TECHNICAL NOTE - DRAFT

SESTRAN REGIONAL MODEL 2012 DEVELOPMENT

TOMTOM DATA PROCESSING

IDENTIFICATION TABLE	
Client	Transport Scotland
Project	SEStran Regional Model 2012 Development
Title	TomTom Data Processing
Type	Technical Note - Draft
Date	16/06/2015
Reference number	102936 12
Number of pages	8

TABLE OF CONTENTS

1. INTRODUCTION

1.1.1 SYSTRA was commissioned by Transport Scotland to develop the SEStran Regional Model 2012 (SRM12) and apply the SRM12 to inform the SESplan cross-boundary transport and land use appraisal study. The appraisal primarily considers the implications of proposed land use options within the SESplan area on the transport network, identifying strategic locations where operational issues may arise and testing potential mitigation options.

1.1.2 To assist the calibration and validation of the SRM12 road model, Transport Scotland purchased TomTom road journey time information collated from satellite navigation systems. 'Area-based' TomTom data was supplied through Streetwise in November 2014, with supplementary data sets received in December 2014 and February 2015.

1.1.3 An 'area-based' data set provides a speed / time record for each section of road (or GIS link) where data is recorded. This describes the average speed / time of all recorded vehicles using a given link, irrespective of the specific turning movement at intersections (whereas 'route-based' data records the average speed / time of vehicles using a specific collection of roads and turning movements).

1.1.4 This note documents the content and coverage of the data sets provided and the approach undertaken to use the data sets within the SRM12 development.

2. DATA PURPOSE AND COVERAGE

2.1 Data Purpose

2.1.1 The TomTom data sets were provided as 'networks' and were received in GIS (shapefile) format. These networks were interrogated and used to develop observed journey time information for 14 routes, against which the SRM12 modelled data could be compared. This facilitated the improvement of the model in terms of matching modelled journey times to observed data, and comparing SRM12 outputs with model validation criteria.

2.1.2 The TomTom data sets represented several weekday time periods during 2012, covering the periods between February and June and between September and November. The TomTom morning peak, inter peak and evening peak time periods were used to compare with the SRM12 AM, Inter and PM peak hourly time periods respectively.

2.1.3 As TomTom data provides a database of average speed records along specific roads, these data sets could also be applied to develop speed-flow curve relationships. Developing a new set of specific speed flow relationships was not part of the SRM12 development programme, but from some initial interrogation and comparisons, this is an area that is worthy of further investigation, particularly for key routes.

2.1.4 The 14 routes where TomTom data was required and were used to provide observed journey time comparisons are illustrated in Figure 1.

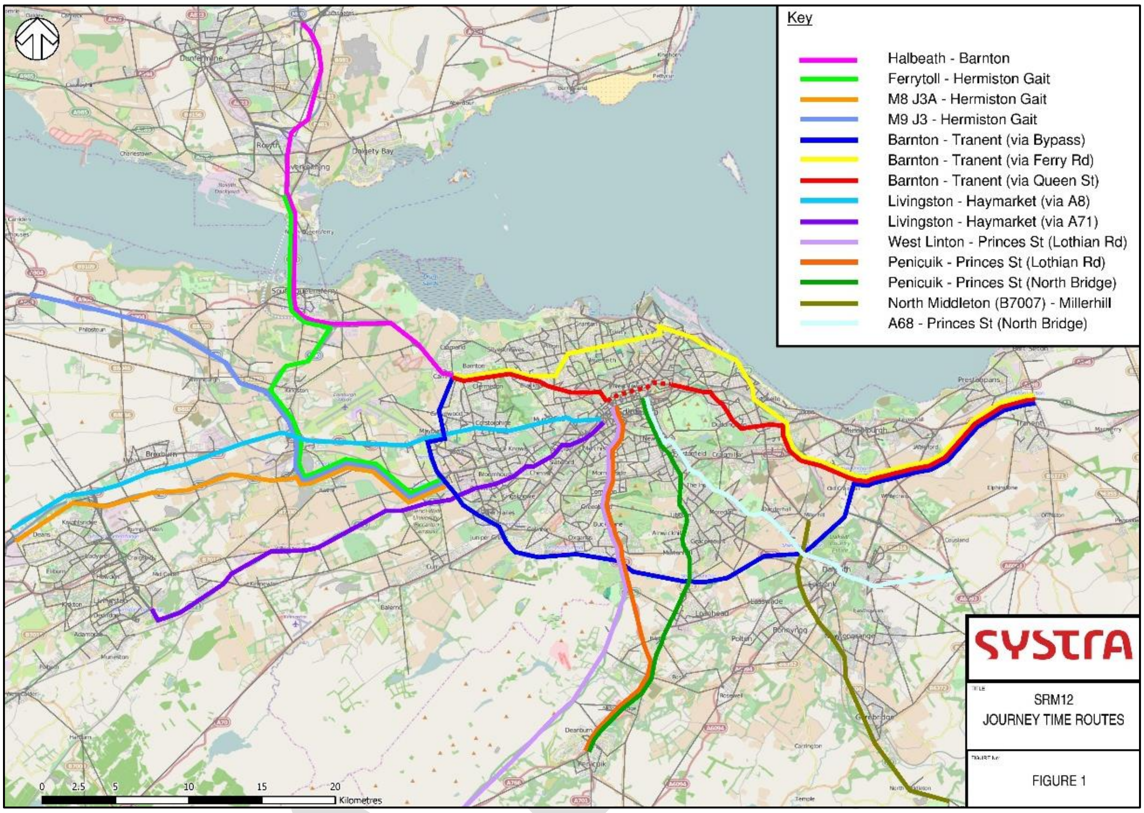

Figure 1. SRM12 Journey Time Routes

2.2 Data Coverage

2.2.1 The network data file contained approximately 12,000 links (one-way) and covered all the major routes in and around Edinburgh. Table 1 lists the main routes covered, and figure 2 illustrates the full geographical coverage of the TomTom routes / links.

2.2.2 With the wider database of roads available from the 'are-based' (rather than route-based) TomTom data set, additional journey time routes could be developed if required. This would require further processing of data for any additional route.

Table 1. TomTom Data Road Coverage

ROAD	FROM	TO
M8	Hermiston Gait	Junction 3A
M9	Junction 1	Junction 3
M90/A90	Edinburgh	Junction 3
A720	Gogar	Old Craighall
A8/A89	Edinburgh	M8 J3A
A71	Edinburgh	West Calder
A70	Edinburgh	B7008
A702	Edinburgh	A766

ROAD	FROM	TO
A703	A702	A701
A701	Edinburgh	Penicuik
A7	Edinburgh	North Middleton
A68	A720	B6458
A1	Edinburgh	East of Tranent

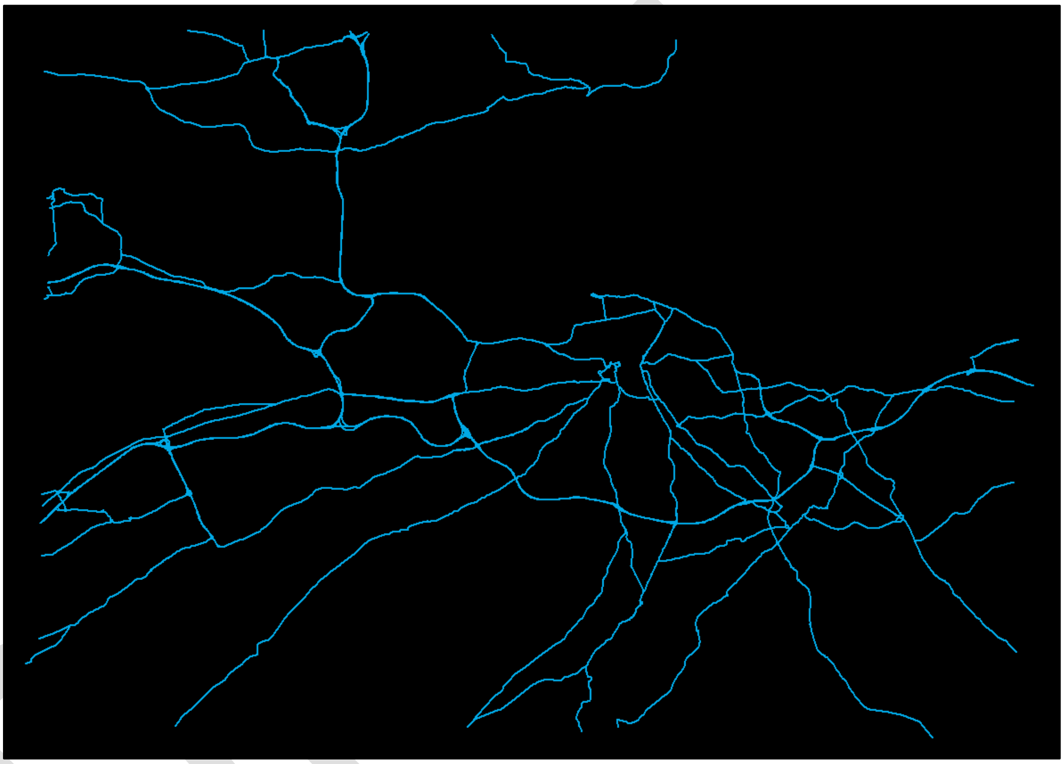

Figure 2. TomTom Data Geographical Coverage

3. DATA UTILISATION & PROCESSING

3.1 Format

3.1.1 The data set comprised two elements: the network in shapefile format and the underlying data associated with each link in database (.dbf) format. There were seven database files received, each relating to a specific time period, including:

- 0700 – 0800
- 0800 – 0900
- 0900 – 1000
- 1000 – 1600
- 1600 – 1700
- 1700 – 1800
- 1800 – 1900

3.1.2 Example data contained within each temporal database file is displayed in Figure 3.

OID	CS7_Id *	CS7_avg_tt	CS7_med_tt	CS7_ratio	CS7_avg_sp	CS7_med_sp	CS7_sd_sp
0	1	2.23	0.97	2.36	34.57	37	10.91
1	2	1.28	1.26	0.8	43.75	44	4.19
2	3	11.23	7.31	0.32	102.67	105	19.01
3	4	1.17	0.95	1.33	37.01	38	9.47
4	5	1.29	1.26	1.01	43.8	44	6.25
5	6	5.42	4.88	0.89	24.52	19	13.66
6	7	0.8	0.5	1.01	26.55	28	14.72
7	8	0.5	0.44	0.79	33.3	34	8.6
8	9	4.23	3.87	0.16	57.41	59	10.83
9	10	4.74	4.39	0.84	51.42	52	9.82
10	11	4.87	4.36	0.24	52.63	53	10.15
11	12	8.11	7.54	1	46.43	47	7.28
12	13	12.77	9.64	0.66	27.62	24	17.02
13	14	10.22	8.25	0.99	41.02	43	10.39
14	15	3.04	2.36	0.21	56.92	58	10.79
15	16	4.22	3.55	1.12	33.39	35	10.1
16	17	4.02	2.97	1	46.02	46	12.41

Figure 3. Example TomTom Data Received Format

3.1.3 Each temporal database file contained approximately 30 columns of data. The majority of these related to link speed, with percentile speeds given in increments of five centiles from 5% to 95%.

3.2 Data Choice and Interpretation

3.2.1 For each link, both time and speed data were provided. Arithmetic averages and medians were provided for travel time; harmonic averages and medians were provided for speed.

3.2.2 Note that when speed data was cross-referenced with link distance and converted to time, it was found that the calculated time differed from the reported time within the table. This is due to the difference in the method that the arithmetic mean and harmonic mean are calculated. It was decided that the dedicated travel time records for each link would be used rather than those inferred from the reported speeds and distance.

3.2.3 A number of detailed spot checks were undertaken to compare differences in the average and median travel time records with local network knowledge. This suggested that the median data may underestimate some travel times (possibly as the median calculation may tend to reduce the impact of the most severe periods of congestion / longer travel times). Therefore, the average travel time data was considered to be more robust, and hence was used as the main observed data comparison.

3.2.4 Note that (unlike floating car surveys) TomTom data sets are not provided with a commentary of general network conditions or incidents that may affect the capture of data. With the months of data provided, there is potential that network incidents (such as accidents or roadworks) are captured within the average speed / time records. If so, these would tend to overestimate comparison times with model data – as the SRM12 is designed to represent average conditions across the network.

3.2.5 Considering these potential issues, in general, it was felt that modelled times should fall somewhere between median and average travel times, potentially closer to the average TomTom data values.

3.3 Data Processing

3.3.1 Within the original data format, the data table associated with the network file did not provide an indication of the directionality of two-way links. This means that, spatially, two-way links simply consisted of two one-way links positioned on top of each other.

3.3.2 To resolve this issue, ArcGIS was used to determine which end was the 'start' and 'end' points of each link. To establish the direction of a link, Arc's *Calculate Geometry* function was used to generate X and Y co-ordinates for the start and end points, with new data columns created to hold this information.

3.3.3 If the X co-ordinate of a link end point was a higher value than the X co-ordinate of the link start point, then this meant the link was an eastbound link and vice-versa. Similarly, if the Y co-ordinate of a link end point was a higher value than the Y co-ordinate of the link start point, then this meant the link was a northbound link and vice-versa. Furthermore, the relative magnitudes of the X difference and Y difference indicated whether the link was predominantly running in an east-west direction or a north-south direction. This method established the directionality of all links in the network.

3.3.4 The next task was to join the data table associated with the network with the temporal database data table. This was undertaken within ArcGIS using the link identification number that was provided within both data sets.

3.3.5 The network data table does not contain a structured geographic order of links, i.e. links geographically adjacent to each other did not necessarily have adjacent identification numbers. Indeed, in many cases, adjacent links could have identification numbers that were different by a value of several thousand.

3.3.6 To extract the required links, sections of the route were selected using Arc's *Selection by Lasso* function, which allowed a freeform shape to be drawn encompassing all the links required. This required several small sections to be selected independently to form the complete route, with care taken to avoid selecting unwanted links. It should be noted at this stage that, even though only links in one direction were required for each route, it was not possible to prevent selection for links for both directions, as they were positioned identically. This would be rectified in the next part of the process.

3.3.7 Once all links for a given route were selected, relevant data was input to a spreadsheet. Using the co-ordinate information attached to each (one-way) link, the direction was added in a new column and the data sorted by direction. A visual check of the route was undertaken in Arc to check for any links whose direction may be inconsistent with that of the route as a whole, e.g. at a roundabout where an eastbound route may have a short link that faces westbound.

3.3.8 The final step was to ensure only links for the relevant direction were included within the final combined total route travel time calculation. The majority of routes were divided into sub-sections to make them more manageable and allow for more detailed comparisons of differences in observed and modelled travel times.

3.3.9 A final review was undertaken during the comparison of modelled and observed travel times. This check related to the distance of the modelled and observed sections. If, for example, an error had been made in terms of extraction of data for each route (or section of route), then it could likely be identified on inspection of the relative lengths of the modelled and observed data.

4. SUMMARY

4.1.1 The TomTom journey time data was an extremely useful source of modelled travel time information to inform the calibration and validation of the SEStran Regional Model (SRM12).

4.1.2 It provided a comprehensive data set for comparison purposes and the 'area-based' nature would provide further data for additional validation if required.

4.1.3 In general, the data sets appeared to reflect local knowledge of road network conditions in and around Edinburgh. Although for some specific routes and data sets, journey time differences between time periods appeared less representative.

4.1.4 One issue relates to the level of data processing required to develop the original data into a more useable format, and select and combine the required data, particularly within urban areas. It is understood that since processing these data sets, further refinements to the data supply procedure have been made, (such as providing road names) which should help with further data analysis.

4.1.5 The time required to process the data also has to be set against the provision of a sizeable data set, which could be used to provide a much fuller comparison of journey times across the modelled area, and potentially refine speed flow curve relationships.

4.1.6 A further issue is the ability to interpret the frequency of changes in journey time in specific time periods on the various days data is recorded. Potential provision of a fuller data set with individual averages for each recorded day would aid data interrogation. This may also enable some indication of journey time reliability or network incident detection, providing a wider understanding of the data sets.

APPROVAL

Version	Name		Position	Date	Modifications
1.1	Author	Laurence Bacon	Senior Consultant	26/05/2015	
	Checked by	Jeff Davidson	Associate Director	16/06/2015	
	Approved by	Jeff Davidson	Associate Director	16/06/2015	
2	Author			DD/MM/YY	
	Checked by			DD/MM/YY	
	Approved by			DD/MM/YY	