

# **A large-scale multi-modal implementation of MATSim for the Nelson Mandela Bay Metropole**

by

**JAMES VAN VELDEN**  
29299838

Submitted in partial fulfilment of the requirements  
for the degree of

**BACHELORS OF INDUSTRIAL ENGINEERING**

in the

**FACULTY OF ENGINEERING, BUILT ENVIRONMENT  
AND INFORMATION TECHNOLOGY**

**UNIVERSITY OF PRETORIA**

October 2012

## Executive Summary

The modelling of private vehicle and freight traffic using agent-based simulation (ABS) has been very successful in South Africa, but transit, or public passenger transport, has comparatively lagged behind. This research project implemented a large-scale multi-modal Multi-Agent Transport Simulation Toolkit (MATSim) model, which includes transit, for the Nelson Mandela Bay Metropole (NMBM). To achieve this, the MATSim simulation environment is investigated and available General Transit Feed Specification (GTFS) data for the region is matched to the network and transformed into a MATSim transit schedule. A comparison of planned and simulated transit leg times indicate that for short legs, simulated travel times are higher than planned, while for longer legs, simulated travel times are shorter than planned. These values are expected to improve as transit data accuracy and model calibration are refined. In addition to the comparison of planned and simulated values, several analyses provide detailed results for and insight into the NMBM transit system. This project, as the first large-scale agent-based transit simulation in South Africa, represents the first step in a process towards providing models which can accurately capture the intricacies of our public transport environment. Several research projects are currently under way which aim to model minibus taxis - only once this integral part of our system is included will agent-based transit simulation be considered as truly representative, but even now it affords the opportunity to gain unparalleled insight into current *and* future transit systems.

# Table of Contents

<b>List of Abbreviations and Acronyms</b>	<b>4</b>
<b>1 Introduction</b>	<b>5</b>
1.1 Transport in South Africa . . . . .	5
1.2 Transport Modelling in South Africa . . . . .	6
1.3 Research Design and Methodology . . . . .	6
1.4 Outline of Document . . . . .	7
<b>2 Literature Review</b>	<b>8</b>
2.1 The Multi-Agent Transport Simulation Toolkit . . . . .	8
2.1.1 The MATSim Process . . . . .	8
2.1.2 The MATSim Building Blocks . . . . .	9
2.1.3 The MATSim Input Components . . . . .	10
2.2 The Transit Schedule . . . . .	11
2.2.1 Transit Data Models . . . . .	12
2.2.2 The General Transit Feed Specification . . . . .	12
2.3 The Transit Router . . . . .	13
<b>3 Data Preparation</b>	<b>17</b>
3.1 The Matching Process . . . . .	18
3.1.1 Flow of Data and Initial Matching . . . . .	18
3.1.2 Solution Modification . . . . .	20
<b>4 Analysis and Discussion of Simulation Results</b>	<b>24</b>
4.1 Transit Router Configuration and Performance . . . . .	24
4.2 Simulation Outcome . . . . .	26
4.2.1 Comparison of Planned and Simulated Route Leg Times . . . . .	26
4.3 Detailed Results for a Single Route . . . . .	31
4.4 Transit Accessibility . . . . .	37
<b>5 Conclusion</b>	<b>39</b>
5.1 Brief Summary of Findings . . . . .	39
5.2 Looking to the Future . . . . .	39
<b>A Transit Schedule DTD</b>	<b>42</b>
<b>B The General Transit Feed Specification (GTFS)</b>	<b>44</b>

# List of Figures

1.1	Taxonomy of paratransit (Joubert, 2012) . . . . .	5
2.1	The building blocks of MATSim . . . . .	8
2.2	Entity Relationship Diagram (ERD) for a MATSim transit schedule . . . . .	11
2.3	Entity Relationship Diagram (ERD) for a GTFS feed . . . . .	13
2.4	Generation of a transit router network in MATSim (Adapted from Rieser (2010)) . . . . .	14
2.5	Starting stops available for different search radii . . . . .	15
2.6	Modification of a plan by the transit router (Adapted from Rieser (2010)). . . . .	16
3.1	NMBM road and rail network . . . . .	17
3.2	The matching process . . . . .	18
3.3	Data Flow Diagram (DFD) for the transfer process . . . . .	19
3.4	Data Flow Diagram (DFD) for the matching process . . . . .	19
3.5	Range of stop facilities in NMBM . . . . .	20
3.6	Example of an aggregate stop location used by Algoa Bus . . . . .	21
3.7	Different routes using the same stop facility . . . . .	21
3.8	Unedited solution . . . . .	22
3.9	Edited solution . . . . .	22
4.1	Number of trips throughout the day for different modes . . . . .	28
4.2	Comparison of planned and simulated route leg times . . . . .	29
4.3	Comparison of simulated and planned stop facility arrival times . . . . .	30
4.4	Algoa Bus Company route 748 . . . . .	31
4.5	Vehicle occupancy for route 748 . . . . .	33
4.6	Activity locations for agents using line 21 . . . . .	34
4.7	Planned activities for a single agent . . . . .	35
4.8	Simulated activities for a single agent . . . . .	36
4.9	Walking time to an activity location after alighting a transit vehicle . . . . .	38
5.1	Relative number of passengers entering a stop facility . . . . .	40

# List of Tables

4.1	Performance statistics for different transit router configurations (100 iterations) .	25
4.2	Route 748 details . . . . .	32
B.1	The GTFS “feed” files and associated content. . . . .	44

# List of Abbreviations and Acronyms

ABS	Agent-based simulation
DFD	Data Flow Diagram
DTD	Document type description
ERD	Entity Relationship Diagram
GPL	GNU Public License
GPS	Global Positioning System
GTFS	General Transit Feed Specification
MATSim	Multi-Agent Transport Simulation Toolkit
NHTS	National Household Travel Survey
NMBM	Nelson Mandela Bay Metropole
VDV	Association of German Transport Operators

# Chapter 1

## Introduction

### 1.1 Transport in South Africa

South Africa, in its efforts to develop integrated, effective and sustainable transport solutions, faces certain unique complexities, many attributable to its turbulent past. The South African transport environment is a complex mix of first and third world transport technologies, infrastructure and economic participants, shaped by past socio-political decisions (Fourie, 2009). The Apartheid policy of forced relocation moved millions of citizens to the periphery of major cities and towns, creating what is colloquially known today as townships. These mostly low income areas present a challenging distribution of demand for public transport services: affluent communities with access to private transport are close to economic activities in city centra, while lower income groups without access to private transport are situated on the periphery. Connecting the townships via traditional public transport services such as rail and bus is not always viable due to the distances involved, which in turn results in inadequate and often unreliable public transport services. This provides an ideal environment for the development of paratransit (Joubert, 2012).

Paratransit services span different levels of spatial and temporal flexibility, as illustrated in Figure 1.1, and are to be found mainly in cities of the developing world where transit demand exceeds supply.

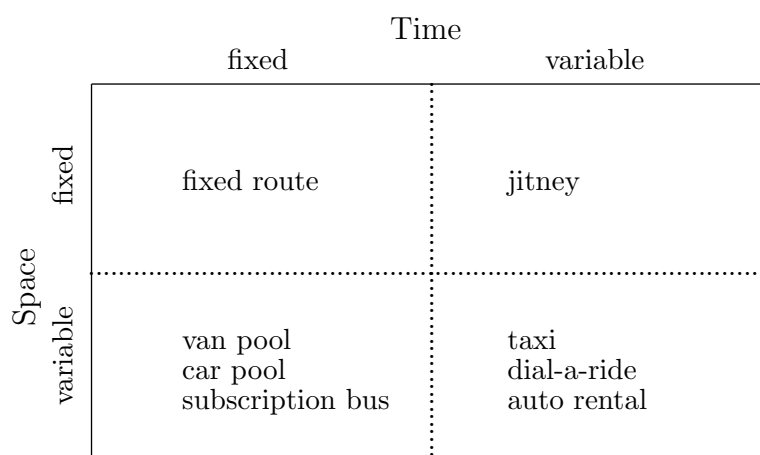


Figure 1.1: Taxonomy of paratransit (Joubert, 2012)

In South Africa, the paratransit minibus taxis form a major and integral part of the overall transit system, resembling the jitney service in following predetermined routes without a fixed

schedule. Highly accessible and flexible, minibus taxis bridge the divide between transit demand and public transport availability. One of the major challenges facing the South African transport environment is the lack of integration between formal public transport services and paratransit minibus taxis. Instead of the different modes coordinating for the benefit of the transit passenger, they are in constant competition and rivalry with one another. Government efforts to regulate and integrate the minibus taxi industry have been largely unsuccessful.

These challenges require that transportation management and infrastructure investment be based on well-informed decisions anticipating the effect on the system as a whole, as the consequences of poor decisions may be costly and long term.

## 1.2 Transport Modelling in South Africa

The need for detailed, dynamic results and major technological improvements have led to the development of simulation as a modelling approach suited to modern traffic systems (Rieser, 2010). Agent-based simulation (ABS) models represent every person in the area of interest as an object, or agent. Each agent is given a set of attributes and activities, but crucially, it may decide on its own how it behaves within the greater system while adhering to certain constraints. Fourie (2009) identified ABS as the only modelling paradigm capable of capturing the complex interactions of the South African transport system. An initial implementation of the Multi-Agent Transport Simulation Toolkit (MATSim), for private vehicle traffic in the Gauteng Province showed the potential of the approach. Since the initial implementation, several researchers have added to the body of knowledge in a process of continuous improvement. Notable additions to the initial implementation include improving the simulation accuracy by incorporating additional, more realistic, activities into the initial transport demand (Van der Merwe, 2011) and adding freight to the simulation (Joubert et al., 2010).

One of the major advances has been the inclusion of transit into the MATSim functionality. Previously, the impact of decisions on multi-modal systems could not be simulated accurately, leaving such questions as mode-switching unanswered. Rieser (2010) describes in detail the extension of MATSim to incorporate transit, and then proceeds to develop and refine an approach to add public transport vehicles to a simulation. A large-scale multi-modal simulation of the greater Zurich region is undertaken, showing the applicability of the integrated simulation in a large, urban context. Wevell (2011) investigates the applicability of MATSim in the South African transit environment, laying the foundation for a large-scale multi-modal implementation.

## 1.3 Research Design and Methodology

The modelling of private vehicle and freight traffic using ABS has been very successful in South Africa, but transit system modelling has comparatively lagged behind. *This* research project aims to implement a large-scale multi-modal MATSim model, which includes transit, for the Nelson Mandela Bay Metropole (NMBM).

NMBM was chosen as a study area due to its traffic characteristics. Compared to areas like Gauteng, through traffic is minimal, reducing simulation complexity. The Metropole has two official transit operators, as well as a paratransit minibus taxi system. The formal transit providers are Eastern Cape Metrorail, which operates a railway from Uitenhage to Port Elizabeth, and the Algoa Bus Company which operates routes in the greater NMBM area. For the purpose of this project, only the formal transit operators will be modelled as the current ability to model the paratransit taxi industry is limited.



This project will follow the framework developed by Rieser (2010) in building and analysing the transit simulation. Available General Transit Feed Specification (GTFS) data for the region is matched to the traffic network and transformed into a MATSim transit schedule. GTFS2TransitSchedule, a MATSim extension, is used in this regard. The simulation can now be executed and analysed. Schedule data is compared to simulated values and components of the transit system are investigated using the simulation results.

## **1.4 Outline of Document**

Chapter 2 presents an analysis of the MATSim simulation process and transit data requirements, specifically the transit schedule. In addition, the General Transit Feed Specification (GTFS) and its interface with the MATSim transit schedule are investigated. Chapter 3 describes the process of matching GTFS data to a network and transforming it into a transit schedule. Chapter 4 presents an analysis of the simulation results. Planned schedule data is compared to simulated values and insights into the transit system are derived. Chapter 5 summarises the findings.

## Chapter 2

# Literature Review

### 2.1 The Multi-Agent Transport Simulation Toolkit

Current public transport assignment models adapt network assignment models to work for public transport traffic. However, these models do not include the interaction between public transport services and private transport. The Multi-Agent Transport Simulation Toolkit (MATSim), an agent-based simulation technology, supports both private and public transport, ensuring that large-scale, integrated simulations are now possible. MATSim is written in the Java programming language and distributed under the GNU Public License (GPL), making it free for download and use (MATSim, 2012).

#### 2.1.1 The MATSim Process

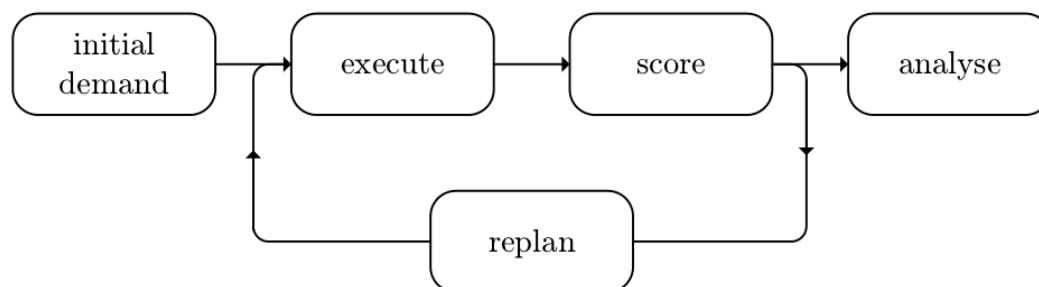


Figure 2.1: The building blocks of MATSim

Figure 2.1 shows the schematic control flow of MATSim. The process starts with the initial demand describing an initial set of plans for the population of agents. These plans are fed into the mobility simulation, starting an iterative process of executing, scoring and replanning the agent plans until a steady state system is reached. This iterative process is based on the concept of a co-evolutionary algorithm. As discussed in Fourie (2009), this provides a means of generating and improving populations of solutions to solve optimisation problems. This is accomplished using mechanisms inspired by biological evolution such as mutation and selection. In MATSim's context, the population refers to the population of plans associated with each agent which will adapt to the transport environment.

### 2.1.2 The MATSim Building Blocks

This section explores the MATSim building blocks in more detail.

#### The Controller

The Controller manages the data flow and interaction between individual modules during a simulation. It can be configured to allow for the simulation of different scenarios, system configurations and output requirements (Rieser, 2010).

#### Initial Demand

MATSim requires a synthetic population of agents, each with individual transport related attributes and daily activity plans (Van der Merwe, 2011). To facilitate accurate simulations, the population of agents should be representative of the actual socio-demographic structure. Agents plan their day by making decisions with regard to a set of choice dimensions. Balmer (2007) identifies the following choice dimensions:

- **Mode choice:** Should I walk, drive or use public transport?
- **Route choice:** Which route should I take to get to the University?
- **Location choice:** Where should I train?
- **Activity type choice:** Should I go and see a movie?
- **Activity chain choice:** Should I go swimming before or after work?
- **Activity starting time choice:** When should I train today?
- **Activity duration choice:** How long should I train?
- **Group composition choice:** Who should I take along?

Each agent's plan, or detailed schedule for the day, represents a single point in a multi-dimensional space defined by the choice dimensions. The initial demand describes the initial plans of all simulated agents. While some choice dimensions, such as route, time and mode may be optimised during the simulation, others are not. For these choice dimensions, the initial demand must provide meaningful values for a realistic simulation (Rieser, 2010).

#### Simulation

The traffic flow simulation, or mobility simulation, is responsible for executing the agents' plans within the transport system, while respecting physical constraints such as speed limits, the flow and storage capacity of a road and transit vehicle capacity (Rieser, 2010). Events are generated for every notable change in the state of the system and represent the output of the mobility simulation.

An essential input into the mobility simulation is the traffic carrying network. This network is described by a series of nodes and links. In a road network, nodes correspond to intersections, while links correspond to road segments. In a rail network, links correspond to the rails, while nodes correspond to the switches or stops (MATSim, 2012).

## Scoring

MATSim uses a simple utility based approach to calculate a plan score (Balmer, 2007):

$$U_{total} = \sum_{i=1}^n U_{perf,i} + \sum_{i=1}^n U_{late,i} + \sum_{i=1}^n U_{travel,i}$$

where  $U_{total}$  is the total utility for the executed plan,  $n$  is the number of activities,  $U_{perf,i}$  is the positive utility for performing activity  $i$ ,  $U_{late,i}$  is the negative utility for arriving late at activity  $i$  and  $U_{travel,i}$  is the negative utility for travelling to activity  $i$ . Each utility value is a time dependent function of arbitrary complexity (Fourie, 2009). Total utility increases for time spent performing activities, while it decreases for time spent travelling and arriving late at activity locations.

## Replanning

Fourie (2009) describes replanning as a sequence of configurable algorithms, analogous to the mutation and selection processes found in EAs. Balmer (2007) lists a number of algorithms used to select from and then expand the population of plans. These algorithms ensure that the population of plans gradually adapt to the transport environment by both mutating along the time, route and mode choice dimensions and discarding plans which are not performing well. This process of adaptation leads to improvements for average plan scores and travel times. It will eventually reach a relaxed state where changes in agent plans do not produce any further improvement in total utility, balancing the marginal utilities.

### 2.1.3 The MATSim Input Components

MATSim requires certain input components to perform a transit simulation (MATSim, 2012). These include:

- **The Simulation Configuration**  
The configuration specifies the parameters of the simulation, such as the number of iterations, which replanning modules to use, etc.
- **The Multi-modal Network**  
The multi-modal network consists of links and nodes which represent the road and rail infrastructure of the transport system under observation. For each link, the transport mode, or modes, which can use that link, must be specified.
- **The Plans/Population**  
The plans file describes the simulated population and its travel demand. Agents must have the option to use transit in performing their plans.
- **The Transit Schedule**  
The transit schedule represents the public transport schedule of all transit providers in the area under observation. Included items are the stop locations, departure times and routes.
- **The Transit Vehicles**  
The available public transport vehicles need to be described. Both the attributes of different vehicle types (capacity, speed, length etc.) and the individual vehicles are specified.

Optional input components include:

- **The Facilities**

Facilities can broadly be described as buildings where agents perform activities. If facilities are used, it is possible to define spatially differentiated opening times, thereby improving the simulation realism.

- **Counts**

Count data is used for validation. The allows for simulated traffic to be compared to real world data from counting stations.

All the requirements should be saved in individual MATSim-specific XML files.

## 2.2 The Transit Schedule

The transit schedule represents the central data structure for the simulation of public transport (Rieser, 2010). It contains information about both the available infrastructure (stop facilities) and operational data (transit lines, routes and departure times).

The **stop facilities** describe locations where passengers can enter or alight transit vehicles, such as bus stops and train stations. Stops are described by a unique identifier, a coordinate (for estimating agent walking distance), a reference to the network link from where it can be accessed and an attribute which specifies if other vehicles can pass the transit vehicle while it is stopped. A **transit line** is a regularly available connection between two end points. In the real world, a transit line is often associated with a number (e.g. bus 14). In MATSim, a transit line has an identifier and can have multiple associated transit routes. A **transit route** describes one specific route of a transit line. Most transit lines have at least two transit routes, one in each direction . Additional routes may be required when different travel times are anticipated during the course of the day, or when certain areas are only serviced by the transit line during peak times. The route is described as both a list of **transit route stops** with arrival and departure time offsets (for passenger route planning) and a network route (for the transit vehicle driver). The mode of transport associated with the route is also specified. A set of **departures** is stored for each transit route, specifying the time at which a specified vehicle starts the associated route.

The data is structured hierarchically, with many 1 :  $n$  relationships. One transit schedule may consist of multiple transit lines, one transit line contains several transit routes, and each route usually has several departures. Figure 2.2 illustrates the data structure as an Entity Relationship Diagram (ERD).

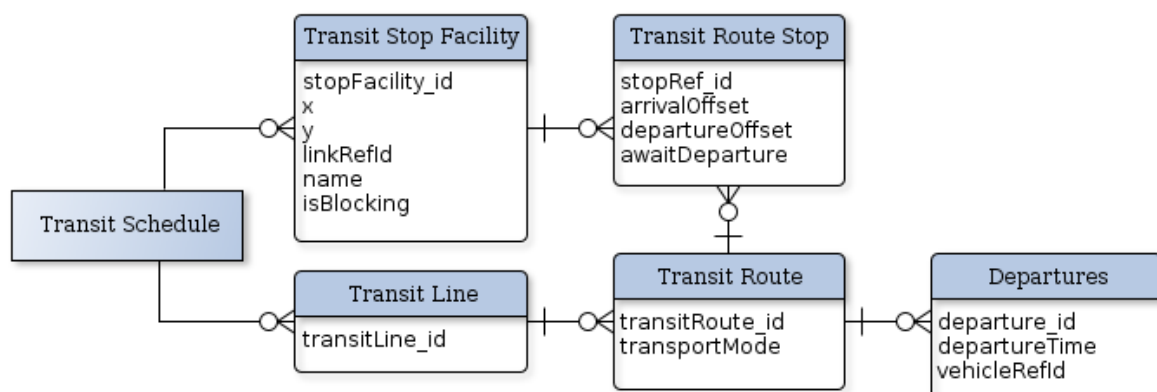


Figure 2.2: Entity Relationship Diagram (ERD) for a MATSim transit schedule

For use in the simulation, the transit schedule has to be stored in a MATSim specific XML format. Appendix A contains the document type description (DTD). This specifies, for the XML transit schedule, precisely where elements may appear and what their attributes and contents may be.

### 2.2.1 Transit Data Models

The preparation of a transit schedule for use in MATSim often requires a substantial amount of resources. In addition to acquiring data from transit operators, the data must be converted to MATSim's specified format. The process of acquiring data for the entire transit system is often hampered by the absence of a common format used by operators for storing and exchanging transit data. However, as most transit systems and operations bear several similarities, standardised models are available. These data models become absolutely essential when seamless interoperation between different operators is required. Some of the most widely used models include:

- **Transmodel**, which was developed as a comprehensive conceptual model to underpin the development of transport information systems and streamline the exchange of transit data within the highly interconnected, complex European environment (Kizoom and Miller, 2008).
- **VDV**, published by the Association of German Transport Operators (VDV), provides recommendations regarding data structures and file formats for the exchange of transit related data in Germany (Rieser, 2010).
- **TransXChange**, used in the United Kingdom, was developed for the *Transport Direct Portal*, an internet-based multi-modal journey planner (Kizoom and Miller, 2008).
- **The General Transit Feed Specification**, developed by *Google Inc.* for *Transit on Google Maps* enables transit agencies to easily capture and share their transit data with transit users and developers.

### 2.2.2 The General Transit Feed Specification

The General Transit Feed Specification (GTFS) represents a fairly recent, but nonetheless very important development in standardised transit data modelling. As stated, it was initially developed by *Google Inc.* to allow transit operators to upload their schedule data for use in *Google Maps*, but quickly evolved into a commonly used data-sharing specification. Its popularity can probably be traced to its simple structure: GTFS focuses on schedule information *relevant* to the transit passenger. This allows for a much simpler structure than Transmodel or TransXChange equivalents, which have to capture detailed operational data as well.

Data is stored in a series of text files, each containing information relating to a particular aspect of the transit system. Wevell (2011) states that not all GTFS feed data may be applicable to all situations and further identifies that data which is relevant for performing MATSim simulations in a South African context. The data structure of a GTFS feed suitable for transit simulation purposes in South Africa is shown in Figure 2.3. The complete set of GTFS feed files and their associated content is shown in Appendix B.

The GTFS feed captures data related to one or more transit agencies, which operates according to a weekly schedule defined in `calendar.txt`. Information regarding each agency's transit lines are contained within `routes.txt`, while the MATSim transit route corresponds to trips within `trips.txt`. The trip is further described in `stop_times.txt`, which contains information such as the trip stop sequence and time profile (equivalent to the transit route stops of the

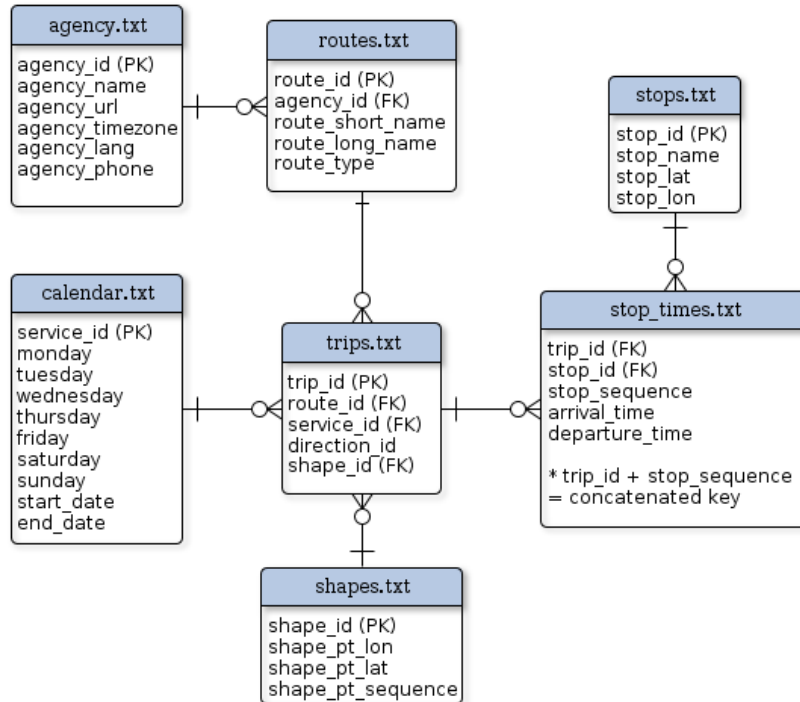


Figure 2.3: Entity Relationship Diagram (ERD) for a GTFS feed

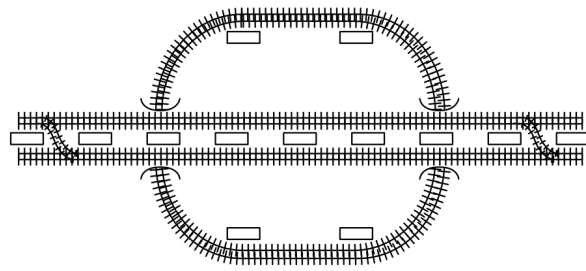
transit schedule). Global Positioning System (GPS) trace data contained in `shapes.txt` shows the transit vehicle location at regular intervals and can be used for network matching purposes.

### GTFS2TransitSchedule

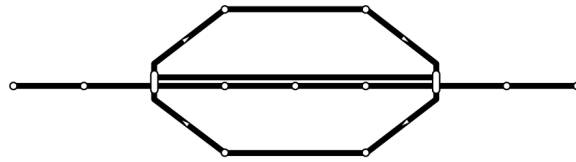
GTFS2TransitSchedule is a MATSim extension designed to assist in the conversion and map-matching of schedules given in GTFS format to MATSim’s transit schedule format. It takes as input the set of GTFS files and associated multi-modal network. For each GTFS trip, a simple map-matching algorithm is applied to generate a preliminary transit route solution (path or sequence of connected network links) and a set of stop-link relationships (Ordonez and Erath, 2011). Should manual editing be required, the proposed solution can be visualised, edited and verified for each route. The solution is modified by manipulating path or stop-link relationship data (MATSim, 2012). Upon completion of the process, the required MATSim-specific XML files are created.

## 2.3 The Transit Router

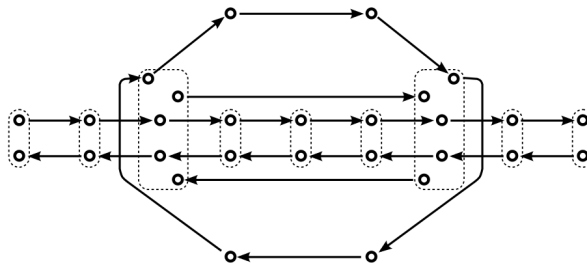
The transit router is used in the initial demand and replanning phases of MATSim to calculate and assign meaningful routes to agents. As described by Rieser (2010), data contained in the transit schedule is used to create a new logical topology or *transit network*. Stop facilities represent network nodes. For each transit route, links are added, connecting the stops of the route. Next, connecting links are added to stops (within a configurable distance) to allow for line changes. Figure 2.4 illustrates the steps followed in creating the transit network.



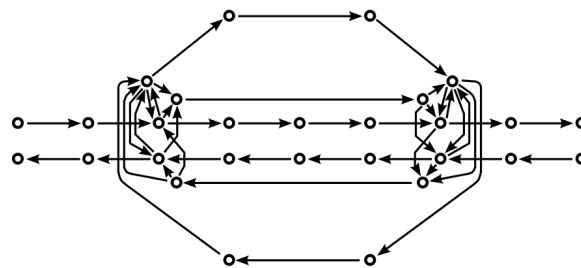
(a) Physical infrastructure. Squares represent stop facilities.



(b) Line map showing a circular line, an express line and a regular line.



(c) Transit router network without transfer links. Dotted areas group nodes belonging to the same stop facility.



(d) Complete transit router network with transfer links.

Figure 2.4: Generation of a transit router network in MATSim (Adapted from Rieser (2010))



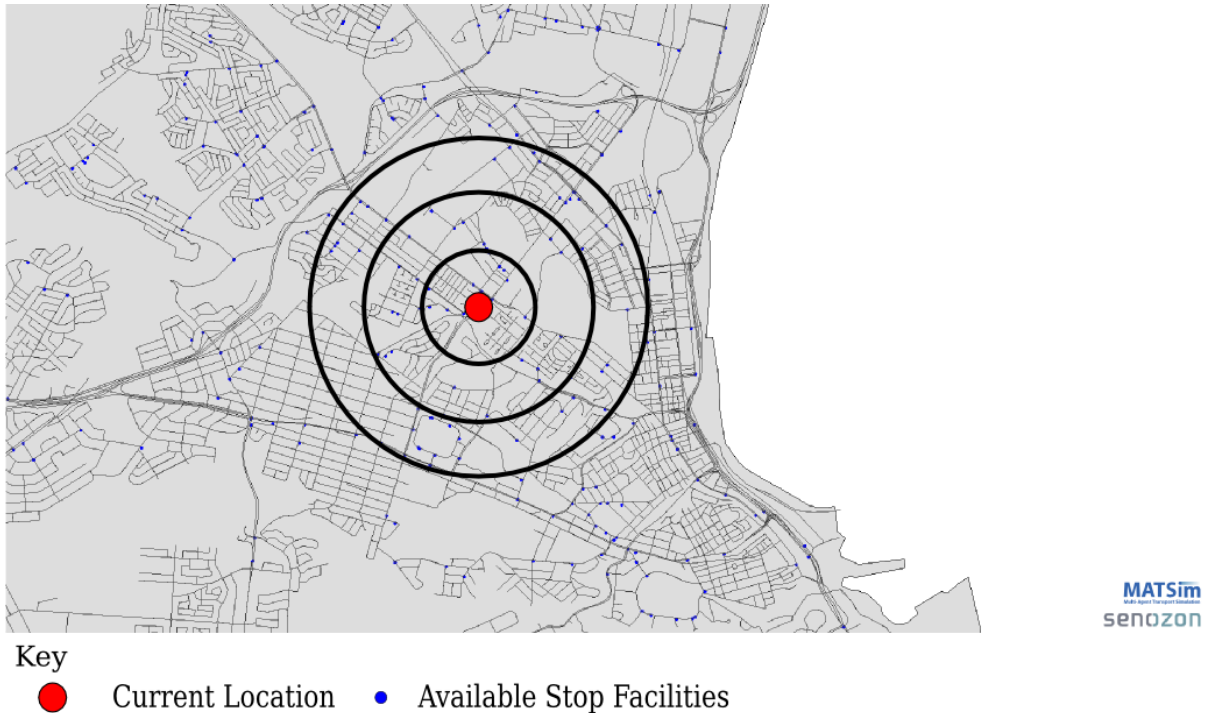


Figure 2.5: Starting stops available for different search radii

The transit router now has to search for a connection. A shortest path algorithm is applied to find the ‘best’ connection for a certain transit user. The term ‘best’ is used as, in the simulation, an agent does not plan his day based on the available transit offerings (as most transit users do) - instead, the agent will only check the transit network for available offerings upon completing an activity. This occurrence is negated as the agent’s time choice dimension will also be modified during replanning, meaning that the agent will eventually find the most suitable connection.

To increase computational efficiency, the transit router can be configured by specifying the following input parameters:

- The **search radius** determines which stops an agent will consider when searching for a transit route - all stops within this distance from an agent’s ending activity location will be considered as potential stops to start a transit trip. Figure 2.5 shows the number of stops which may be considered from an ending activity location for different values.
- The **connection distance** determines the maximum (beeline) walking distance to stops which might be used to continue the journey when switching transit line or route. All stops within this distance, from a given stop, are considered as transfer stops.

The transit router returns the best connection as a list of *stages*. A stage is defined as a continuous part of a leg (journey between activities) where a person uses one specific mode of transportation. In MATSim, plans consist of an alternating list of activities and legs. In Figure 2.6, the agent’s initial plan consists of home and work activities connected by transit legs. The transit router identifies a multi-staged connection, bounded by walking legs to and from stop facilities. In the routed plan, which replaces the initial plan for simulation purposes, zero length transit interaction activities are added between legs, and the plan can be used for simulation purposes.

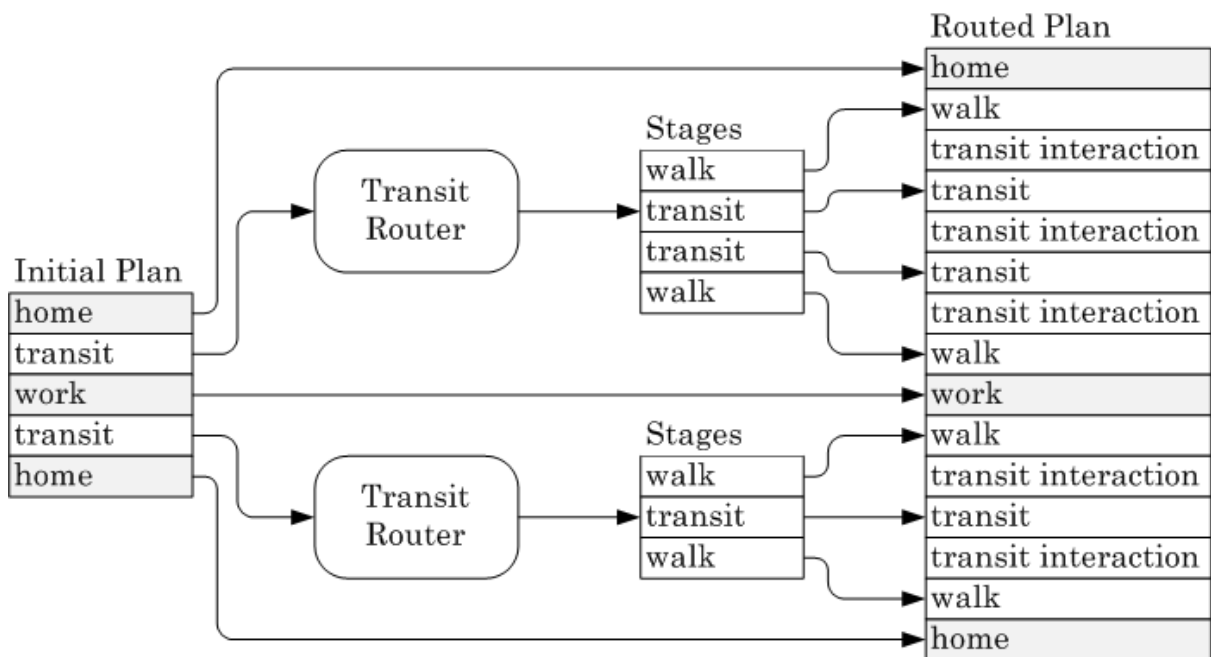


Figure 2.6: Modification of a plan by the transit router (Adapted from Rieser (2010)).

## Chapter 3

# Data Preparation

The main transit data for this research project was provided by Algoa Bus Company. It consists of a GTFS feed specifying 1678 routes relating to 246 lines and 914 unique stop locations. Additional transit data for Eastern Cape Metrorail was obtained via the GTFS Data Exchange ([www.gtfs-data-exchange.com](http://www.gtfs-data-exchange.com)). It consists of a GTFS feed specifying 12 routes relating to a single line and 10 unique stop locations.

For use in MATSim, this transit data has to be matched to a network and then transformed into a transit schedule. The network was derived from OpenStreetMap (OpenStreetMap Development Team, 2012), a free geographic data and mapping service. It consists of 44235 nodes and 95080 links representing the NMBM road and rail transport network, and includes information regarding the different transport modes which may use each link.

Figure 3.1 provides a visual display of the network.

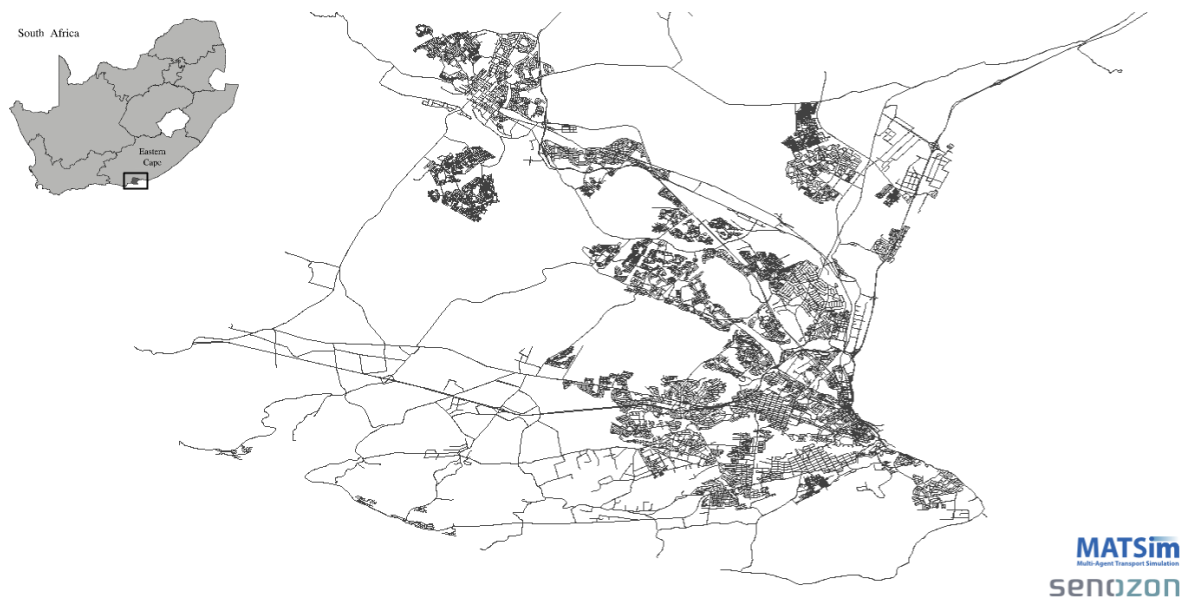


Figure 3.1: NMBM road and rail network

## 3.1 The Matching Process

The matching process attempts to pair GTFS transit information with the network to produce a MATSim transit schedule, as illustrated in Figure 3.2. It is important to note that these data structures are inherently different, and not all data contained in GTFS feeds is required for transit simulation. The matching process is therefore not simply a conversion of data, but rather involves processing and combining relevant data to create information required during transit simulation. GTFS2TransitSchedule, a MATSim extension, was used to assist in this process.

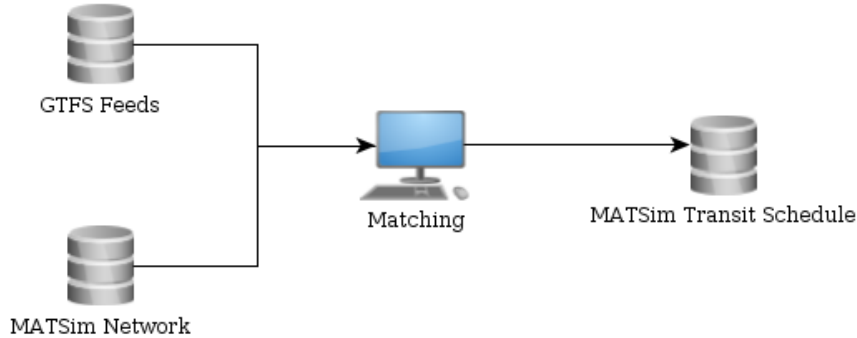


Figure 3.2: The matching process

### 3.1.1 Flow of Data and Initial Matching

Understanding the flow of data is paramount if GTFS2TransitSchedule is to be effectively used. Two distinct data processes are utilised in gathering the information required for a transit schedule.

The first process, illustrated in Figure 3.3, involves transferring data which does not require matching from the GTFS data structure to the transit schedule data structure. This includes the different stop locations, route information and departure times. In general, entity identifiers are maintained for ‘equivalent’ data structures, such as `routes.txt` & `TRANSITLINE`, `trips.txt` & `TRANSITROUTE` and `stops.txt` & `TRANSITSTOPS`. Specific flows of importance include `route_type`, specified in `routes.txt`, which determines the `TRANSPORTMODE` of the matching `TRANSITROUTE` entry, and `arrival_time` at the first stop of the trip which sets the `DEPARTURETIME` in `DEPARTURES`.

The next process, illustrated in Figure 3.4, involves matching the GPS trace data contained in `shapes.txt` and the stop locations specified in `stops.txt` with the network. The GTFS data structure was designed to facilitate the inclusion of transit information in map and routing services - therefore, routes are simply described by the stops serviced (`stop_times.txt`) and not the actual route followed, as is required for simulation purposes. Fortunately, the GPS trace data contained in `shapes.txt` shows the transit vehicle location at regular intervals, facilitating the network matching process. This semi-automatic process starts with the application of a simple map-matching algorithm. Its objective is to generate a preliminary solution (path or sequence of connected network links and a set of stop-link relationships) for one route, taking into account its stop sequence profile and sequence of GPS points (Ordonez and Erath, 2011). This preliminary solution can then be modified by editing the sequence of connected network links or set of stop-link relationships via the user interface.

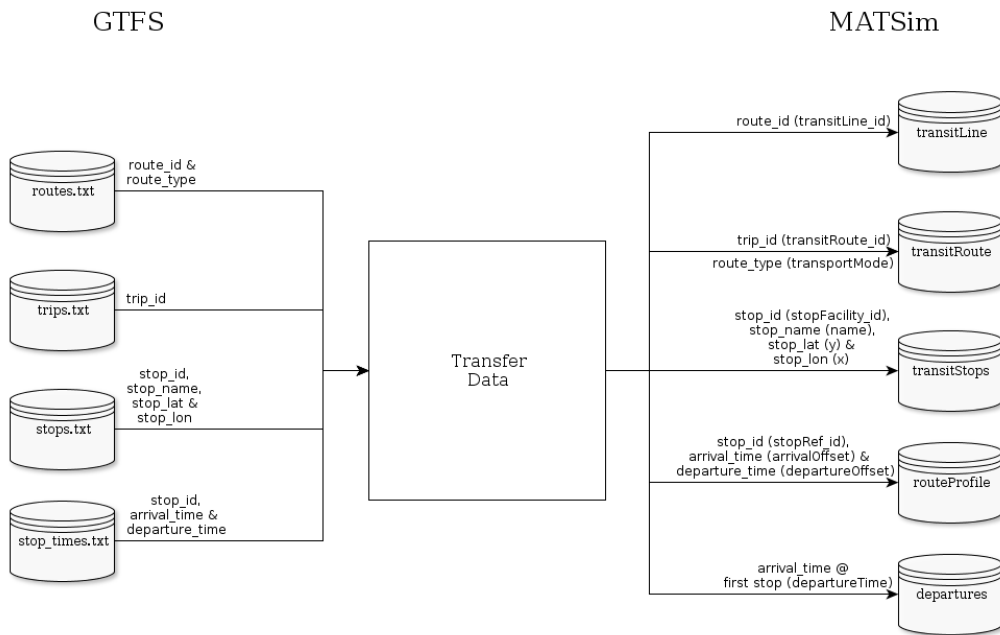


Figure 3.3: Data Flow Diagram (DFD) for the transfer process

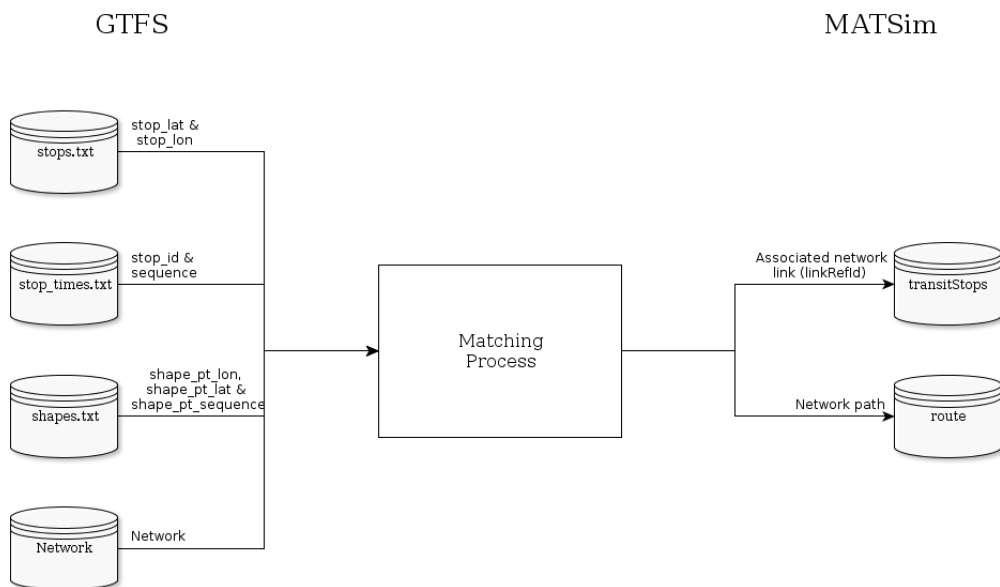


Figure 3.4: Data Flow Diagram (DFD) for the matching process

Given the peculiarities of South African transit, such as the widespread absence of stop facility infrastructure and a network which does not (yet) capture the extended road network, substantial editing was required.

### 3.1.2 Solution Modification

#### Modifying Stop-Link Relationships

*Google Maps*' satellite imagery and *Street View* were used extensively as a tool in identifying the stop-link relationships. Stop facility infrastructure (if any) or informal stop facilities could be located and then matched to the correct network link. Using this approach meant that actual stop facilities could be identified, within a fair degree of accuracy, from more than 1100km away. To illustrate the range of stop facilities encountered, Figure 3.5 displays four of NMBM's stop facilities. Figures 3.5a and 3.5b show a bus shelter and road intake, purpose built stop infrastructure similar to that found worldwide. In contrast, Figure 3.5c and 3.5d show two of the common informal stops, with no clear demarcation or signage to indicate their presence.



(a) Formal example 1



(b) Formal example 2



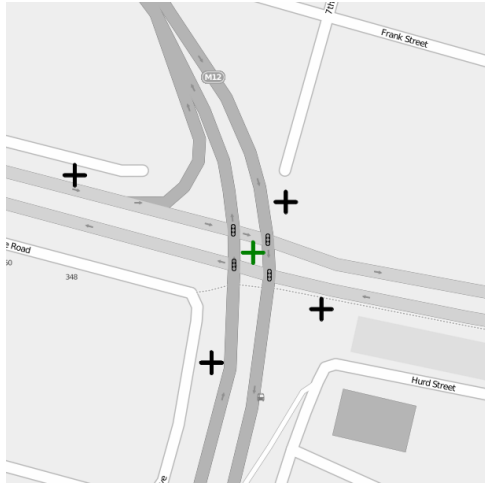
(c) Informal example 1



(d) Informal example 2

Figure 3.5: Range of stop facilities in NMBM

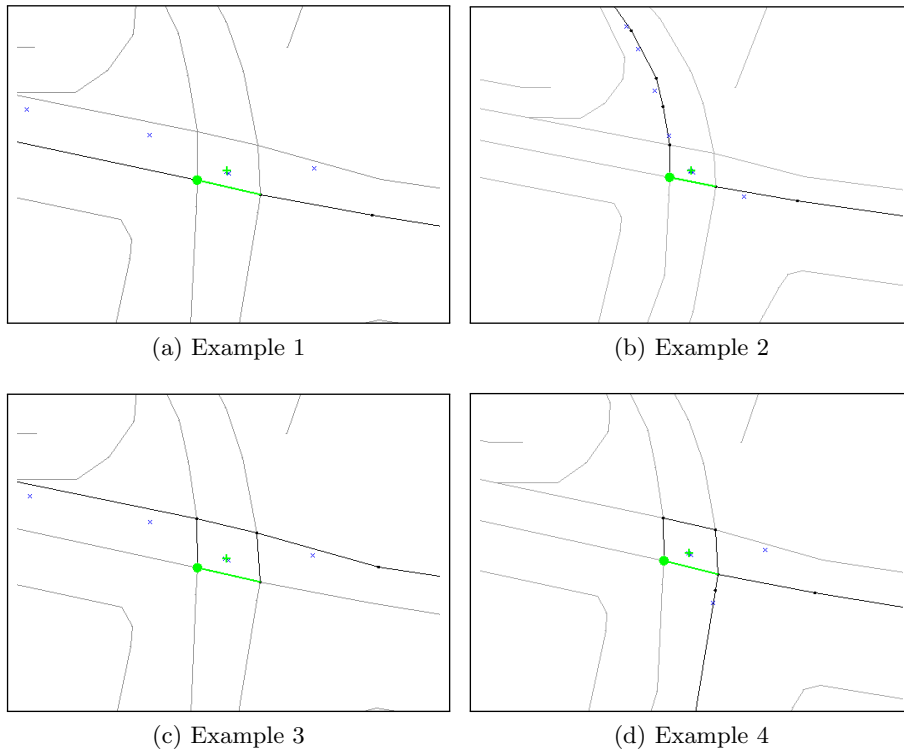
The MATSim transit schedule is designed to assign each stop facility (which may feature in several routes) to only one link in the network (which should belong to the path of all routes in which this stop features). While this may work well for established systems with clearly demarcated stop facilities (known to the operator), it creates problems in the South African context where operators only know the aggregate stop locations (which actually consist of several stop facilities). An example of this is shown in Figure 3.6 where Algoa Bus uses an aggregate stop location to describe their route profiles, while several stop facilities are actually present. The aggregate stop location replaces four actual stop facilities in their GTFS feed.



Legend:  $\blackplus$  actual stop  $\greenplus$  aggregate stop

Figure 3.6: Example of an aggregate stop location used by Algoa Bus

This occurrence necessitated small deviations from the actual routes to complete the matching process. As shown in Figure 3.7, all routes accessing the aggregate stop must include its referenced network link (shown in green) as part of the network path, whether this reflects the actual route or not. Taking into account that most routes consist of several hundred links, the deviations are not expected to measurably affect simulation accuracy.

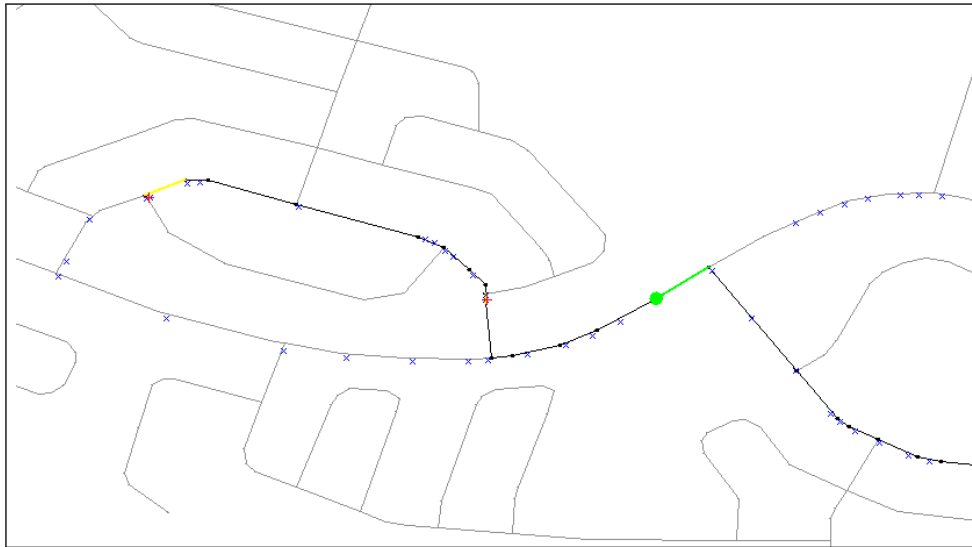


Legend:  $\blackline$  Path  $\greenline$  Stop link  $\greenplus$  Stop  $\bluecross$  GPS point

Figure 3.7: Different routes using the same stop facility

## Modifying the Path

The map-matching algorithm used to identify an initial solution was not designed to account for the peculiarities of public transport routing or network inaccuracy (Ordonez and Erath, 2011), but simplifies the routing process to a great extent. Figure 3.8 displays an unedited portion of a route. Using the stop and GPS trace sequence as a guide, the route may be modified to represent the actual path, as displayed in Figure 3.9.



Legend: / Path / Stop link + Stop × GPS point

Figure 3.8: Unedited solution



Legend: / Path / Stop link + Stop × GPS point

Figure 3.9: Edited solution



After the solution has been sufficiently modified, it is verified to ensure its correctness. GTFS2TransitSchedule sequentially checks if:

1. the path is joined;
2. the path is without U-turns;
3. there are any repeated links;
4. every stop of the route has a stop-link relationship;
5. every stop is related to a link inside the path;
6. the sequence of related links matches the sequence of correspondent stops in the route profile;
7. the first link of the path is related to the first stop on the route profile.

Should it be required, verifications 2), 3) and 5) can be disabled as the actual route often includes U-turns, repeated links or peculiar stop locations. Should the solution pass the verification, the transit route is saved, if not, further modification is required.

With the help of GTFS2TransitSchedule's semi-automatic process described above, 1678 routes relating to 246 lines and 914 unique stop locations for Algoa Bus and 12 routes relating to a single line and 10 unique stop locations for Eastern Cape Metrorail were matched to the network and transformed into a MATSim transit schedule.

## Chapter 4

# Analysis and Discussion of Simulation Results

A 10% sample population was used for the multi-modal simulation, representing 116 511 agents. Agents were allowed to modify time and route choice dimensions in the simulation; however, they were not allowed to change their mode of transport for a given journey. This reflects the context of the Nelson Mandela Bay Metropole (NMBM) and South Africa in general, where private vehicle users are unlikely to switch to public transport and transit users generally lack an alternative. Transit agents may freely change between transit lines, routes and departure times, trying to find the least cost path with regards to walking-, in-vehicle-, travel-, transfer- and waiting-time. Additional costs, such as bus fares, were not considered.

Car, bus and train trips were simulated jointly using the mobility simulation and could therefore cause and experience network congestion. All other modes, including walk and taxi, were teleported. As discussed in Agarwala and Zilske (2012), upon encountering a trip of this type the mobility simulation will use the beeline distance and a mode specific factor to calculate an expected travel time. Trips of this type are only executed logically, meaning that no detailed information about their interaction within the greater traffic system is available.

Before presenting and analysing the results, it is important to note that the implemented simulation model has not been extensively calibrated. Finding the “best” values for a simulation of this size is a complex topic in itself (See Flotterod et al. (2009)). As such, these results set out to illustrate the value and benefit of using a large-scale agent- and activity-based multi-modal simulation model.

### 4.1 Transit Router Configuration and Performance

Simulation performance and accuracy are influenced by the input parameters specified. As discussed in Chapter 2, the transit router matches agents to transit routes for use in the mobility simulation. The area within which it searches for transit services and the maximum connection distance (between services) can be configured as follows:

- The **search radius** determines which stops an agent will consider when searching for a transit route - all stops within this distance from an agent’s ending activity location will be considered as potential stops to start a transit trip.
- The **connection distance** determines the maximum (beeline) walking distance to stops which might be used to continue the journey when switching transit line or route. All stops within this distance, from a given stop, are considered as transfer stops.

Table 4.1: Performance statistics for different transit router configurations (100 iterations)

	Standard	Extended
Search radius (m)	1000	3000
Connection distance (m)	100	1000
Avg. time per iteration (mm:ss)	04:37	08:51
Avg. time for replanning (mm:ss)	01:26	03:20
Avg. time for traffic flow simulation (mm:ss)	03:00	05:15
Number of events in iteration 100	12 901 514	12 910 605
Number of routes not found	171	102
Number of agents stuck	807	890
Average score	173.6	175.5

Table 4.1 compares the performance (efficiency and effectiveness) of the simulation over 100 iterations for two configurations. In the first configuration a search radius of 1km and connection distance of 100m is used. This represents standard values, based on European transit systems. In South Africa, these distances will be *much* greater, as transit users often walk long distances to and between transit services. Accordingly, the second configuration has tripled the search radius to 3km and increased the connection distance to 1km.

This higher level of detail comes at a computational price. The extended configuration has an average iteration time of 8 minutes 51 seconds, nearly double the 4 minutes 37 seconds of the standard configuration. As expected, the number of agents for whom routes could not be found dropped by 68% to 102 for the extended configuration. Somewhat counter-intuitively, the average score and number of “stuck” agents (agents who arrived late for their transit service and now perpetually wait at the bus stop) increased. This phenomenon is attributable to the way MATSim handles transit agents for whom no route could be found. Instead of removing them from the simulation, they are teleported to the next activity location, and may therefore continue their journey. As the extended configuration has resulted in less agents without routes (by searching for far removed routes), more agents become “stuck” when missing their transit service, which will result in a low score.

## 4.2 Simulation Outcome

A total of 700 iterations were required for the simulation to reach a relaxed state. This is significantly more than is required for an equivalent private vehicle or freight simulation. The high number of iterations required can be traced to the way MATSim “matches” transit users to an available service. As explained in Chapter 2, agents rely on the combination of time and route adaptations to find a service best suited to their needs.

Figure 4.1a shows the number of simulated car trip departures, arrivals and the number in progress throughout the day. As expected, clear morning and afternoon peaks appear as agents travel to and from work and education activities. The morning peak occurs at 07h30 with 1767 trips in progress, while an afternoon peak of 1691 is observed at 17h10 (based on the 10% sample). In comparison, the number of simulated transit trips proved much more stable throughout the day, as shown in Figure 4.1b. The morning peak occurs at 07h20 with 1153 trips in progress, while an afternoon peak of 1389 is observed at 14h20 (based on the 10% sample). It is interesting to note that the morning and afternoon peak transit demand starts earlier, ends later, and is generally more spread out than that of cars. This is rooted in the nature of transit, with users often needing time to walk to and from stop facilities after completing activities and generally travelling greater distances than car users. Another observation is the impact of school children using transit, in effect increasing midday demand on their return from school.

### 4.2.1 Comparison of Planned and Simulated Route Leg Times

As a decision maker one would be interested to know how close one’s transit data, in this case GTFS data, is to the actual travel time of the vehicles. To analyse this using an agent-based model, one can take two different perspectives. Firstly, the actual (simulated) leg travel time, that is the time travelled between two consecutive stops on any route, can be compared against the planned leg travel time contained in the GTFS data. In this case we treat each leg independently. Additionally, the actual (simulated) arrival time can be compared against the planned values contained in the GTFS data for each stop on the route, taking into account the knock-on effect a delay on one leg will have on the subsequent arrival times.

For the first perspective, Figure 4.2 compares the planned and simulated leg travel times in the form of a scatterplot, each point representing a different leg. In a perfect world all the points should be very close to the diagonal, the solid line representing a 1:1 correlation between simulated and planned leg times. Points above the line represents legs where the simulated leg times are longer than planned, indicating instances where the GTFS underestimates the travel time. Conversely, points below the line represents simulated leg travel times that are shorter than planned, indicating an overestimation in the GTFS data. As we use a log scale on both axes, we provide two reference lines, the first is above the 1:1 reference line and represents the 2:1 correlation where simulated travel times are double that of the planned times, and the second is below the reference line and represents the 1:2 correlation where simulated travel times are half that of the planned times. Applying a linear fit to the data in Figure 4.2, delivers the equation:

$$Y_{\text{Simulated}} = 0.53X_{\text{Planned}} + 68 \mid R^2 = 0.25$$

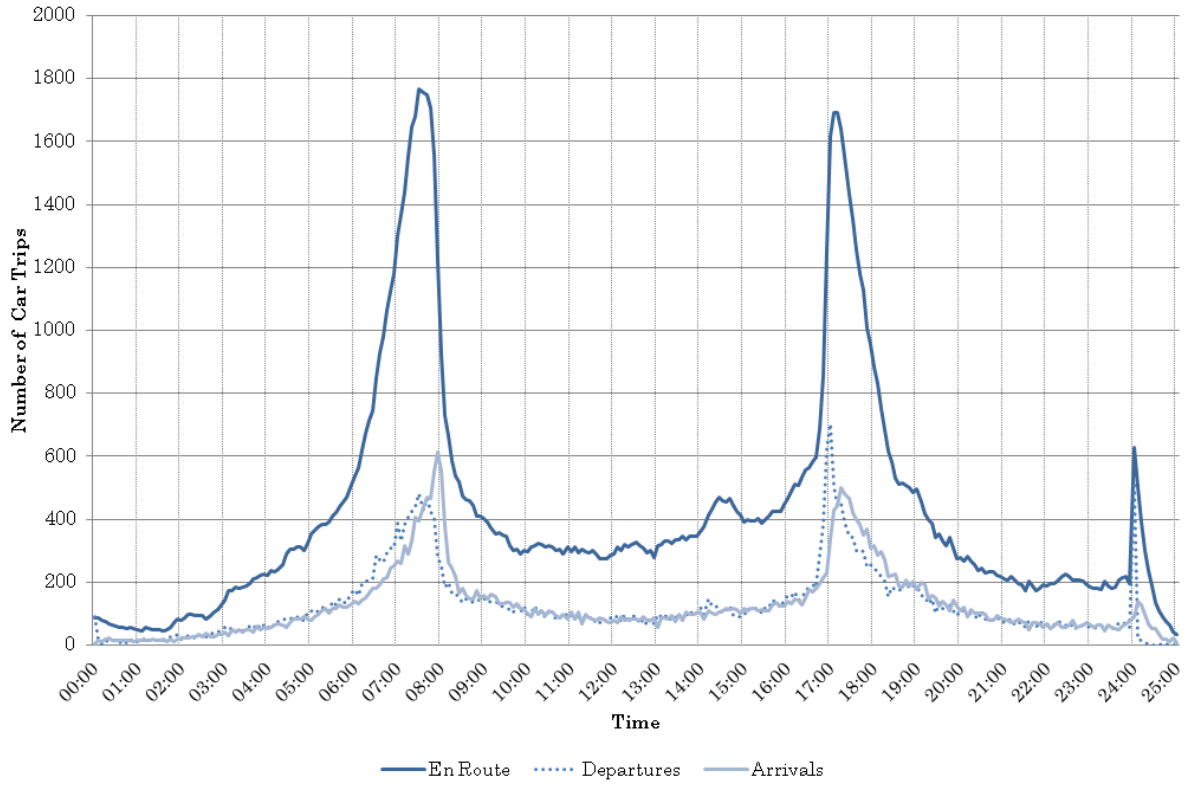
We find the slope of the line to be 0.53 with a y-intercept of 68 seconds and a square error ( $R^2$ ) of 0.25. The higher the  $R^2$  goodness of fit, the more realistic the planned leg times. Although the linear fit is arguably much worse than we would like, it does reveal some interesting insights. The slope of 0.53 suggests that for short legs we under-estimate the travel time, resulting in simulated travel times being higher than planned. This may indeed be plausible since the

impact of decelerating to a stop, and accelerating from the stop have more impact on shorter legs. Another possible explanation is that stops are usually closer together resulting in short inter-stop leg distance, in areas with higher demand. Such areas may also represent higher traffic densities and, as a result, more congested travel times.

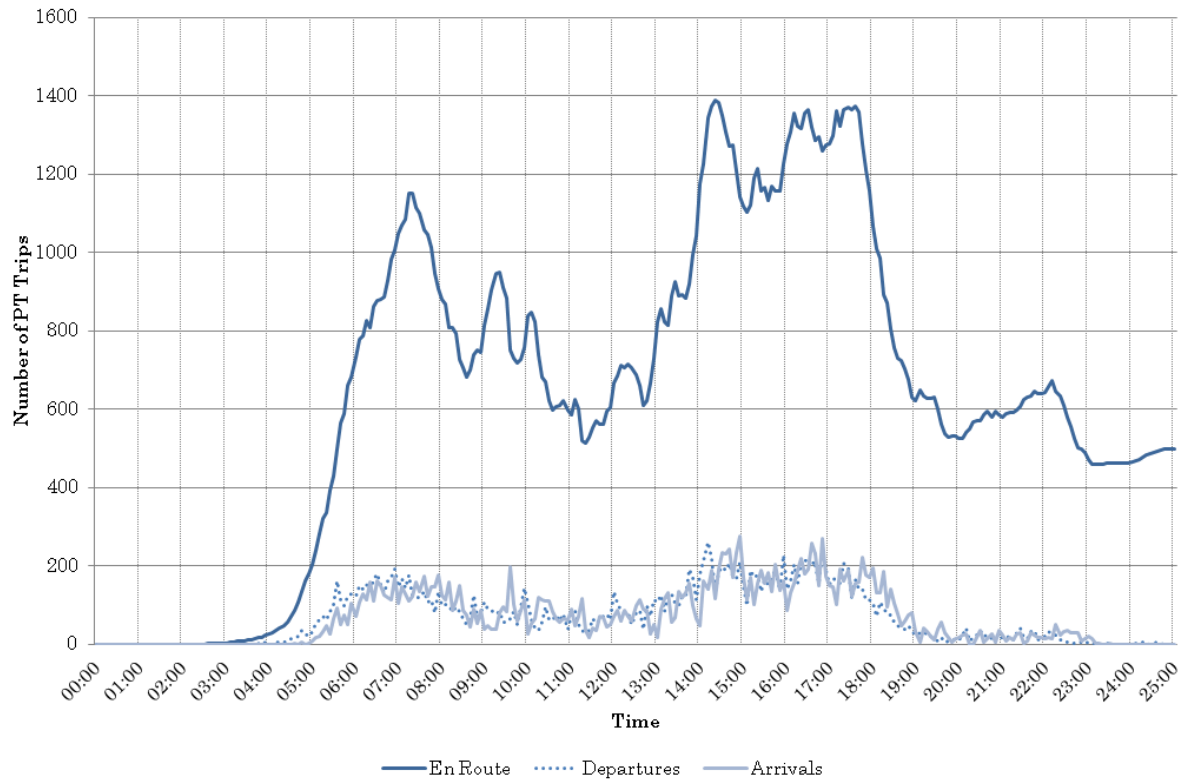
Conversely, the lower-than-one slope suggests we over-estimate the travel times on longer legs. Given the urban form, this too may indeed be plausible. Many commuters using bus transit services live in peripheral townships, many of which are connected to the urban core via higher order roads with speed limits in excess of the assumed travel speed in denser urban areas. As a result, buses in reality may travel quite a bit faster on these longer legs.

For decision-makers having to plan routes and schedules this is a valuable tool. It lends itself to prepare more accurate schedules which, in turn, may result in commuters experiencing the service offering as more reliable. The objective of using the linear fit is twofold. Firstly, one would aim to get the slope closer to unity, and the intercept closer to zero. That would mean that overall the simulated leg times are closer to the planned times. Secondly, but as importantly, one must reduce the spread of the points around the fitted line. Even if the linear fit results in a line  $y = 1x + 0$ , a large spread means that for every over-estimated leg, there is one (or more) underestimated leg. The  $R^2$  metric is useful to evaluate the spread, and the aim would be to increase it to be much closer to 1.0. It would be unrealistic to expect it to be 1.0 since traffic conditions and vehicle speeds are simply too uncertain.

The leg travel times reported in Figure 4.2 treat each leg independently. In reality though, a delay on one leg will have a knock-on effect on subsequent arrival times on the route. To analyse this, one can map the simulated arrival times at each stop against the planned arrival times in what is often called a space-time diagram (See Figure 4.3). In the first example (Figure 4.3a) we depict route 748 where the simulated times are very close to the planned times. Even though the route started a few minutes late at the first stop (6548), it was indeed able to catch up and end less than a minute early at the last stop (1051). The second example, in Figure 4.3b, however, shows the knock-on effect quite clearly. Route 1502 departed on time from stop 6822, setting off at 07h35 (during the morning rush hour), but soon became severely delayed as it progressed along its route, arriving at its final stop (5326) at 08h58, more than 35 minutes late. For this route, on legs with a planned travel time of less than 2 minutes the simulation averaged 2.5 times the planned values. This result supports the under-estimation of planned travel times for short legs observed in Figure 4.2, likely enhanced by the prevailing congestion during the morning rush hour.



(a) Car trips



(b) Transit trips

Figure 4.1: Number of trips throughout the day for different modes

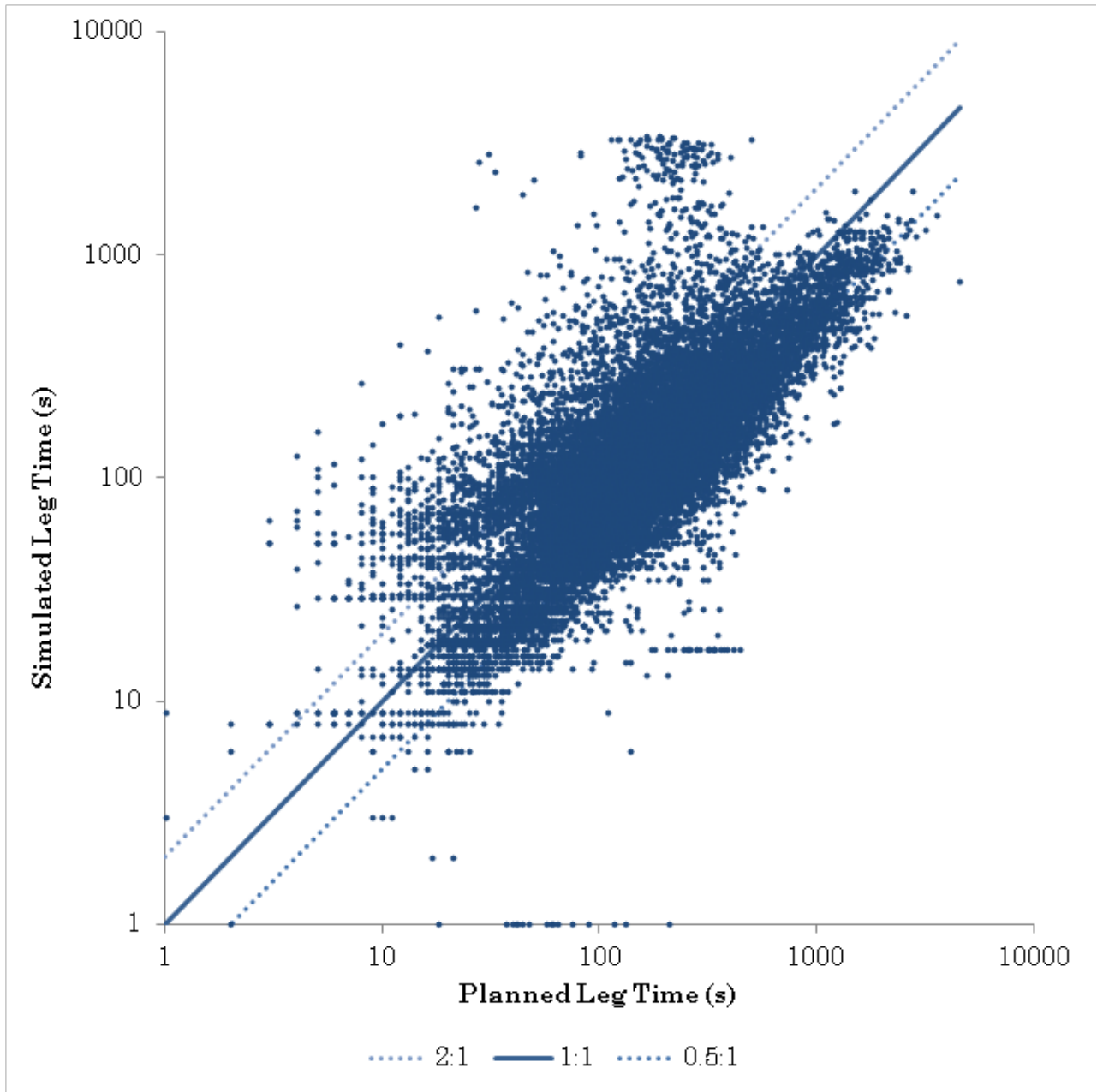
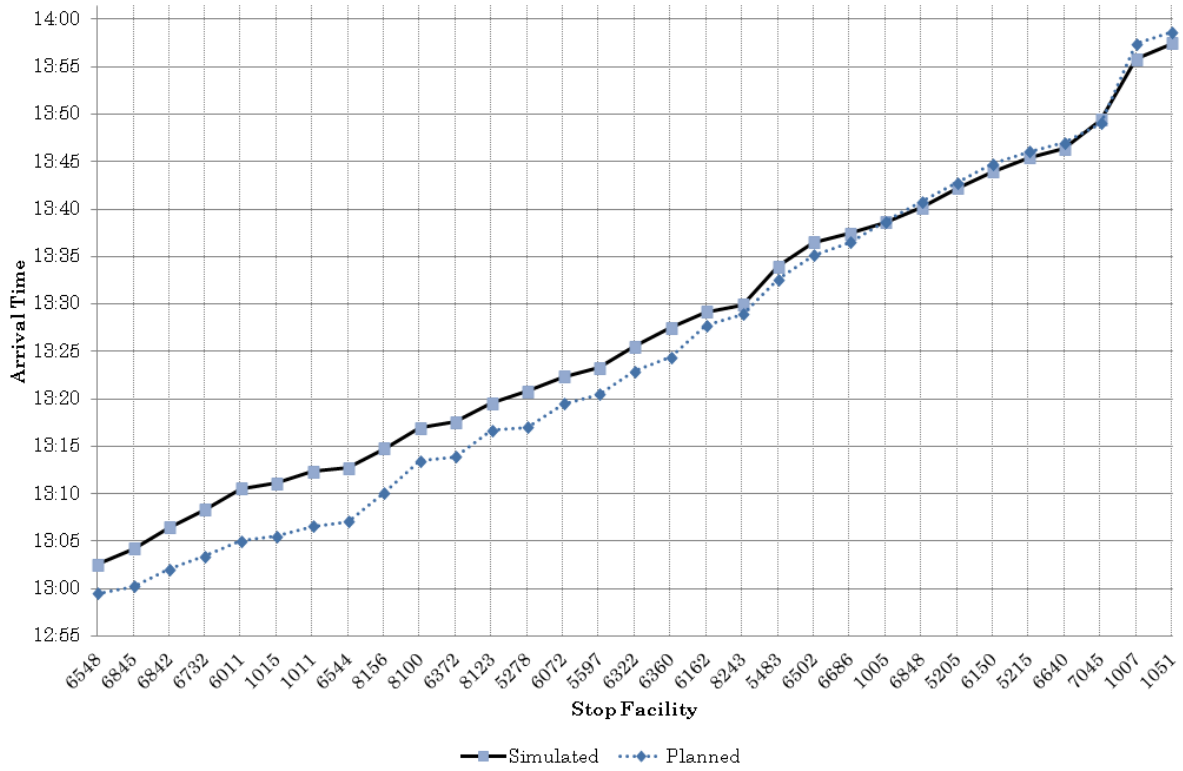
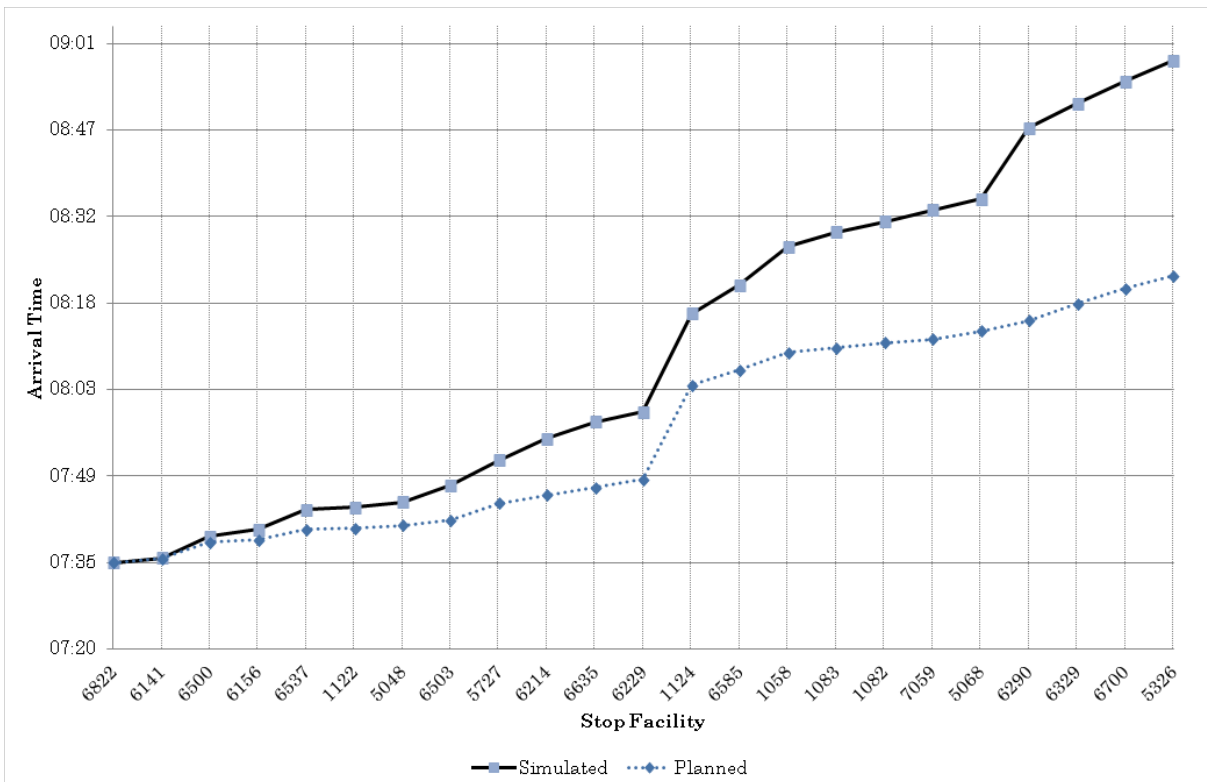


Figure 4.2: Comparison of planned and simulated route leg times



(a) Route 748



(b) Route 1502

Figure 4.3: Comparison of simulated and planned stop facility arrival times



### 4.3 Detailed Results for a Single Route

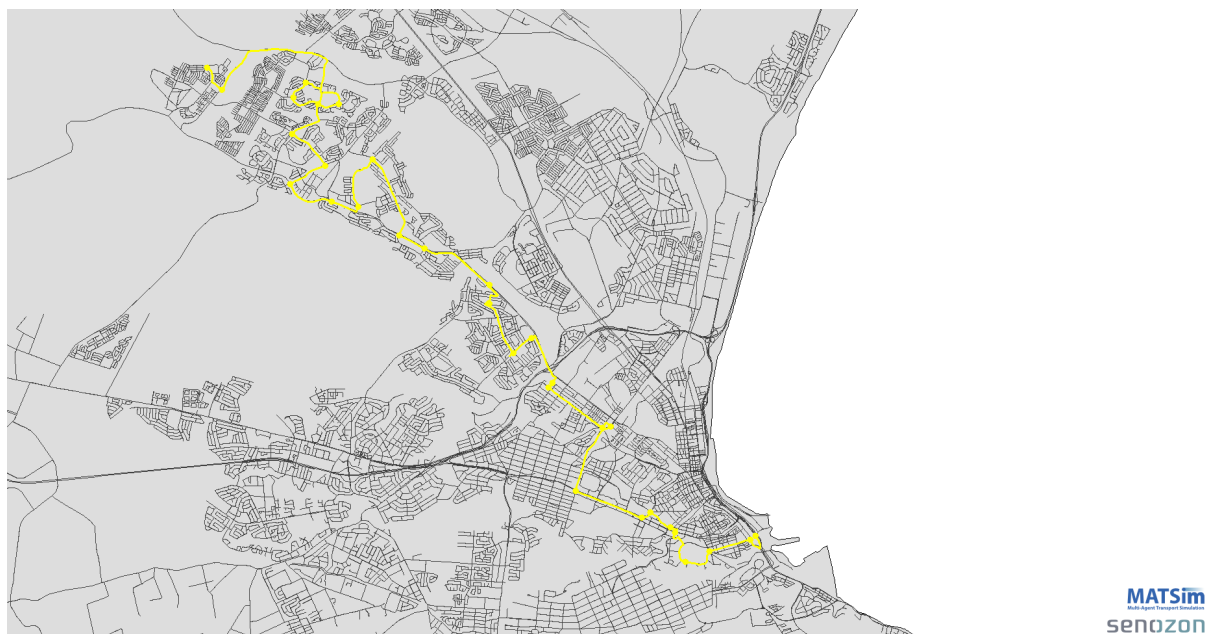


Figure 4.4: Algoa Bus Company route 748

In addition to comparing the correlation between transit data and actual travel times, decision makers may be interested in using agent-based simulation to extract information related to their services. The use of events, which capture every change in system state, permits detailed and disaggregate analyses. The analyses one may wish to perform, however, depend heavily on the research focus and decision support required. To illustrate the potential of this approach, a single route will be investigated thoroughly.

Algoa Bus Company's route 748, part of line 21, spans 35km through central Port Elizabeth servicing 31 stops along the way. It starts at MSB Station in the South-East and ends at Coral Road in the North West (See Figure 4.4). The simulated versus planned arrival times, described in Table 4.2, are discussed in Section 4.2.1.

Beyond providing an accurate schedule, transit operators must ensure that sufficient vehicle capacity is available to meet passenger demand, whilst simultaneously minimising operating costs. Figure 4.5 describes route 748's observed vehicle occupancy. The first 11 stops are hallmarked by an occupancy of less than 15%. A somewhat lower initial occupancy is expected for a transit service as passengers enter at various points along the route, but this may also be indicative of sub-optimal route configuration. Occupancy rapidly increases at stops 5278 and 5597, reaching the capacity limit of 60 agents at stop 5843. Additional agents wanting to board are unable to do so, and now have to wait for the next available service. This simulation constraint may contradict the real world behaviour where buses are often loaded beyond their capacity. Remaining high for the next 7 stops, occupancy dropped sharply at stop 6640 where 38 agents disembarked. All agents have disembarked by stop 1007, even before the bus has completed its route. Algoa Bus Company will aim to achieve high occupancy levels throughout their routes, using their limited resources efficiently. In this case, it may consider shortening the route and increasing frequency to improve initial occupancy levels and prevent capacity shortages at later stops.

Stop #	Stop Name	Stop ID	Arrival Time		Difference (s)	Early / Delayed
			Planned	Simulated		
0	MSB Station	6548	13:00:00	13:03:01	181	Delayed
1	Whites Road	6845	13:00:45	13:04:45	240	Delayed
2	Western Road	6842	13:02:33	13:06:59	266	Delayed
3	St Georges Hospital	6732	13:03:57	13:08:53	296	Delayed
4	5 Ways	6011	13:05:32	13:11:07	335	Delayed
5	Eastbourne	1015	13:06:03	13:11:40	337	Delayed
6	Gipson	1011	13:07:08	13:12:56	348	Delayed
7	Mount Road	6544	13:07:36	13:13:20	344	Delayed
8	Cape Road / Langenhoven	8156	13:10:39	13:15:21	282	Delayed
9	Mercantile	8100	13:14:04	13:17:34	210	Delayed
10	Highfield Road	6372	13:14:28	13:18:11	223	Delayed
11	Schauderville	8123	13:17:20	13:20:13	173	Delayed
12	Ablett Street	5278	13:17:36	13:21:25	229	Delayed
13	Beetlestone Road	6072	13:20:07	13:22:59	172	Delayed
14	Kobus Road	5597	13:21:07	13:23:56	169	Delayed
15	Gelvandale	6322	13:23:35	13:26:12	157	Delayed
16	Hartebees Street	6360	13:25:06	13:28:12	186	Delayed
17	Cleary Park	6162	13:28:26	13:29:55	89	Delayed
18	Standford / Rensburg	8243	13:29:40	13:30:38	58	Delayed
19	Arcadia Crossing	5483	13:33:21	13:34:41	80	Delayed
20	Marock Road	6502	13:35:53	13:37:16	83	Delayed
21	Salsoneville	6686	13:37:15	13:38:13	58	Delayed
22	24 Crossing	1005	13:39:26	13:39:23	-3	Early
23	William Slammert Drive	6848	13:41:35	13:40:58	-37	Early
24	21 Crossing	5205	13:43:34	13:43:00	-34	Early
25	Chatty 5	6150	13:45:36	13:44:45	-51	Early
26	Barberry Drive	5215	13:46:55	13:46:18	-37	Early
27	Pincushion Avenue	6640	13:47:50	13:47:14	-36	Early
28	Butterlily Avenue	7045	13:49:53	13:50:18	25	Delayed
29	Nooitgedacht	1007	13:58:20	13:56:45	-95	Early
30	Coral Road	1051	13:59:35	13:58:26	-69	Early

Table 4.2: Route 748 details

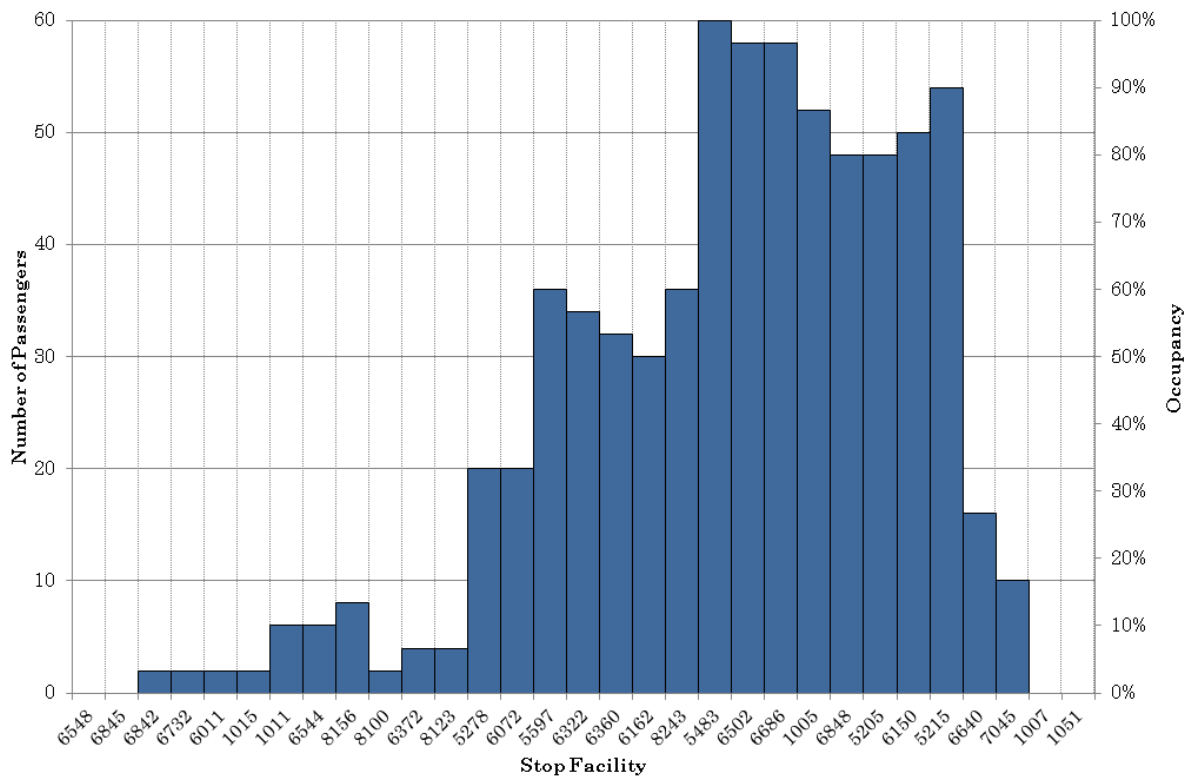


Figure 4.5: Vehicle occupancy for route 748

Before considering changes to a route, decision makers may wish to investigate the demand for a service. Figure 4.6 shows the activity locations of agents using line 21 (of which route 748 forms part) at some point in their day. This may be for them to get to work, get home from school or connect with another transit service. Figure 4.6c shows the schools and universities attended by agents using this line. This is a very important service for scholars, who represent 68% of the total demand. Given the route's proximity to schools, this is to be expected. People going to work make up a further 23% of demand, with shopping and leisure making up the rest of the activity demand.

Agents using this line had an average of 6 transit interaction activities during their day (see Figure 4.6d) - that is either using a transit service or transferring between services. This high number shows line 21 to be a very important connecting line. Agents from dispersed home locations (see Figure 4.6a) use feeder routes in combination with line 21, mostly for short distances, to complete their journey. Lines of this type need frequent, predictable service to prevent passenger delays and missed connections.

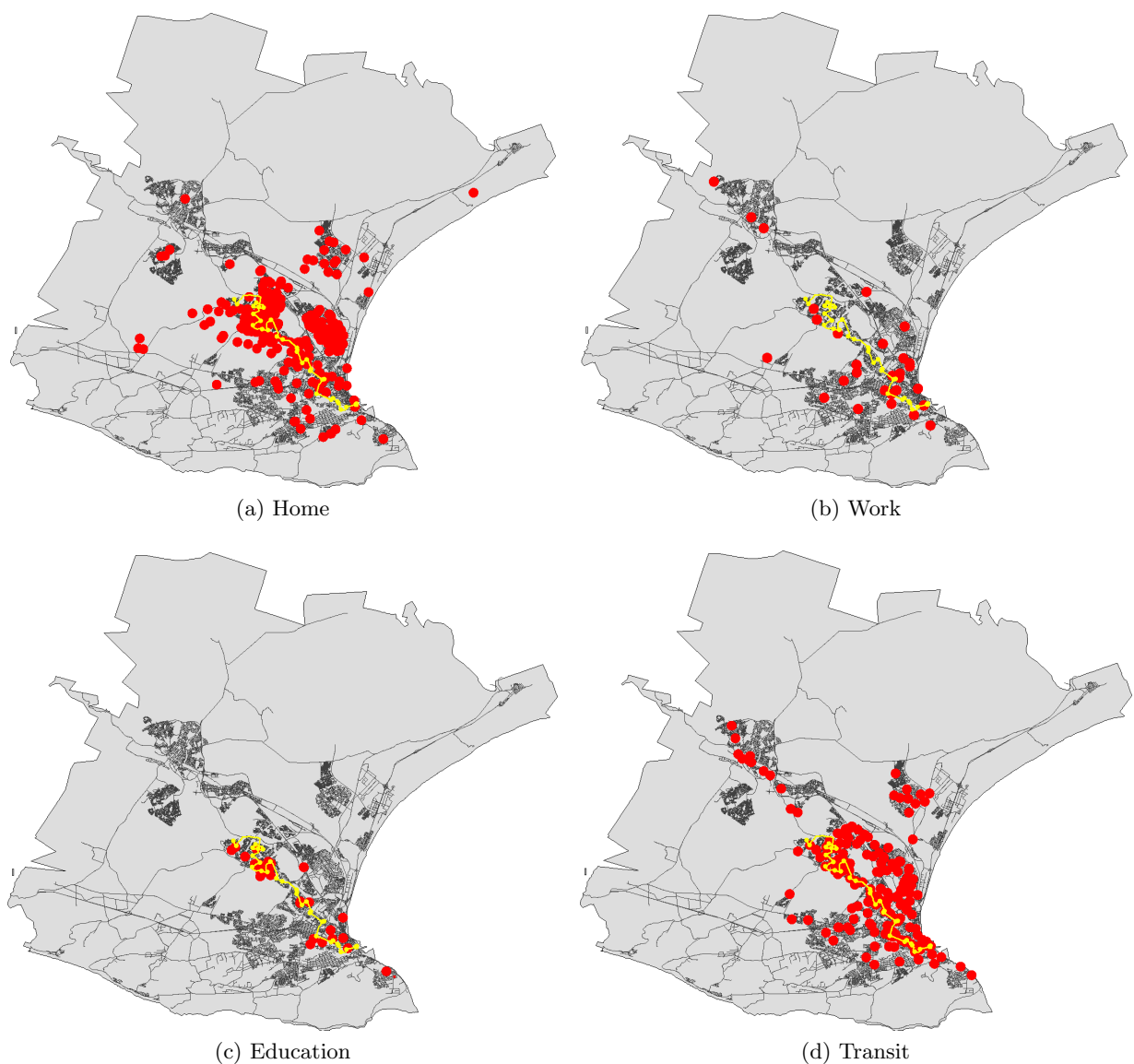


Figure 4.6: Activity locations for agents using line 21

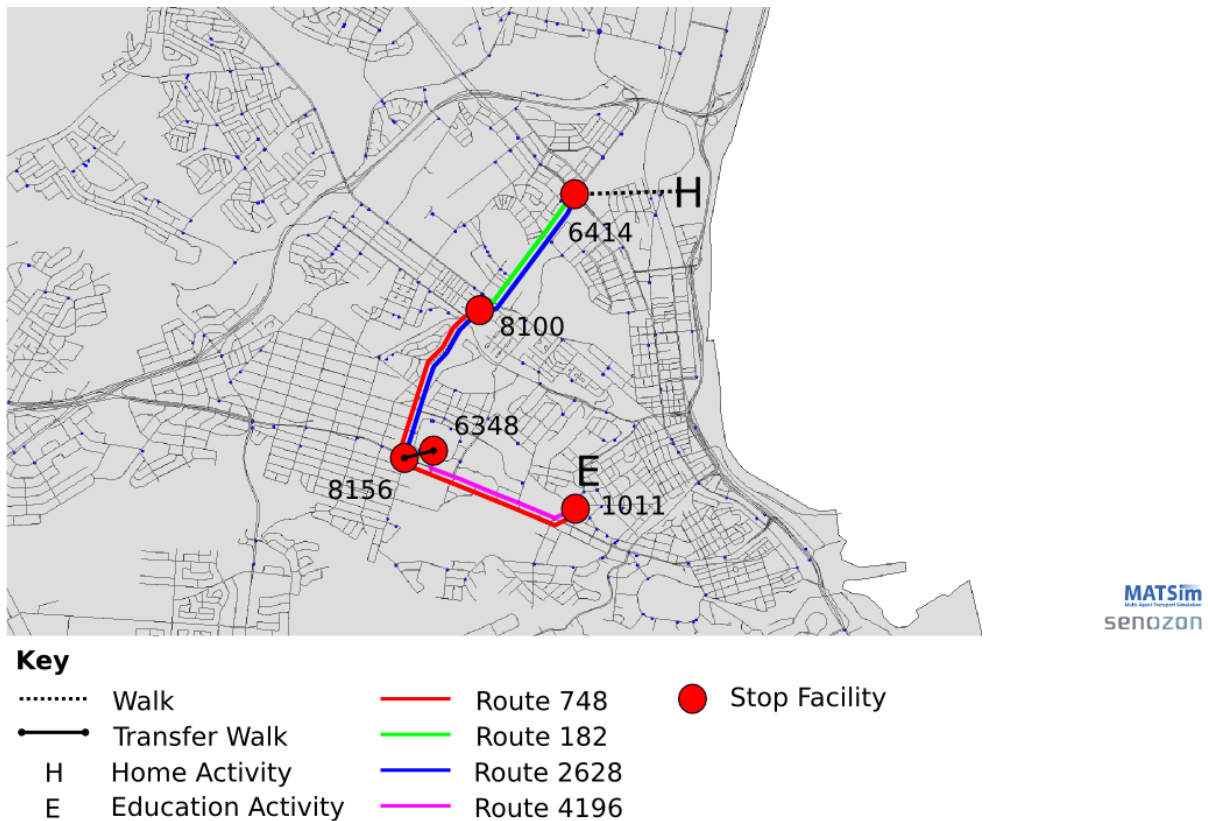


Figure 4.7: Planned activities for a single agent

Figure 4.6 shows the activity locations of agents using line 21. It is difficult, however, to fully understand demand without also looking at the sequence of activities, or activity chain, making up a transit user's demand. To achieve this, one can investigate the plan of a single agent using the line. Agent 693913 is a 14 year old male who wishes to use route 748 on his way home from school. His full plan is illustrated in Figure 4.7 and set out below:

- Walk from home to stop 6414 at 06h30, arriving at 07h10;
- Travel on bus **line 70, route 2628** to stop 8156, arriving at 07h27;
- Walk to stop 6348, arriving at 07h36;
- Travel on bus **line 100, route 4196** to stop 1011, arriving at 07h42;
- Walk to school, arriving at 07h49 and finishing at 12h55;
- Walk to stop 1011, arriving at 13h00;
- Travel on bus **line 21, route 748** to stop 8100, arriving at 13h08;
- Travel on bus **line 4, route 182** to stop 6414, arriving at 13h14;
- Walk home, arriving at 13h55.

In the simulation, as in life, things didn't go exactly according to plan. As shown in Figure 4.8, the agent:

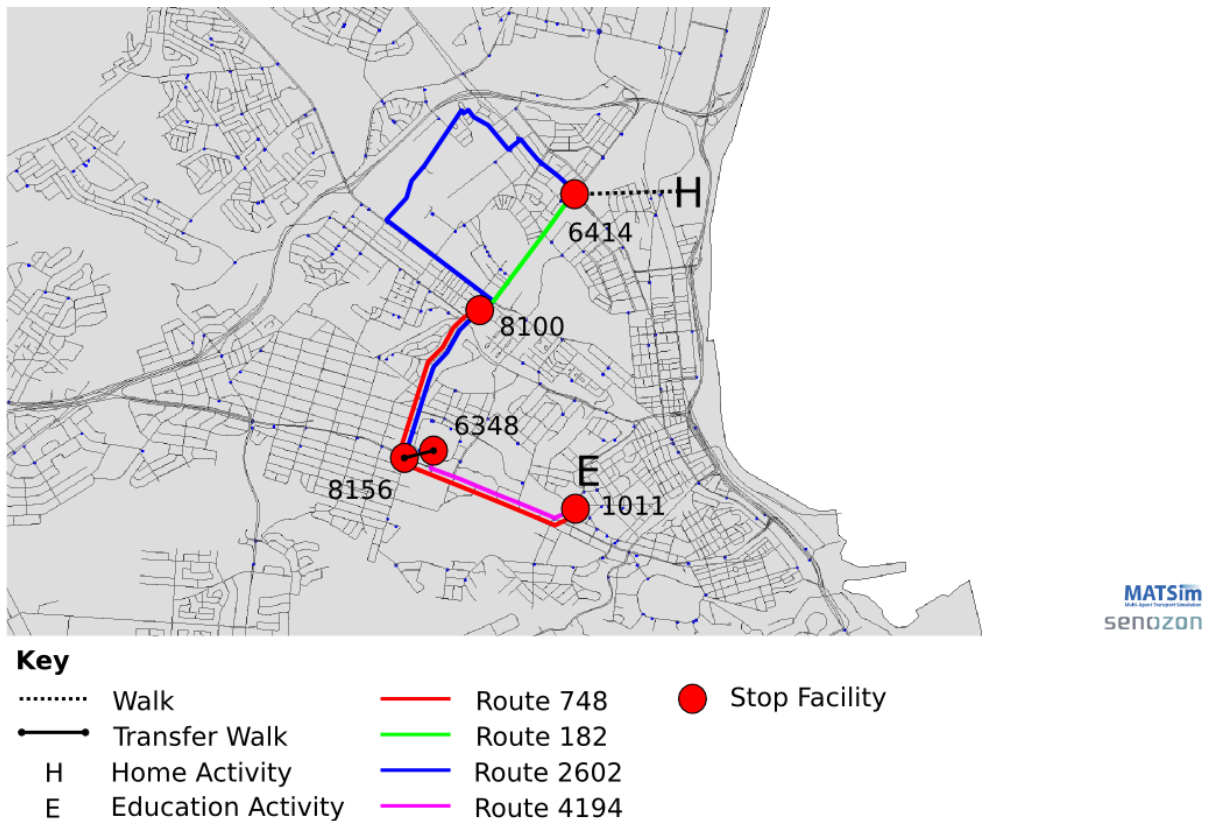


Figure 4.8: Simulated activities for a single agent

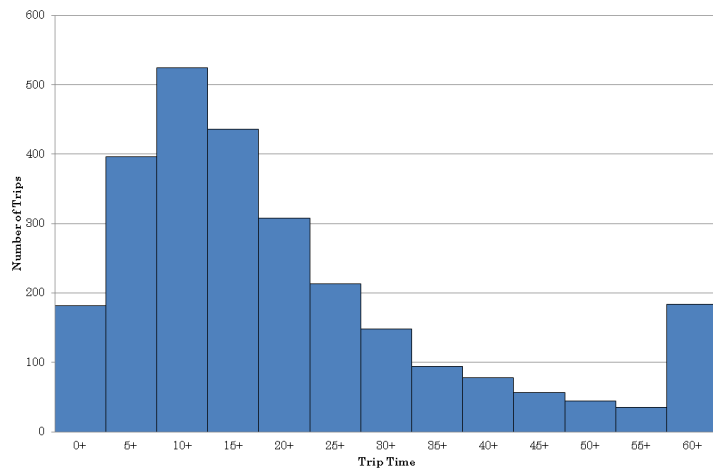
- Walked from home to stop 6414 at 06h22, arriving at 07h03;
- Travelled on bus **line 70, route 2602** to stop 8156, arriving at 07h58. The bus he planned to take (*route 2628*) was severely delayed, therefore he had to wait for another bus of the same line driving to the same stop;
- Walked to stop 6348, arriving at 08h07;
- Travelled on bus **line 100, route 4194** to stop 1011, arriving at 08h29. The bus he planned to take (*route 4196*) had already departed, therefore he had to use another bus of the same line driving to the same stop;
- Walked to school, arriving (very late) at 08h36 and finishing at 12h55;
- Walked to stop 1011, arriving at 13h01;
- Travelled on bus **line 21, route 748** to stop 8100, arriving at 13h17;
- Travelled on bus **line 4, route 182** to stop 6414, arriving at 13h20;
- Walked home, arriving at 14h01.

Analyses at this level of detail may seem trivial at first glance, but they give insight into the everyday transit system which may be lacking from higher level analyses. Ultimately, the goal is to get *individuals* to where they want to go as quickly and easily as possible. Using disaggregate simulation tools, such as MATSim, we can see the impact of system decisions on the individual.

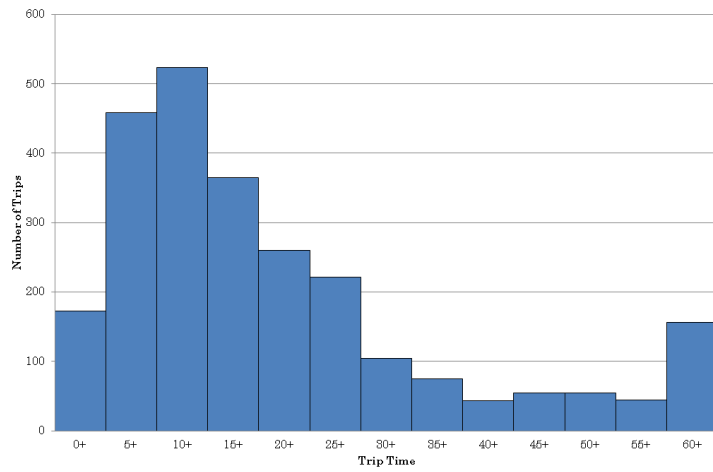
## 4.4 Transit Accessibility

Understanding transit service accessibility is key for decision makers. Transit services need to be available both spatially and temporally at both the point of origin and the destination for transit to be considered. A transit decision maker may also set certain strategic objectives for accessibility. In South Africa, the National Land Transport Strategic Framework (Department of Transport, 2006) sets a target of access to public transport services within 1 kilometer (about a 15 minute walk) for urban and metropolitan households.

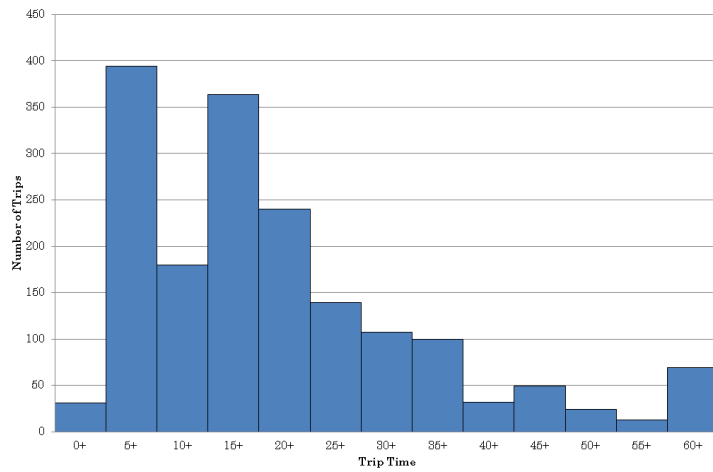
Figure 4.9 explores the time needed to start an activity after alighting a transit vehicle, i.e. walking time from the stop facility to the activity location. It is based on data for transit users in the simulation. Figure 4.9a shows that 57% of transit users are able to acquire transit from their home within the target time of 15 minutes. This is slightly better than the national average of 48% as reported in the National Household Travel Survey (NHTS). 60% of transit agents could access their work locations within 15 minutes of alighting (Figure 4.9b), while only 56% of transit agents could access education locations within this time (Figure 4.9c).



(a) Home



(b) Work



(c) Education

Figure 4.9: Walking time to an activity location after alighting a transit vehicle



# Chapter 5

## Conclusion

This research project implemented a large-scale multi-modal MATSim model for the Nelson Mandela Bay Metropole. The main data preparation involved matching GTFS data for the study area to the MATSim network, creating a transit schedule suitable for simulation purposes.

### 5.1 Brief Summary of Findings

A 10% sample population was used for the simulation. Figure 4.2 compares the planned and actual (simulated) leg, or inter stop, travel times in the form of a scatterplot. Applying a linear fit to the data produced a slope value of 0.53 and y-intercept of 68 seconds, indicating that for short legs travel time is under-estimated by transit agencies, resulting in simulated travel times being higher than planned, while for longer legs, travel time is over-estimated by transit agencies, resulting in simulated travel times being shorter than planned. The  $R^2$  goodness of fit value of 0.25 is much lower than we would like. These values are expected to improve as transit data accuracy and model calibration are refined.

Several analyses are undertaken which attempt to shed light on components of the transit system. Investigating arrival times, occupancy levels and demand at line level illustrate the use and applicability of disaggregate analyses for transit decision makers. In addition, measuring service accessibility offers insight into the state of the overall transit system performance which may be hard to obtain through other means.

### 5.2 Looking to the Future

This project, as the first large-scale agent-based transit simulation in South Africa, represents the first step in a process towards providing models which can accurately capture the intricacies of our public transport environment. At present, the relative unavailability of accurate and complete data presents a major obstacle to transit modelling in South Africa. As shown in Figure 3.6, this data is often not available to the transit operators themselves. It is hoped that projects such as this one may encourage these operators and decision makers to invest in and benefit from agent-based simulations. Several research projects are currently under way which aim to model minibus taxis - only once this integral part of our system is included will agent-based transit simulation be considered as truly representative, but even now it affords the opportunity to gain unparalleled insight into current *and* future transit systems.

Looking further afield, we discover that agent-based models allow you to answer questions related to dynamic mobility markets with a high level of detail. As an example we may consider Figure 5.1 which identifies the busiest stop facilities for the study area. This information may be valuable for determining optimal advertisement placement, especially given that we can distinguish between different customer groups down to individual level.



Figure 5.1: Relative number of passengers entering a stop facility

# Bibliography

- Agarwala, A. and Zilske, M. (2012). Person-based dynamic traffic assignment for mixed traffic conditions. VSP working paper 08-01, TU Berlin, Transport Systems Planning and Transport Telematics, see [www.vsp.tu-berlin.de/publications](http://www.vsp.tu-berlin.de/publications).
- Balmer, M. (2007). *Travel demand modeling for multi-agent transport simulations: Algorithms and systems*. PhD thesis, Swiss Federal Institute of Technology (ETH).
- Department of Transport (2006). National Land Transport Strategic Framework. *Government Gazette*.
- Flotterod, G., Chen, Y., Rieser, M., and Nagel, K. (2009). Behavioral calibration of a large-scale travel behavior microsimulation. In *Proceedings of the meeting of the International Association for Travel Behavior Research*, Jaipur, India. International Association for Travel Behavior Research.
- Fourie, P. J. (2009). An initial implementation of a multi-agent transport simulation for South Africa. Master's thesis, University of Pretoria.
- Joubert, J. W. (2012). Paratransit: perpetual pain or potent potential? *Draft paper*.
- Joubert, J. W., Fourie, P. J., and Axhausen, K. W. (2010). Large-scale agent-based combined traffic simulation of private cars and commercial vehicles. *Transportation Research Record*, 2168:24–32.
- Kizoom, N. and Miller, P. (2008). A transmodel based xml schema for the Google Transit Feed Specification. *Kizoom Ltd*.
- MATSim (2012). Matsim. URL [www.matsim.org](http://www.matsim.org).
- OpenStreetMap Development Team (2012). Open Street Map, The Free Wiki World Map., URL [www.openstreetmap.org](http://www.openstreetmap.org).
- Ordonez, S. and Erath, A. (2011). Semi-automatic tool for map-matching bus routes on high-resolution navigation networks. 16th International Conference of Hong Kong Society for Transportation Studies, Hong Kong.
- Rieser, M. (2010). *Adding Transit to an Agent-Based Transportation Simulation*. PhD thesis, Technischen Universitat Berlin.
- Van der Merwe, J. (2011). Agent-based transport demand modelling for the South African commuter environment. Master's thesis, University of Pretoria.
- Wevell, D. (2011). Implementing MATSim transit simulation in a South African context.

# Appendix A

## Transit Schedule DTD

The following document type definition (DTD) specifies the XML grammar used to store transit schedules in MATSim.

```
<?xml version="1.0" encoding="utf-8"?>

<!ELEMENT transitSchedule (transitStops?,transitLine*)>

<!ELEMENT transitStops (stopFacility)*>

<!ELEMENT stopFacility EMPTY>
<!ATTLIST stopFacility
id          CDATA #REQUIRED
x           CDATA #REQUIRED
y           CDATA #REQUIRED
linkRefId   CDATA #IMPLIED
name        CDATA #IMPLIED
isBlocking  (true|false) "false">

<!ELEMENT transitLine (transitRoute)*>
<!ATTLIST transitLine
id          CDATA #REQUIRED>

<!ELEMENT transitRoute (description?,transportMode,routeProfile,route?,departures)>
<!ATTLIST transitRoute
id          CDATA #REQUIRED>
<!-- transitRoute.id must be unique within a transitLine only. -->

<!ELEMENT description (#PCDATA)>

<!ELEMENT transportMode (#PCDATA)>
<!-- describes the mode of which vehicles are that serve that line/route -->

<!ELEMENT routeProfile (stop)*>

<!ELEMENT stop EMPTY>
```

```

<!ATTLIST stop
refId          CDATA    #REQUIRED
departureOffset CDATA    #IMPLIED
arrivalOffset  CDATA    #IMPLIED
awaitDeparture (true|false) "false">
<!--
* stop.departureOffset is required for all stops but the last.
The offsets are to be added to the departure times of the single
departures listed in the transitRoute.
* stop.refId are id-references to facilities. -->

<!ELEMENT route          (link)*>

<!ELEMENT link           EMPTY>
<!ATTLIST link
refId          CDATA    #REQUIRED>

<!ELEMENT departures     (departure)*>
<!-- the single departures along that transitRoute -->

<!ELEMENT departure      EMPTY>
<!ATTLIST departure
id             CDATA    #REQUIRED
departureTime  CDATA    #REQUIRED
vehicleRefId   CDATA    #IMPLIED>
<!-- the departure.id must be unique within a transitLine only -->

```

## Appendix B

# The General Transit Feed Specification (GTFS)

Table B.1: The GTFS “feed” files and associated content.

Filename	Required	Description
agency.txt	Required	One or more transit agencies that provide the data in this feed.
stops.txt	Required	Individual locations where vehicles pick up or drop off passengers.
routes.txt	Required	Transit routes. A route is a group of trips that are displayed to riders as a single service.
trips.txt	Required	Trips for each route. A trip is a sequence of two or more stops that occurs at specific time.
stop-times.txt	Required	Times that a vehicle arrives at and departs from individual stops for each trip.
calendar.txt	Required	Dates for service IDs using a weekly schedule. Specify when service starts and ends, as well as days of the week where service is available.
calendar-dates.txt	Optional	Exceptions for the service IDs defined in the calendar.txt file. If calendar-dates.txt includes ALL dates of service, this file may be specified instead of calendar.txt.
fare-attributes.txt	Optional	Fare information for a transit organization’s routes.
fare-rules.txt	Optional	Rules for applying fare information for a transit organization’s routes.
shapes.txt	Optional	Rules for drawing lines on a map to represent a transit organization’s routes.
frequencies.txt	Optional	Headway (time between trips) for routes with variable frequency of service.
transfers.txt	Optional	Rules for making connections at transfer points between routes.
feed-info.txt	Optional	Additional information about the feed itself, including publisher, version, and expiration information.