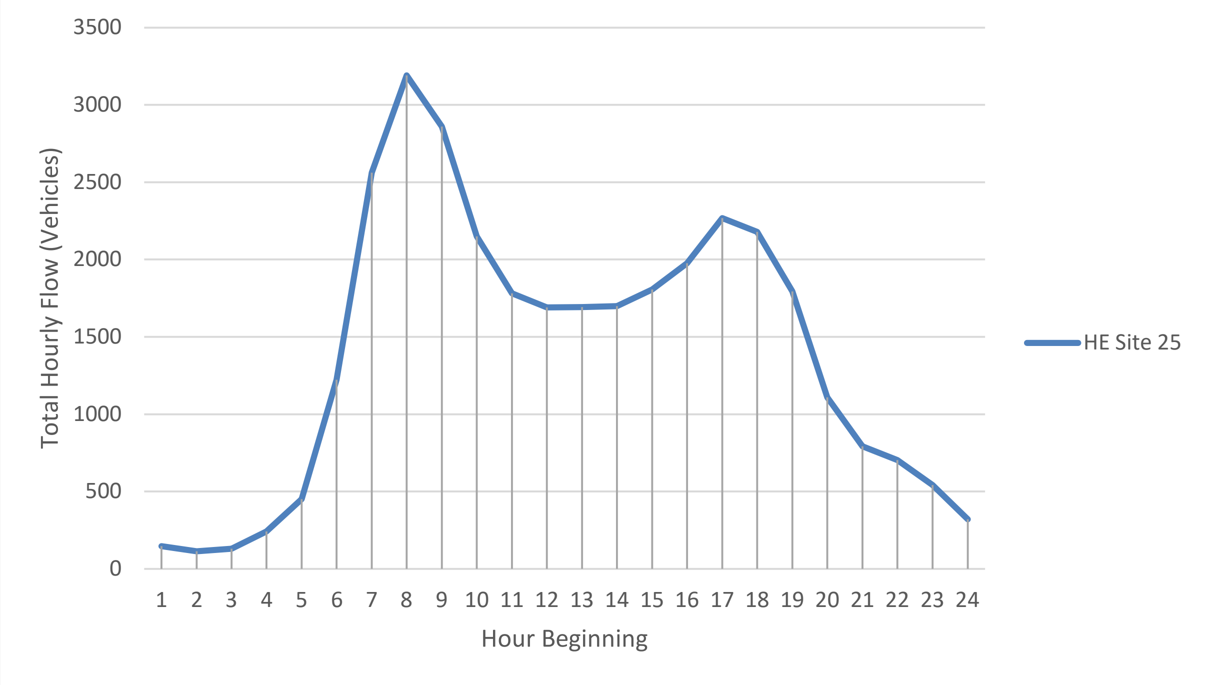


Figure 2-59: HE ATC Site 25

3 Manual Classified Counts

3.1 Overview

3.1.1 Manual classified counts (MCC) are counts undertaken at junctions, normally utilising video technology and are used to record all turning movements through a junction.

3.1.2 Surveys are undertaken for a 12-hour period (0700 to 1900) and for the purposes of the model development, should represent a typical average day, therefore avoiding school holidays and avoiding any major events or road works that could distort the data. Typically, they are undertaken on Tuesday, Wednesday or Thursday.

3.1.3 The surveys for the HTM were undertaken on Thursday the 23rd of May 2019.

3.1.4 The data has been fully classified into car, LGV, OGV1, OGV2, PSV, Motorcycles and cycles and is provided in 15-minute periods in excel format and is provided on DVD accompanying this report. This includes summary hourly turn flows provided by way of flow diagrams.

3.2 Survey Locations

3.2.1 The locations of the MCC surveys undertaken are set out in Table 3-1. Figure 3-1 shows a plan of the MCC surveys data locations. The surveys were undertaken on Thursday the 23rd of May 2019 being within the two-week period within which the new ATC surveys were undertaken.

3.2.2 The main purpose of the MCC data is to inform the matrix estimation process as part of the matrix development process.

Table 3-1: MCC locations

MCC Reference	Location
1	A24 Dorking Road / A281 / A24
2	A264 / A264 Crawley Road / B2195
3	A264 / Sullivan Drive / A2220 Horsham Road / A264
4	A281 / Old Guildford Road / A24 NB On-slip / A24 SB Off-slip / A281 Guildford Road / A24 SB On-Slip / A24 NB Off-slip
5	A272 / A24 Worthing Road
6	A24 NB On-slip / A24 SB Off-slip / A24 SB On-slip / East Wolves Farm / A24 NB Off-slip
7	A283 Storrington Road / A24 A283 The Pike / A24
8	Stane Street / A29

MCC Reference	Location
9	A272 / A29 / A272 West Street / A29
10	Albian Way / A281
11	A264 / A23 / M23 NB On-slip / M23 SB Off-slip / B2114 Brighton Road / A23 SB On-slip / A23 NB Off-slip
12	A272 Cowfold Road / London Road / A23 NB On-slip / A23 NB Off-slip
13	A272 Cowfold Road / A23 SB Off-slip / Crossways / A272 / A23 On-slip
14	A283 / A2037
15	Hop Oast Roundabout
16	A29 Stane Street/High Street
17	A281 / A29 Roman Road / A281 Guildford Road
18	Ingfield Manor School / A29 / A264 Horsham Road / A29
19	B2133 / A29
20	A283 Station Road / A29 London Road / A283 Lower Street / A29 Roman Road
21	A283 High Street / B2139 School Hill / A283 Manley's Hill
22	A283 Washington Road / Water Lane / Chanctonbury Ring Road
23	Clays Hill / A283 / Roman Road / The Street / A283 / Maudlin Lane
24	A272 Station Road / A281 Brook Hill / A272 / A281
25	A281 High Street / A2037 High Street / A281 Brighton Road

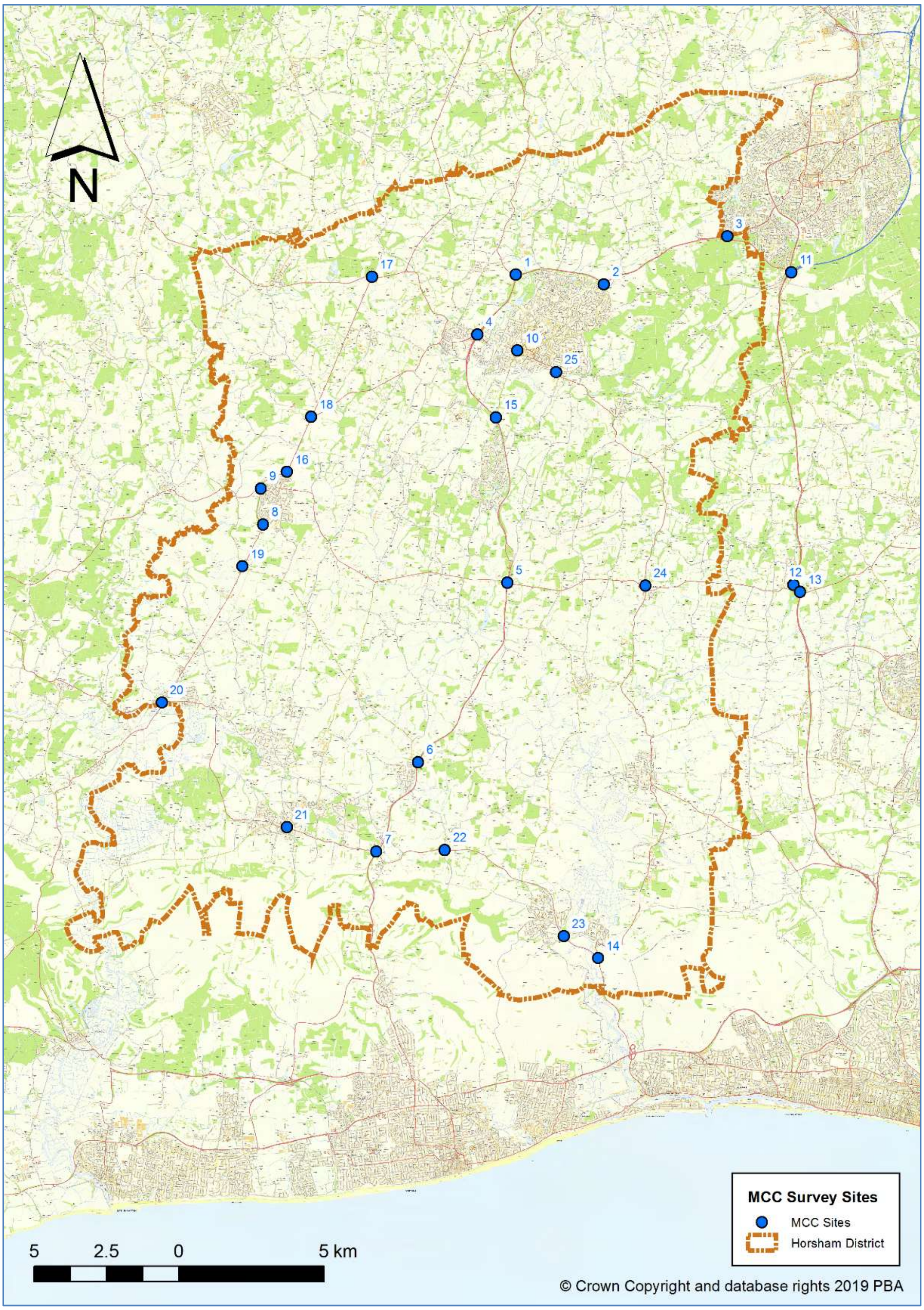


Figure 3-1: MCC Surveys data Locations

4 Journey Time Data

4.1 Overview

4.1.1 Journey time data for model update is sourced from Trafficmaster data via the Department for Transport (DfT) covering the period May 2019.

4.1.2 Trafficmaster journey time data has been sourced from in vehicle GPS tracking data and is broken into links with 15-minute segment. Trafficmaster data is made up of a mixture of vehicles.

4.1.3 Journey time data is used to validate the model through comparison of modelled journey times and observed data from this data source, on a number of routes within the model area. In total, 12 routes have been defined for the HTM which cover the main links within the study area.

4.1.4 Journey time routes for validation have been defined and the relevant journey time data for the AM peak hour (08:00 to 09:00), IP average hour (10:00 to 16:00) and PM peak hour (17:00 to 18:00) extracted from the full data for each of these routes. The data used was for the neutral weekdays Tuesday to Thursday.

4.1.5 The journey time routes are described in Table 4-1 and can be seen in Figure 4-1. The raw Traffic Master data is included in the DVD accompanying this report.

Table 4-1: Journey Time Routes

Route Number	Description	From	To
1	A272 – Petworth to Billingshurst	A283	A29
2	A281 - Cowfold to Horsham/A24	A272	A24
3	A264	M23 J11	A24
4	A272 – Billingshurst to Bolney	A272 High Street	A23
5	A24 North Horsham	A272	Horsham A264
6	A24 South Horsham	Findon	A272
7	A29 Madhurst to Billingshurst	A284	West Street
8	A281 – Alfold Crossways to Horsham	Alfold By-Pass	A24
9	Harwood Road, Horsham	A264	Albion Way
10	A283 – Petworth to Botolphs	Golden Square	A2037
11	A29 Billingshurst to Ockley	A272	B2126
12	A24 – Capel to Horsham	Wolves Hill	A264

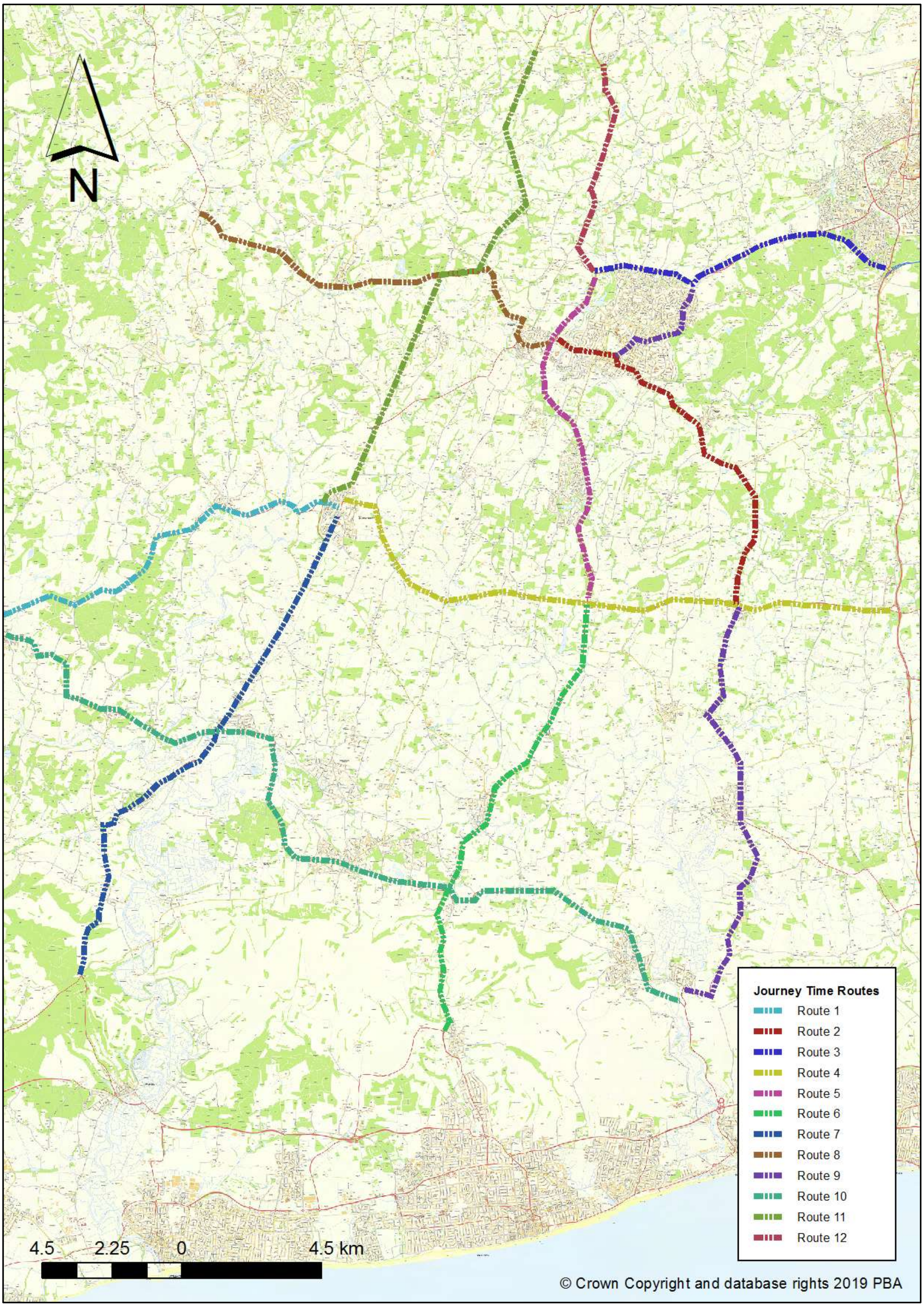


Figure 4-1: Journey Time Routes

4.1.6 Table 6–2 summarises the results for the AM peak hour for each of the eight routes by direction. This data will be used in validating the journey times for HTM AM peak model. The average speed for each route was determined from the route distance and average journey time and is given both in kilometres per hour and miles per hour. Speeds are easier to understand than journey times and give context as to the operation of the road network.

Table 4-2: Summary of AM Peak hour Traffic

Route	Direction	Length (km)	Observed Avg. JT (mm:ss)	Lower 15% JT (mm:ss)	Upper 15% JT (mm:ss)	Avg. Speed (kph)	Avg. Speed (mph)
1	Eastbound	13.0	12:26	10:34	14:18	63	39
	Westbound	13.2	13:19	11:19	15:19	59	37
2	Northbound	12.7	17:29	14:52	20:07	44	27
	Southbound	12.7	16:39	14:09	19:09	46	28
3	Eastbound	10.5	11:05	09:25	12:45	57	35
	Westbound	10.4	09:12	07:49	10:35	68	42
4	Eastbound	20.0	22:04	18:46	25:23	54	34
	Westbound	19.8	20:55	17:46	24:03	57	35
5	Southbound	11.8	08:41	07:23	09:59	82	51
	Northbound	11.7	11:27	09:44	13:10	61	38
6	Northbound	15.6	11:22	09:40	13:04	82	51
	Southbound	15.6	10:33	08:58	12:08	89	55
7	Northbound	17.9	21:38	18:24	24:53	50	31
	Southbound	18.0	21:08	17:58	24:18	51	32
8	Eastbound	8.6	07:56	06:45	09:07	65	41
	Westbound	8.6	08:04	06:51	09:17	64	40
9	Northbound	4.0	11:50	10:04	13:37	20	12
	Southbound	3.9	07:20	06:14	08:26	32	20
10	Eastbound	29.9	35:01	29:46	40:16	51	32
	Westbound	29.6	35:46	30:24	41:08	50	31
11	Northbound	17.2	15:03	12:48	17:19	69	43
	Southbound	17.5	15:57	13:33	18:20	66	41
12	Northbound	7.6	07:12	06:07	08:17	64	40
	Southbound	7.6	07:20	06:14	08:25	63	39

4.1.7 Similar analysis for the PM average hour is shown in Table 6-3. The rural nature of the network is also evident with speeds generally below or just above 30 mph. The more strategic Route A24 shows speeds of the order of just above 50 mph in both directions. The results appear reasonable.

Table 4-3: Summary of PM average hour Traffic

Route	Direction	Length (km)	Observed Avg. JT (mm:ss)	Lower 15% JT (mm:ss)	Upper 15% JT (mm:ss)	Avg. Speed (kph)	Avg. Speed (mph)
1	Eastbound	13.0	11:38	09:54	13:23	67	42
	Westbound	13.2	12:27	10:35	14:19	64	40
2	Northbound	12.7	15:43	13:22	18:05	48	30
	Southbound	12.7	17:36	14:57	20:14	43	27
3	Eastbound	10.5	08:56	07:35	10:16	71	44
	Westbound	10.4	10:52	09:14	12:30	57	36
4	Eastbound	20.0	19:38	16:41	22:35	61	38
	Westbound	19.8	19:41	16:44	22:38	60	38
5	Southbound	11.8	08:26	07:10	09:41	84	52
	Northbound	11.7	09:21	07:57	10:45	75	47
6	Northbound	15.6	10:15	08:43	11:48	91	57
	Southbound	15.6	10:48	09:10	12:25	87	54
7	Northbound	17.9	18:27	15:41	21:13	58	36
	Southbound	18.0	18:28	15:41	21:14	59	36
8	Eastbound	8.6	07:33	06:25	08:41	69	43
	Westbound	8.6	07:20	06:14	08:26	70	44
9	Northbound	4.0	10:44	09:07	12:20	22	14
	Southbound	3.9	06:29	05:31	07:28	36	23
10	Eastbound	29.9	30:14	25:42	34:46	59	37
	Westbound	29.6	31:31	26:47	36:14	56	35
11	Northbound	17.2	14:27	12:17	16:37	71	44
	Southbound	17.5	14:36	12:24	16:47	72	45
12	Northbound	7.6	07:07	06:03	08:11	64	40
	Southbound	7.6	07:43	06:34	08:53	59	37

4.1.8 Journey time data has been cross-referenced against online travel data sources, in all cases and in all route direction the GPS data falls within other online journey time estimations, this further verifies the accuracy of the collected GPS journey times used within the model calibration.

5 Mobile Network Data

5.1 Overview

5.1.1 Mobile network data (MND) will be used as the main source of data to develop the initial origin and destination matrices for the HTM. This was in preference to undertaking Roadside Interview Surveys (RSI) which has been the traditional method of developing matrices for transport models in the UK for a long time. However, RSI surveys are disruptive to travellers and sample rates can be low leading to less robust matrices. Use of mobile phone data is increasingly being seen as a credible alternative although understanding and experience of using this data for matrix development is still limited.

5.1.2 The origin and destination matrices provide information on where trips start and where they end within the model. The model is split up into a number of zones, which are smaller within the study area and get larger as you get further away. Zone size is also reflective of density of development and the highway network, i.e. zones will be smaller in built up areas and larger in rural areas. The zones can also reflect areas of different land use where these are differentiated i.e. representing residential or employment areas.

5.1.3 The existing Telefonica data was collected within West Sussex County in 2015 for the purposes of providing travel demand data for a transport model for the Crawley district. It has been deemed after further analysis that the existing mobile phone data is still appropriate to use as a base for prior demand in 2015. The analysis undertaken looked to verify if there have been any substantial changes to travel movement or land use within the Horsham District, and it was deemed that since 2015 general travel demand movements caused by changes to transport infrastructure and land use changes have remained similar.

5.1.4 The MND was provided by Telefonica (O2 in the UK) for the West Sussex region for six neutral weeks in April and May 2015. The data as provided was separated into different modes (road, rail and HGV). The data had also been split by purpose into the following categories:

- Non-Home Based (NHB) Trips
- Outbound Home-Based Work (OB_HBW) Trips
- Outbound Home Based Other (OB_HBO) Trips
- Inbound Home-Based Work (IB_HBW) Trips
- Inbound Home Based Other (IB_HBO) Trips

5.1.5 As MND is known to possess some biases the data for the Horsham model area has been validated to ensure that it is fit for purpose. The purpose of the validation is to demonstrate that the mobile phone data is consistent with known sources of trip making data such as census journey to work data and National Travel Survey (NTS) data.

5.2 Initial Validation by Telefonica

5.2.1 Telefonica undertook some initial validation of the mobile phone data prior to releasing it to WSCC. The initial validation is reported in '*West Sussex OD from Mobile Phone Data, Project Report*', 11/03/2016, attached as Appendix A.

5.2.2 The validation looked at the dataset for the whole of West Sussex whereas the HTM is only interested in a subset of this region itself. It has been deemed that the validation checks conducted by Telefonica are sufficient to validate the Horsham Transport Model, where Horsham can be deemed to be representative of similar trip generation rates of the wider West Sussex area.

5.2.3 Telefonica compared a number of different factors in the data to the Census data (including journey to work data) and the National Travel Survey (NTS). These tests included:

- Comparison of home-based origins with zone home population
- Comparison of work-based origins with zone work population
- Analysis of trip purpose split
- Comparison of trips starting and ending per zone (trip symmetry)
- Trip length distribution against the census journey to work and NTS data
- Comparison of travel start time with NTS data
- Comparison of rail mode share with the census journey to work data

5.2.4 Telefonica used these tests to verify that there is a strong correlation and good fit between the mobile data and the census/NTS data. Thereby ensuring that the data is suitable for use. Their results also highlighted some known biases in the mobile data such as the under-representation in short trips up to about 5 miles.

6 Traffic Signal Data

6.1 Overview

6.1.1 Traffic Signal data has been provided by WSCC in the form of signal timing data sheets. The data provided will provide initial signal timing data for input into the SATURN model by time period. This included information on phase movements by stage including pedestrian phases, minimum and maximum green times, inter-green times and cycle times.

6.1.2 Figure 6-1 shows the locations of signal junctions in Horsham. Signal data was provided for these junctions. As stated, the timings will be used as initial inputs to the model and it is expected that these might change during the calibration and/or validation process. The data is included in electronic format.

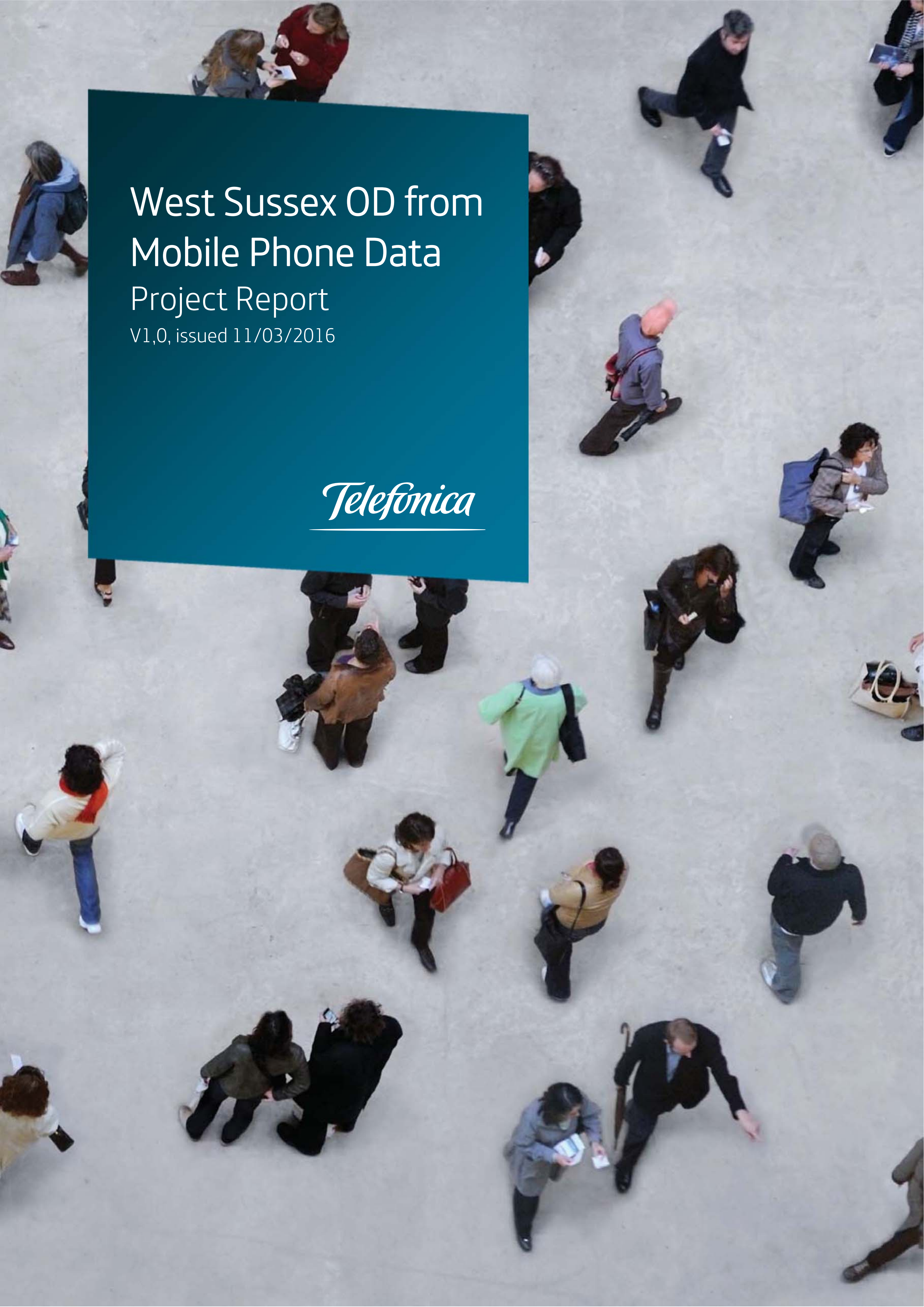
Figure 6-1 – Horsham Traffic Signal Sites

Appendix A Mobile Phone Data Report

West Sussex OD from Mobile Phone Data

Project Report

V1,0, issued 11/03/2016



Telefonica

Confidentiality Statement

Registered office: 260 Bath Road, Slough, Berkshire, United Kingdom, SL1 4DX

Contents

1. Introduction & Project Scope

1.1. Introduction

Telefonica are a mobile network operator (O2 in the UK), providing telephony services to over 22 million UK customers in both the public and private sectors. In order to provide this service Telefonica operate a network which provides continuous nationwide coverage to each customer phone (device). In order to provide efficient service to each phone, the network and phone are in frequent communication. Intimate understanding of these networks allows Telefonica to build contextual understanding of the movement of devices in space and time in the real world, with each phone creating events at specific points in time and space which can be chained into 'breadcrumbs', demonstrating whether each phone is moving or stationary at any point in time.

The result of Telefonica's processing creates a huge and valuable dataset which describes the movement and flow of O2 users across the UK. Devices are tracked anonymously and can be associated with attributes derived from the user's contract (age, gender, contract type and billing address) or their observed behaviour (affluence, lifestyle, home and work location and other points of interest). In aggregate, therefore, mobile phone data provides an effective insight in the movement patterns of the UK population.

Given the nature of mobile phone data, it is able to effectively represent movements on a macro basis across larger areas. The technology is generally better at identifying longer trips and those where the user dwells at their destination for a longer period of time. For this reason, the data should not be used in isolation but should be combined with other data sources prior to application.

Customer privacy is of upmost importance to Telefonica. All events processed are by-products of the core telephony network, and the process does not affect any user's handset. The records are anonymised prior to being stored in the analysis platform, so all analysis of behaviour is done in a completely anonymous separate environment. Outputs from the analysis are aggregated such that no individual level data will be given to clients.

1.2. Scope

Telefonica were requested by WSP to prepare origin-destination matrices for travel in the West Sussex area. Trips were included if they penetrated a cordon shown in the figure below.

Image showing the extent of the model cordon

The trips were allocated to a start and end zone based on a zone system provided by WSP. This consisted mostly of Lower Super Output Areas (LSOAs) inside the cordon, middle super output areas (MSOAs) around the edge of the cordon and districts or regions outside the cordon, as illustrated in the figure below. There were a total of 501 zones.

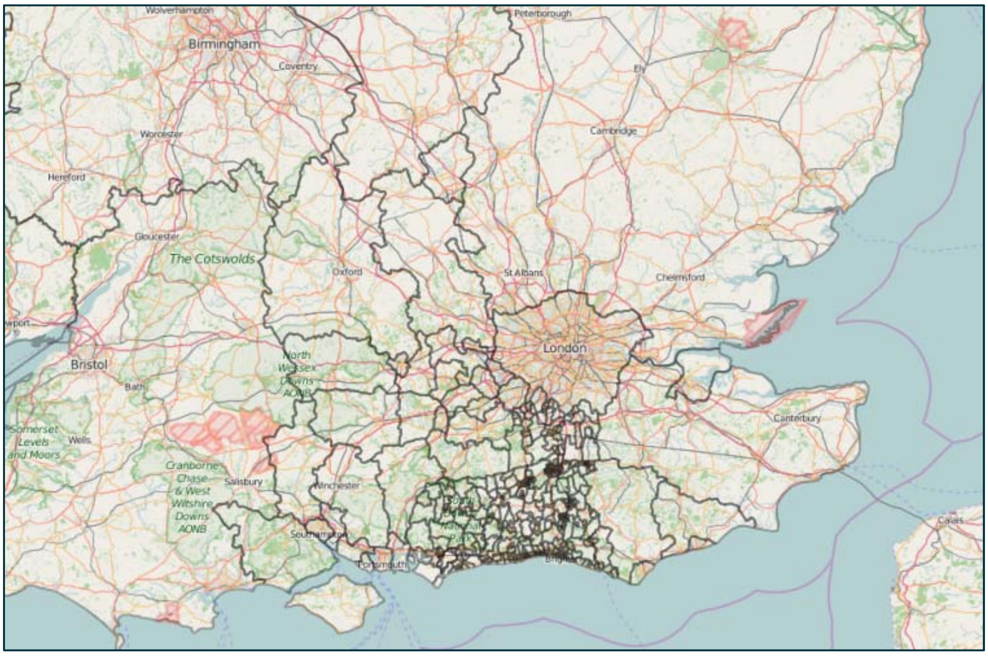


Image showing the zones used to identify trip stand end points

Trips were segmented as follows:

- By mode into rail, HGV and other road, with walk/cycle trips removed
- By purpose into outbound home based work (OB_HBW), inbound home based work (IB_HBW), outbound home based other (OB_HBO), inbound home based other (IB_HBO) and non-home based (NHB)
- By time period into:
  Early off peak (00:00-07:00)
  AM peak (07:00-10:00)
  Interpeak (10:00-16:00)
  PM peak (16:00-19:00)
  Late off peak (19:00-00:00)

  Further outputs were also created represented travel in the AM peak hour (08:00-09:00) and PM peak hour (17:00-18:00).

  Trips were segmented based on the time they entered the model cordon, or started inside the model cordon

1.3. Study Period

Trips were sampled using 23 neutral days in April and May 2015. These were defined as Mondays-Thursdays, excluding bank holidays and school holidays. The following 23 days were included in the final dataset

13/04/15	Monday
14/04/15	Tuesday
15/04/15	Wednesday
16/04/15	Thursday
20/04/15	Monday
21/04/15	Tuesday
22/04/15	Wednesday
23/04/15	Thursday
27/04/15	Monday
28/04/15	Tuesday
29/04/15	Wednesday
30/04/15	Thursday
05/05/15	Tuesday
06/05/15	Wednesday
07/05/15	Thursday
11/05/15	Monday
12/05/15	Tuesday
13/05/15	Wednesday
14/05/15	Thursday
18/05/15	Monday
19/05/15	Tuesday
20/05/15	Wednesday
21/05/15	Thursday

2. Mobile Phone Technology

2.1. Overview of the Cellular Network

A cellular network or mobile network is a wireless network distributed over land areas called cells, each served by at least one fixed-location transceiver which is known as a cell site or base station. In a cellular network, each cell uses a different set of frequencies from neighbouring cells, to avoid interference and provide guaranteed bandwidth within each cell. When joined together, these cells provide radio coverage over a wide geographic area. This enables a large number of portable transceivers (e.g., mobile phones, pagers, etc.) to communicate with each other and with fixed transceivers and telephones anywhere in the network, via base stations, even if some of the transceivers are moving through more than one cell during transmission.

Adjacent cells form groups of cells. The names of these groups depend on the generation of the cells, but for simplicity in this document we will use the 2G grouping which is LAC. LACs overlap and vary in size, depending on the area. Grouping cells into LACs is essential for the collection of event data.

2.2. Event Data

O2 mobiles phones generate "events" as they communicate with the national cell network. Telefonica collects these events on an anonymised basis for the purpose of analysis. Each event is linked to a persistent, yet anonymised user ID. Along with each event, Telefonica also stores a timestamp as well as the cell ID of the cell that recorded the event. In this manner, the spatial and temporal distribution of events can be analysed to determine users' movement patterns. Events can be classified into active and passive events. It is the combination of both of these types of events that allows Telefonica to build a representative, stable dataset. Without the inclusion of passive events, the sample would be biased toward more active users and individual user profiles would be biased towards locations where they made calls.

Active Events

- **Connection events** occur when a user turns their phone on or off, loses or regains connection
- **Call events** occur when a user makes or receives a phone call, or moves between cells when on a call
- **Text events** occur when a user makes or receives a text message

Passive Events

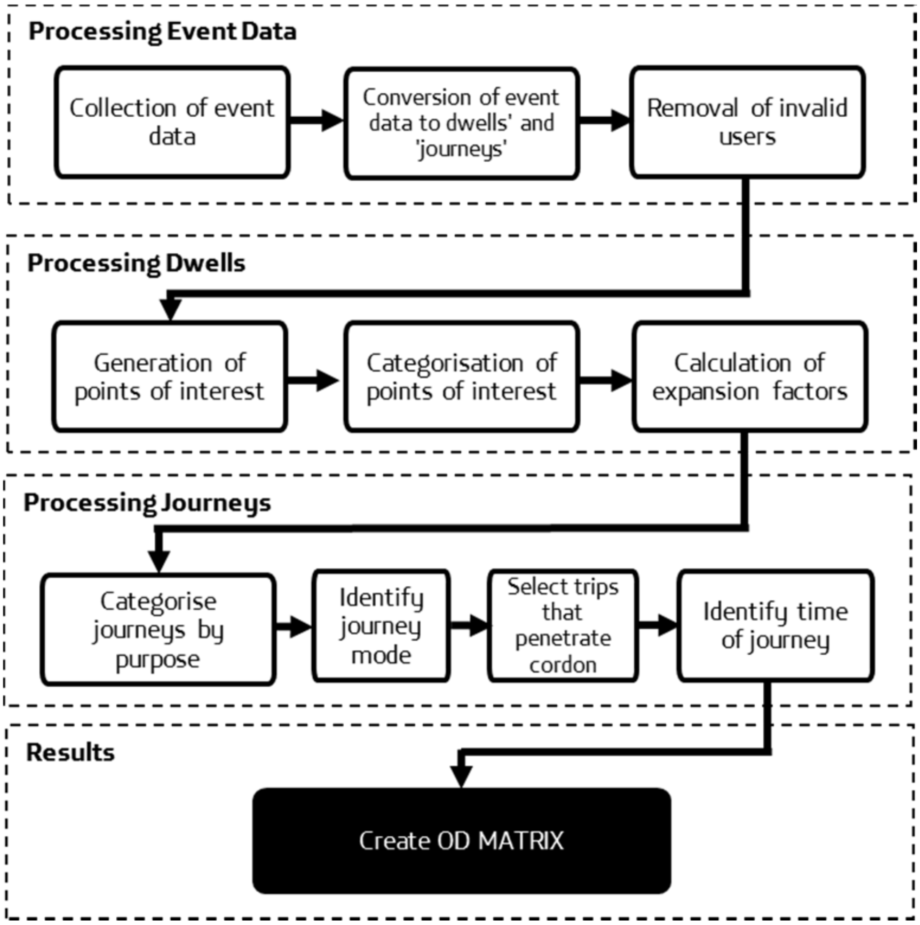
- **Movement events** occur when a user moves from one LAC to another. LACs consist of a number of nearby cells in the same band – so users also create passive events when they transition between 2G/3G/4G coverage. These events ensure that journeys that cover more than one LAC will be recorded by the analysis process. The collection of these events is vital for accurately observing trips and allocating them to the correct mode.

- **Time-based events** occur whenever a user does not create any event for a sustained period of 3 hours. These events ensure that longer dwells are identified even if they are in the same LAC as the previous dwell.

3. Methodology

3.1. Process Overview

The diagram below summarises the process used to create the OD matrix deliverables. Each step is described in more detail in this chapter.



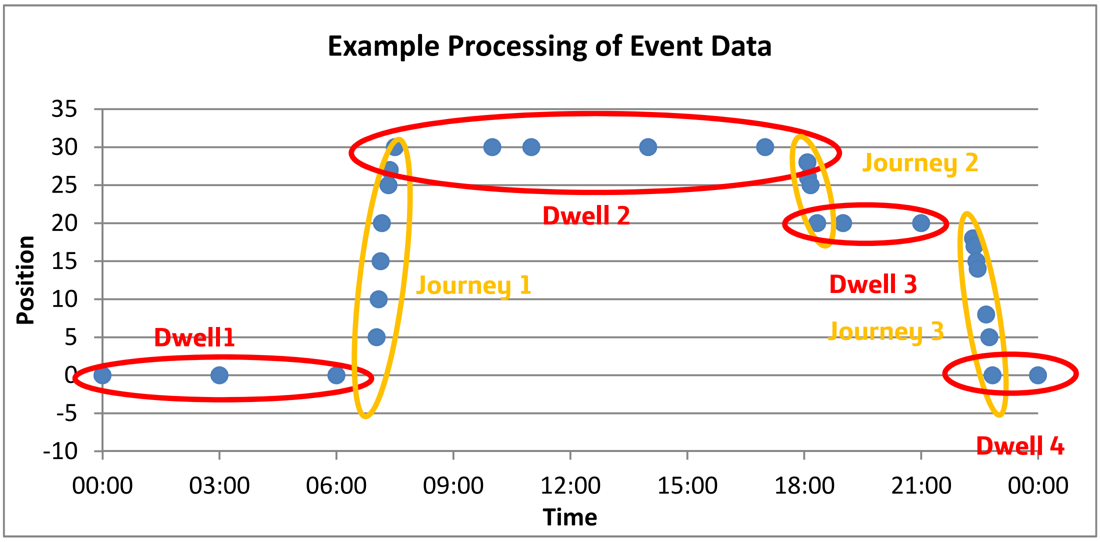
Process diagram of existing methodology

3.2. Collection of event data

As described in section two, mobile phones regularly generate events. These are collected ('probed') by Telefonica for network management and billing purposes. To enable analysis of travel data, the events are also stored in a database for further analysis. Telefonica has access to data relating to the whole of the UK for the last two years, but for the purposes of this project data was analysed for 23 specific weekdays as listed above. It should be noted that although only these 23 days were used to create the OD matrix, data from other days was analysed for some specific purposes, e.g. identifying valid users and home locations.

3.3. Conversion of Event Data to Dwells and Journeys

Telefonica converts the raw event data into 'dwells' (or settles) and 'journeys'. The algorithm that is used for this conversion process takes into account the geographic proximity of events, the propensity for phones to 'flicker' between cells without changing location and the timing of each event. In general, dwells are created whenever a user is assumed to be stationary in one distinct place for at least 30 minutes. The period between two dwells is classified as a journey. The cells of the events which have been combined to make up each settle and each journey are stored as 'via points', which can be interrogated to understand the route of each journey or the location of each settle. Note that journeys represent person trips, and not vehicle trips, due to the nature of mobile phone data.



Processing event data: dwells and journeys

3.4. Removal of invalid users

Events are created by all O2 users, corresponding to about 30% of the UK population or c22m connections. Each user is allocated an anonymised user ID, to ensure their records cannot be traced back to a particular user. The anonymous ID is set up to ensure that it is consistent even if a user changes their SIM card or phone, but if a user leaves O2 their records will cease. To prevent these users from affecting the sample, a filtering process is run to identify a sample of 'stable users' who are consistently present throughout the study period. To be included in the sample users must:

- Be seen at least once in the 10 days before the study period;
- Be seen at least once in the 10 days after the study period;
- Be seen on at least 10 days during the study period

Also at this stage a filter is applied to ensure that only mobile devices are included in the sample – machine to machine (M2M) devices, tablets and GPS units are excluded, since they are less likely to be carried by users at all times. Large business contracts are also removed from the sample to reduce the risk of double counting users who carry two phones.

Users who change phones: the anonymous and president user id is based on a user's telephone number, so they will persist in the data if they change phone or SIM card, providing they keep their number and stay on O2.

3.5. Generation of Points of Interest

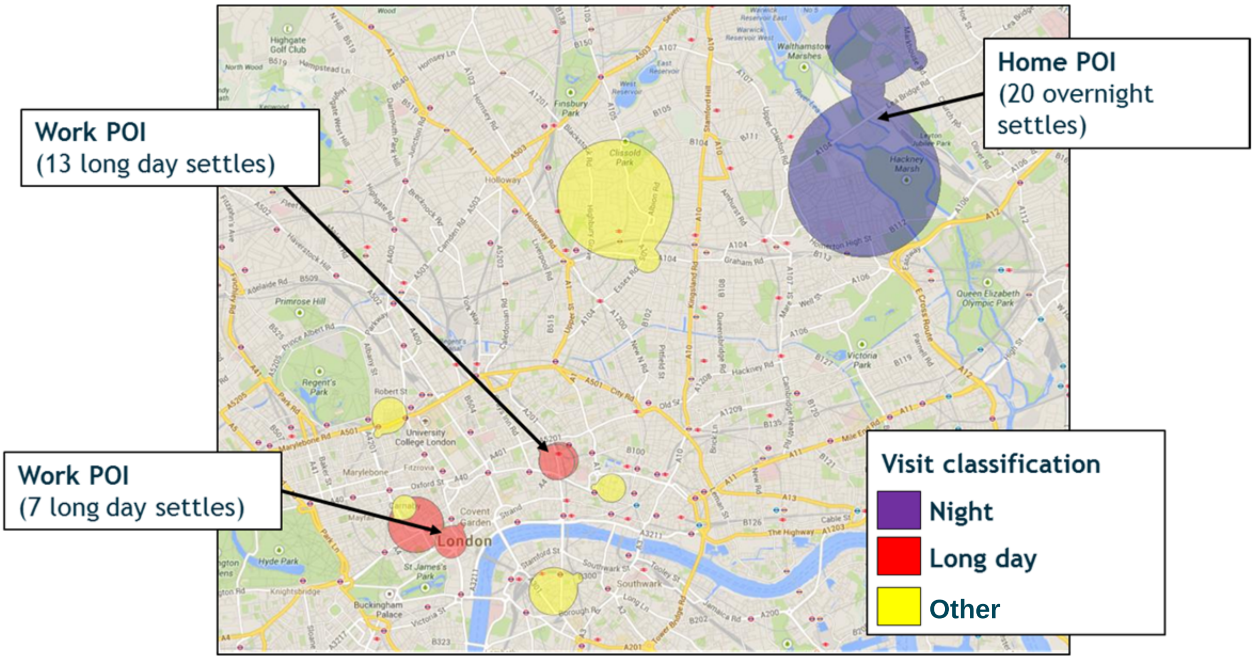
Where a user has multiple dwells which overlap each other, these will be associated with a particular Point of Interest (POI). By analysing all of the dwells associated with a particular POI the position of the POI can be identified with a higher degree of accuracy, because more information will be provided. All of the events associated with a POI will be analysed and the relevant cell geographies will be compared to the zone system supplied by WSP, so that each POI is associated with a zone. Every time a user visits a cell associated with one of their POIs, this will be recorded as a trip to the associated zone.

Zones smaller than MSOA: mobile phone data has been found to be most appropriate for zones equivalent in size to MSOA or larger. Where smaller zones are used, it is often necessary to distribute POIs between the small zones based on a combination of the zone shape, the shape of the cells where the user generated events during their visits to the POI and the population data associated with the zone – this process has been carried out for the LSOA zones provided in this project.

3.6. Categorisation of Points of Interest

Categorisation of POIs is based on the temporal patterns of a user's dwells at each POI throughout the study period. POIs where users spend a large amount of time overnight are classed as home POIs. All users must have a home POI. POIs where users spend long periods of time during the working day are defined as work POIs. All other POIs are defined as 'other' POIs.

Note that this process may create multiple home and work POIs for each user. When this occurs, the most used home POI is flagged as a primary home location (for expansion purposes) but the secondary home/work locations will be retained for the purpose of deriving trip purpose.



Example POI classification

The POI schematic used is designed to detect regular daytime commuters. As such, it may cause small errors relating to users who behaviour in unusual ways:

Working from home: users who work from home will have a home POI, but no work POI.

No fixed place of work: users who have a moving place of work (e.g. plumbers) will not usually have a work POI, unless they spend most of the study period working at the same site. Their trips to work will usually be included in the home-based-other matrix.

Shift workers: users who work unusual hours, e.g. night shifts, will not usually have a work POI- their trips will also be included in the home-based-other matrix.

3.7. Calculation of expansion factors

O2's market share varies across different geographical regions in the UK. To account for this, users are allocated an expansion factor, which relates to how representative they are of the UK population. The process for calculating the expansion is as follows:

- For every valid user (as described in section 3.4), identify their primary home POI. This is defined as the POI at which they spend the most nights during the study period
- Count the number of primary home POIs in each MSOA region of the UK. Intermediate zones are used in Scotland, LSOAs in Northern Ireland.
- For each MSOA, compare the number of primary home POIs with the total census population from 2011. Each MSOA will become associated with an expansion factor which is equivalent to the census population divided by the number of primary home POIs in that MSOA.

- Each user then inherits the expansion factor associated with the MSOA that their primary home POI is located in. This means that the sum of user weights for all the users in the UK will match the census population.
- A small uplift of +3% is applied to each user's weight to account for population growth between 2011 and 2015 (based on ONS statistics)
- Any trips made by each user, regardless of origin or destination, will be scaled up according to the weight of the user

3.8. Categorising journeys by purpose

Journeys are assigned a purpose based on the categorisation of their start and end POI:

Origin POI	Destination POI	Purpose
Home	Work	Outbound Home-Based-Work (OB_HBW)
Work	Home	Inbound Home-Based-Work (IB_HBW)
Home	Home	Outbound Home-Based-Other (OB_HBO)
Home	Other	Outbound Home-Based-Other (OB_HBO)
Other	Home	Inbound Home-Based-Other (IB_HBO)
Work	Work	Non-Home-Based (NHB)
Work	Other	Non-Home-Based (NHB)
Other	Work	Non-Home-Based (NHB)
Other	Other	Non-Home-Based (NHB)

Trip Purpose Categories

Education trips: Telefonica are not able to specifically identify trips made by users aged under 18, so we do not segment education trips in our outputs. However, some pay as you go users are included in the sample, and many of these will be associated with people in education. Education trips made by these users will usually be included in the home based work trips, because they are trips between home and a place where the user regularly spends long periods of time during the time. Many education trips will not be included in the data, either because they are too short (see validation, trip length distributions for details) or because they are made by users who do not carry phones. It is recommended that alternative datasets are used to split out and supplement education trips from the matrices.

Note that education escort trips, where observed, will usually be included in home-based-other trips.

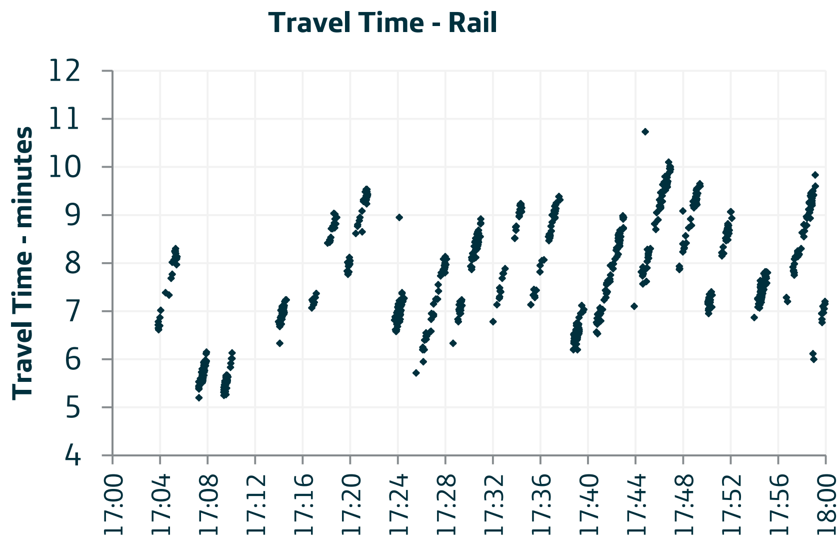
3.9. Identify journey mode

At this stage, the route and characteristics of each journey will be analysed to allocate the journey to one of the following modes:

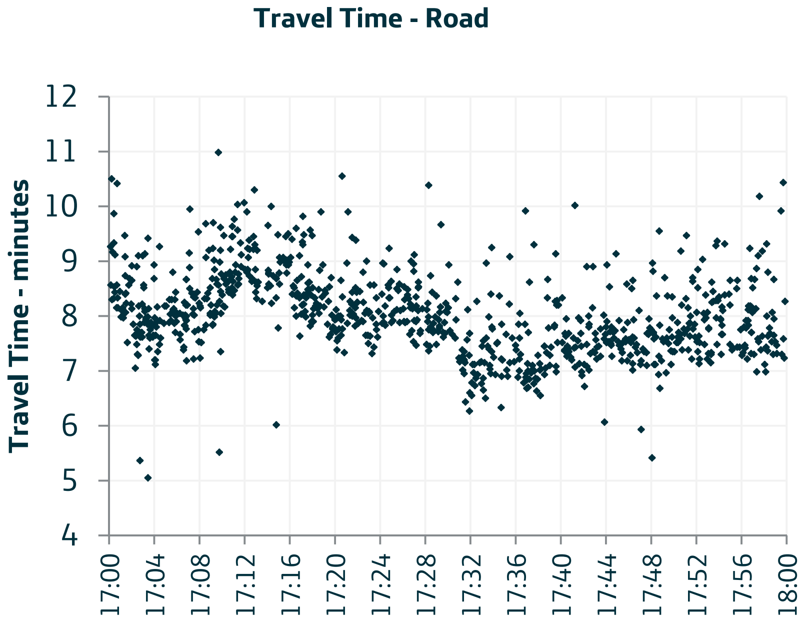
- Air – journeys with a high speed between two airports. These trips are removed from the final matrix. Note that users will create 'dwells' at the start and end airport of their journey, so access trips to airports will be included in the dataset
- Rail – journeys which follow the rail network and which exhibit 'clustering' (see description below) will be allocated to rail
- HGV – trips made by users who travel regularly at lower speeds (<62mph) on the strategic road network will be allocated to the HGV matrix.
- Walk/cycle – short and slow trips will be allocated to walk/cycle and removed from the matrix. Most walk and some cycle trips will be too short to detect using mobile data. Some longer, faster cycle trips may be indistinguishable from road trips and so will be included in the road matrix.
- Road – any remaining trips will be allocated to the road matrix – note that this includes coach, bus and LGV trips as well as car trips.

Park and ride: it is usual practice when processing mobile data to identify trips based on their true origin and destination, defined as points where the user has dwelled for more than 30 minutes. This means that park and ride trips, where the user drives to a station and then travels by train, will not usually be split into two distinct trips but will instead by represented as a single trip in the rail matrix. It is recommended that adjustments are made to the rail matrix if park and ride trips are thought to be a significant component of travel in any part of the model.

Clustering: the distinction between road and rail journeys is accomplished by identifying cell pairs that show characteristic travel time patterns for either of the two modes. When a train crosses the boundary of a LAC, the phone of every O2 customer on board will generate a passive event. These events will occur within very quick succession (depending on the length and speed of the train, as well as the device type and the current state of the mobile network), which will result in the clearly identifiable clustering patterns. When these patterns become apparent for a specific pair of cells, these cell pairs can be classified as "rail". However, when a pattern with no clusters is observed this is indicative of a road pair. On a road, a continuous flow of cars is usually observed and events (i.e., movements from one LAC to another) also occur continually. An algorithm examines the clustering patterns of all the journeys in the system to identify rail and road journeys.



Characteristic clustering pattern of a rail cell pair



The lack of any identifiable clusters indicates a road cell pair

3.10. Select trips that penetrate cordon

Once every journey is associated with a route, these routes are compared against the cordon provided by WSP and only those trips which penetrate this cordon are included in the matrix.

3.11. Identify time of journey

Journeys are allocated to a time band based on their start time (for trips starting within the cordon) or their time of entering the cordon. The time of entering the cordon is derived from the event closest to the cordon, unless there are no events within 3km of the cordon. In this case, the time of entry is inferred from the time and position of the two events closest to the cordon on either side.

3.12. Create OD matrix split by mode

Once all journeys have been allocated a time, purpose and mode it is straightforward to create the OD matrix outputs. Trips are allocated to a time period, mode and purpose and included in the relevant part of the matrix. Note that trips in the AM peak hour (8-9am) and PM peak hour (5-6pm) will be included twice- once the in peak hour matrix and once in the peak period matrix.

Stochastic rounding: to preserve personal data, Telefonica does not provide outputs relating to the movement of individuals. In the context of an origin-destination matrix, this is achieved by creating an average result representing multiple days of observations, and by rounding results to integer values.

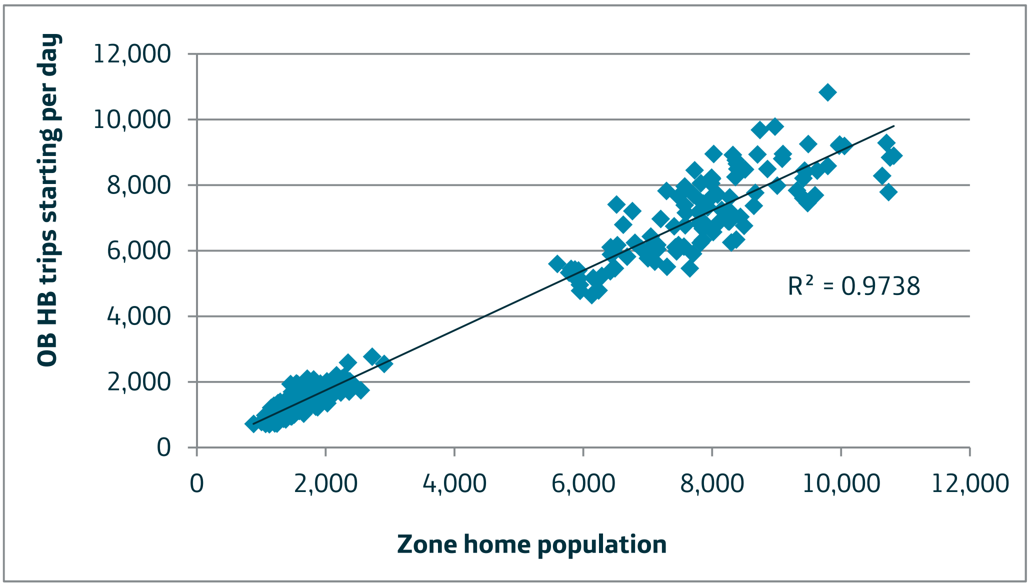
Applying standard rounding methods would cause errors in the outputs because they would cause many cells in the matrix to be rounded to zero, reducing the volume of trips in the data. To avoid this, stochastic rounding is used whereby the probability of a value being rounded up or down depends on fractional part – so a value of 0.1 has a 90% probability of being rounded down to zero and a 10% probability of being rounded up to one. This method of rounding preserve the overall volumes of the matrix (and the size of any part of the matrix large enough for the rounding interval to be negligible) while also preventing the disclosure of individual level data.

4. Validation

Prior to releasing the data Telefonica carries out a range of validation checks to ensure internal consistency and check against relevant alternative data sources. Because trips are only included if they penetrate the model cordon, checks are usually limited to zones which are fully within the cordon.

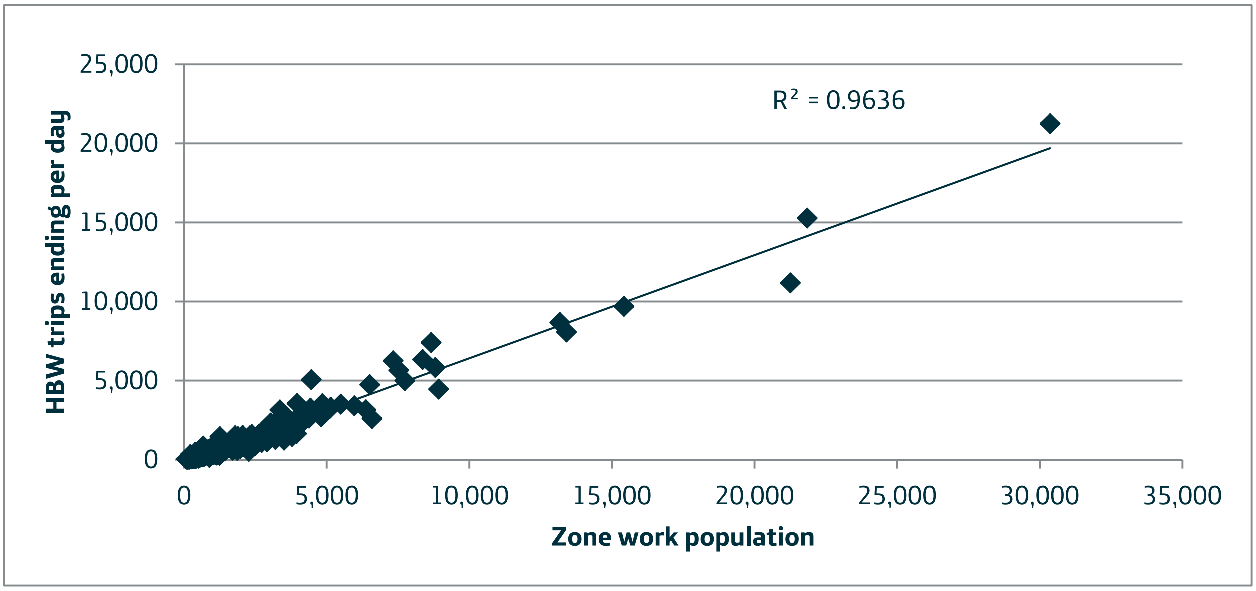
4.1. Comparison of home based origins with zone home population

The scatter graph below shows the number of outbound home based trips starting in each zone within the cordon on an average day in the study period against that zone's home population, based on the 2011 census. As is to be expected, zones with a higher population tend to have more home trip ends per day, with an R2 of 0.97 indicating a strong correlation between the two variables. A similar result was found when checking the number of inbound home based trips ending in each zone against the home population.



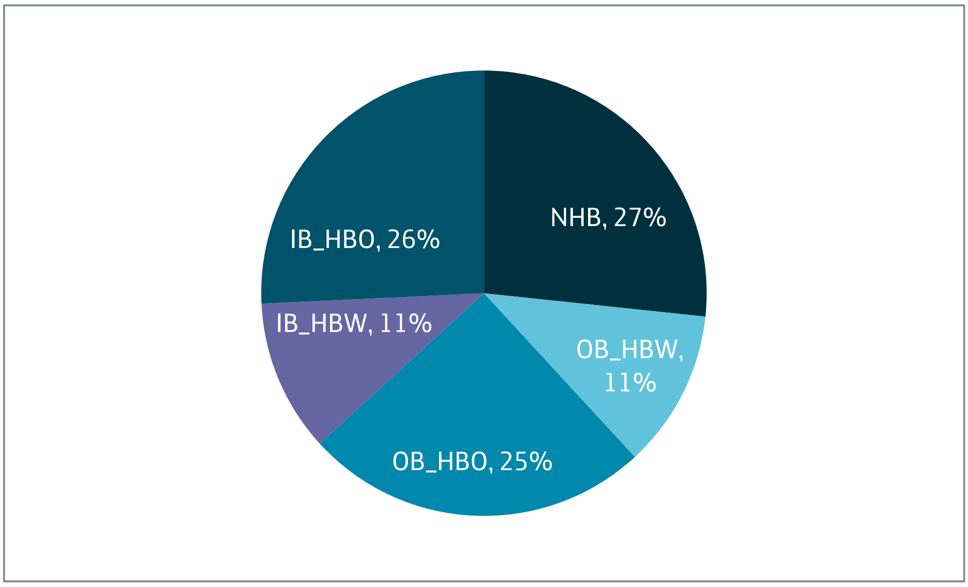
4.2. Comparison of work based origins with zone work population

The following graph shows the number of outbound home based work trips arriving at each zone within the cordon, during a typical day in the study period, against the work population of each zone (based on census workplace statistics). Again a strong correlation is found with an R2 of 0.96. A similar result was found when comparing the number of inbound home based work trips starting in each zone with the zone work population.



4.3. Trip purpose split

The following graph shows the trips starting or ending in zones fully within the cordon, split by purpose:



A direct comparison of this data with secondary data is difficult because there are no publicly available datasets showing trips split by purpose for particular areas of the country. A high level comparison with the DfT's TEMPRO data suggests that the number of non-home-based trips in the mobile phone data is high. This could be due to limitations in the TEMPRO data, although there two other reasons why NHB trips may be over-represented in the mobile phone data:

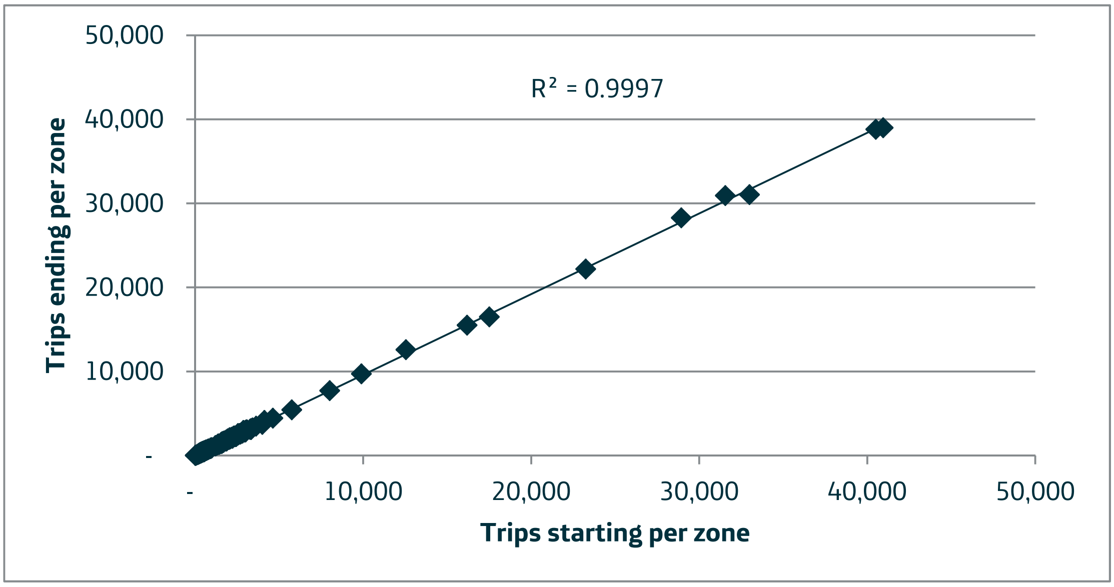
- Home based trips may tend to be shorter than NHB trips, and so more likely to be missing in the mobile phone data
- Education trips are more likely to be home based, and are also more likely to be missing from the mobile phone data (particularly for users without phones).

It is recommended that secondary data sets are used to complement the mobile phone data to correct for the biases described above and increase the proportion of home based trips.

.

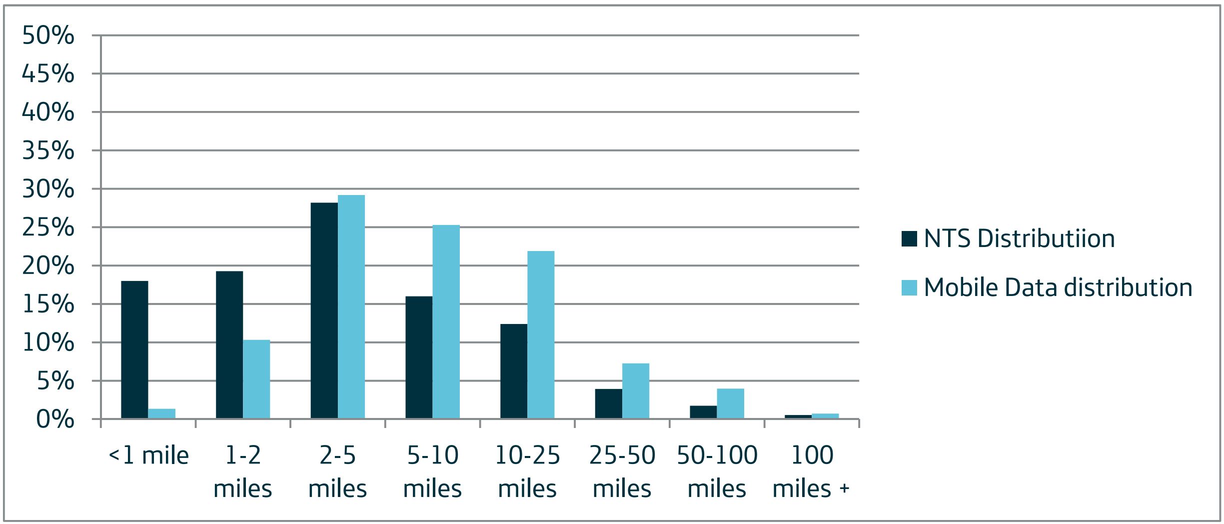
4.4. Trip symmetry

The graph below shows a comparison of the number of trips starting (by all modes and purposes) with the number of trips ending in each zone. As expected a very strong correlation is found, with an R2 of close to 1. A similar result was found when analysing the symmetry of HBW trips only, HBO trips only, and when trips were segmented by mode.

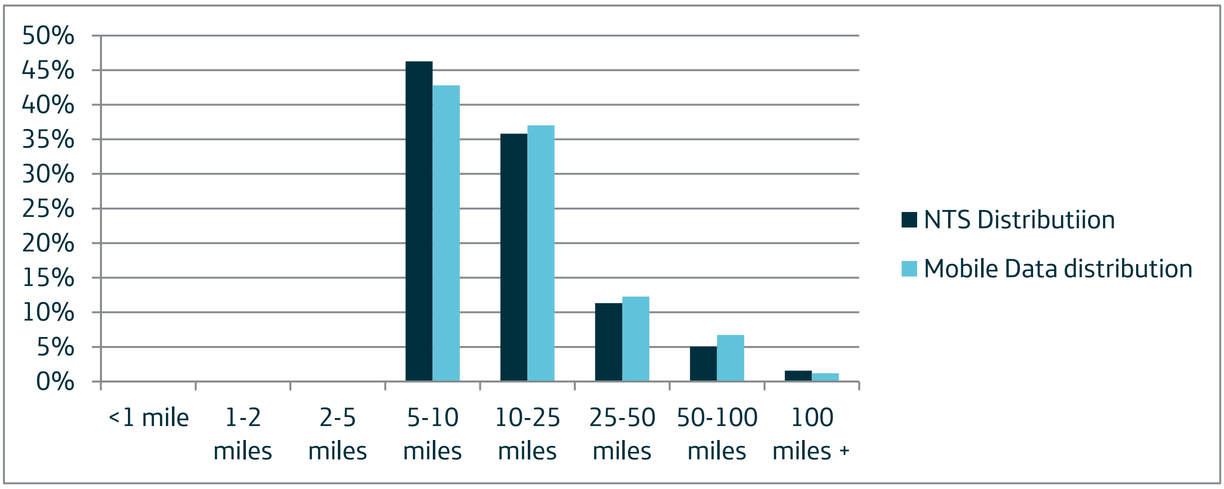


4.5. Trip length distribution compared to National Travel Survey

The graph below shows a comparison of the trip length distribution for trips starting or ending in the cordon (by all modes and purposes) with the trip length distribution reported in the National Travel Survey for South East England (NTS9911).



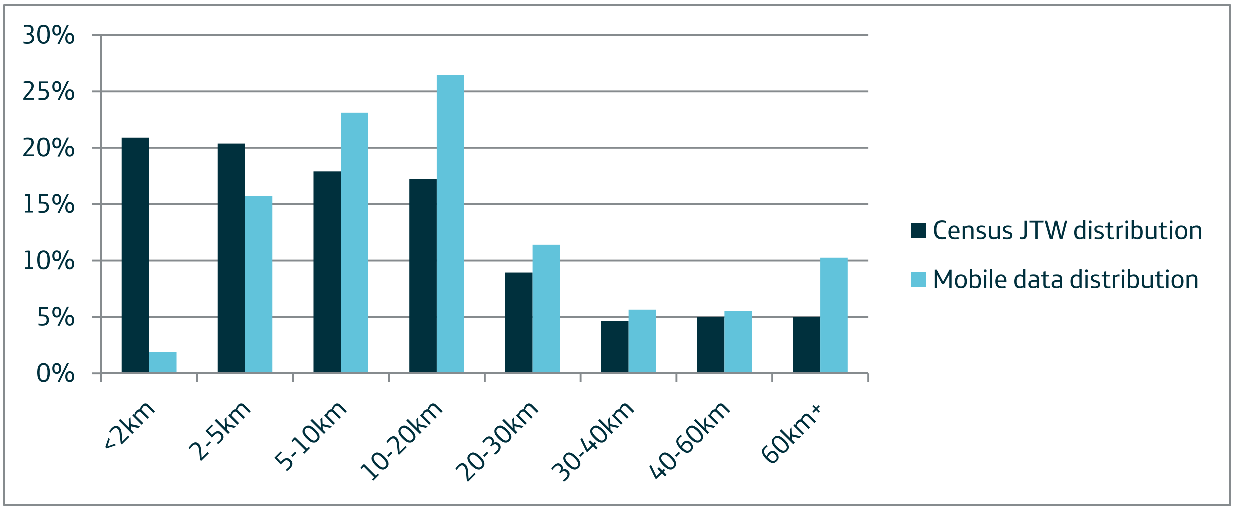
At first glance the match between the two datasets is poor, with the NTS containing more trips below two miles and the mobile data containing more trips above 5 miles. However, a better match is found when comparing only trips above five miles in length:



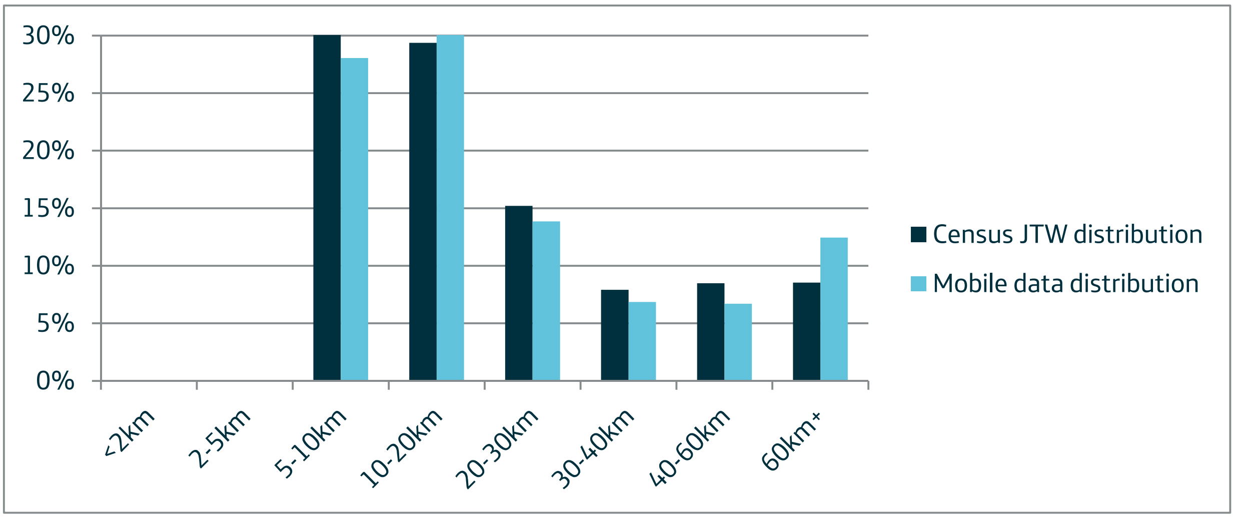
This result indicates that trips above five miles in length are well represented in the mobile phone data, while trips below five miles in length are partially represented in the mobile phone data. This is a well-known limitation of mobile phone data, since trips can only be represented if they move between cells, which in rural areas can be large. It is recommended that secondary datasets are used to correct for this bias in the mobile phone data.

4.6. Trip length distribution compared to census journey to work data

The graph below shows a comparison between the trip length distribution of the outbound home based work trips in the mobile phone data with the commute distances reported for West Sussex in the 2011 census:



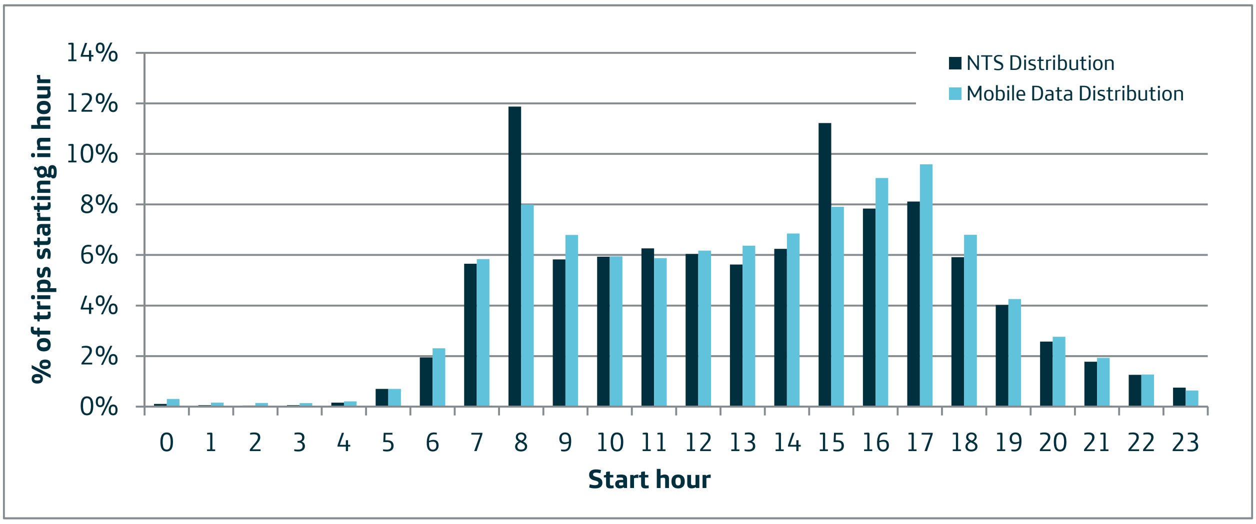
As for comparisons with the census, the match for this range is poor, but improves when only trips above 5km are included in the comparison:



This result indicates that trips above 5km are well represented in the data, suggesting that the point at which trips start to be under-represented in the mobile phone data is somewhere between 2 miles (3.2km) and 5km.

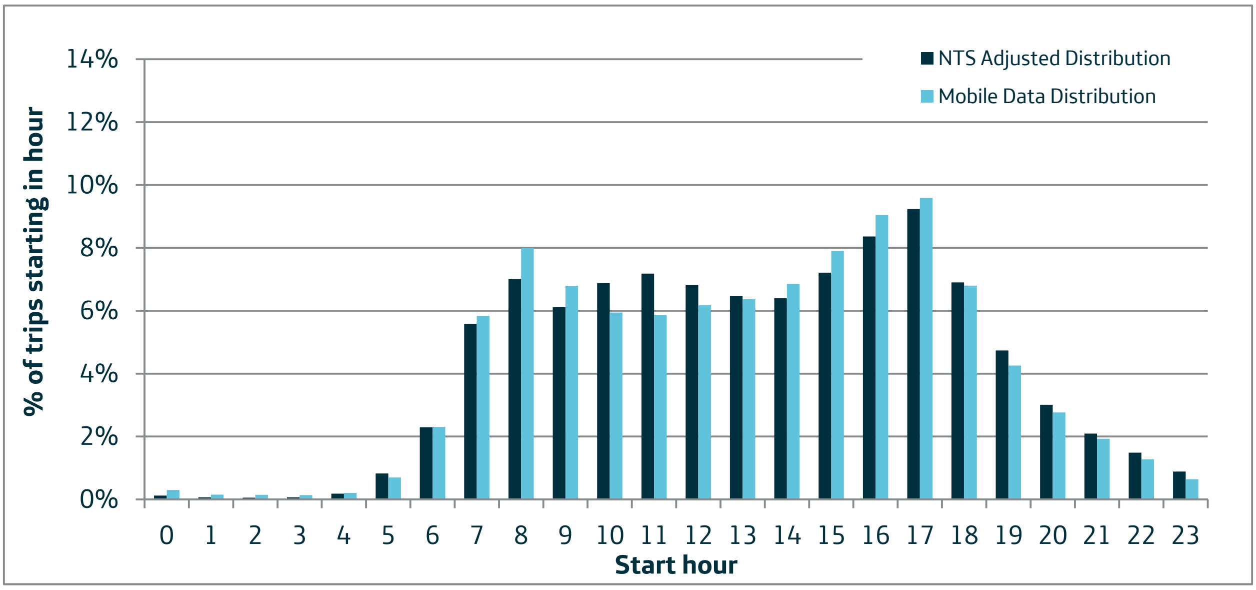
4.7. Comparison of trip start time with National Travel Survey

The graph below shows the percentage of trips starting in each hour in the study against the percentage indicated by the National Travel Survey (NTS). This is not a direct comparison since the NTS result includes trips across the UK, whereas the mobile data results only include trips that start in West Sussex.



This comparison shows a good match between the two datasets, except for trips starting between 8am and 9am and between 3pm and 4pm. Further analysis indicates that this difference could be attributed to education and education escort trips, suggesting that these trips are poorly represented in the mobile data. This may be due to the trips being particularly short, and due to some education trips being made by users without phones. It is recommended that secondary data be used to correct for this bias.

The graph below indicates how this comparison improves when education and education escort trips are removed from the NTS. The comparison is much stronger, especially considering the inconsistency between the areas covered by the surveys:



4.8. Comparison of mode with census journey to work data

Analysis of census journey to work data for the study area indicates that 12% of commute trips starting in the cordon were by rail (excluding walk/cycle trips). By comparison, the mobile phone data indicated 11% of OB HBW trips were rail trips.

Similar analysis was carried out for commuting trips ending in the study area. The 2011 census suggested 7.3% of commute trips ending in the cordon were by rail, while the mobile data suggested 5.5% of trips ending in the cordon were by rail.

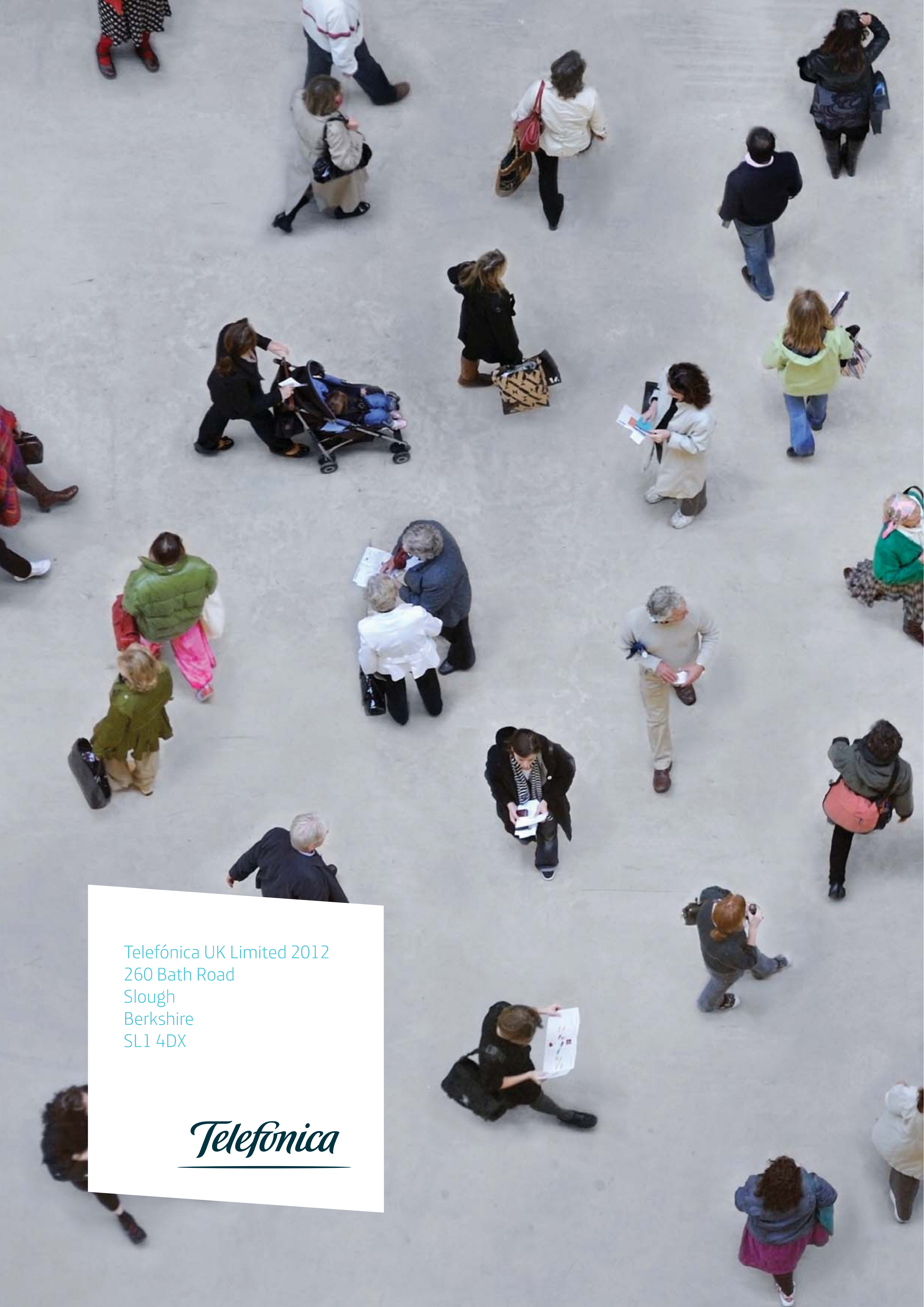
In both cases, the census data indicated a slightly higher proportion of trips by rail than found in the mobile data. This is likely to be due to a small proportion of shorter rail trips being attributed to road. The impact on the road matrix is likely to be small but the impact on the rail matrix may be more significant, and it is recommended that where local rail trips are important secondary data is used to complement the mobile phone data.

5. Summary

The methodology and validation sections have highlighted a number of biases, all of which are recognised limitation of mobile phone data. It is represented that secondary data sources are used to enhance them mobile phone data to correct for them:

- Comparisons with trip length distributions from NTS and the census indicate that trips below 5km are likely to be under-represented in the mobile phone data. However, this will depend on the cell resolution – in urban areas (e.g. Brighton) short distance trips are more likely to be represented, while in rural areas the threshold may be slightly higher. Previous experience suggests that trips between MSOAs are likely to be well-represented, but trips within MSOAs, or between LSOAs, may be under-represented.
- Comparisons with trip start time indicate that education and education escort trips are likely to be under-represented in the mobile phone data. This will partly be a natural consequence of the short-trip bias (since many education trips are short), but may also be due to some education trips being made by people who do not carry phones. Where education trips are included in the mobile phone data, they are likely to be counted as home-based-work trips.
- Comparisons with TEMPRO data indicate that home-based trips are under-represented in the mobile phone data. This is likely to be a combination of the above two biases, since most education and education escort trips are home based, and home based trips may also tend to be shorter than non-home based trips.
- Park and ride trips are represented in the mobile data as rail trips from the initial origin to the final destination. Where park and ride is thought to be a key component of demand, it is recommended that secondary datasets are used to split the rail matrix into rail legs and road access legs.
- Longer distance, faster cycle trips are likely to be included in the mobile data as road trips. These trips are unlikely to have a significant impact on the matrices since this represents a small volume of commutes compared to road modes.
- Some short distance rail trips may be included in the road matrix, due to difficulties in distinguishing mode for short distance trips.

Subject to these biases, the checks have shown that the mobile data provided is internally consistent and compares well to the secondary data it is compared above. The checks described are limited to publicly available datasets and are not intended to be exhaustive, further comparisons with appropriate local datasets are advised prior to applying the matrices to a transport model.



Telefónica UK Limited 2012
260 Bath Road
Slough
Berkshire
SL1 4DX

Telefónica