An initial implementation of a multi-agent transport simulator for South Africa

Pieter Jacobus Fourie

A dissertation submitted in partial fulfilment of the requirements for the degree

Master of Engineering (Industrial Engineering)

in the

Faculty of Engineering, Built Environment, and Information Technology

University of Pretoria
Pretoria, South Africa

June 2009

Abstract

AN INITIAL IMPLEMENTATION OF A MULTI-AGENT TRANSPORT SIMULATOR FOR SOUTH AFRICA

PIETER JACOBUS FOURIE

Supervisor:	Dr. Johan W. Joubert
Department:	Industrial and Systems Engineering
University:	University of Pretoria
Degree:	Master of Engineering (Industrial Engineering)

Transport demand planning in South Africa is a neglected field of study, using obsolete methods to model an extremely complex, dynamic system composed of an eclectic mix of First and Third World transport technologies, infrastructure and economic participants.

We identify Agent-Based Simulation (ABS) as the only modelling paradigm capable of capturing the effects emerging from the complex interactions within the South African transport system, and proceed to implement the Multi-Agent Transport Simulation Toolkit (MATSim) for South Africa's economically important Gauteng province. This report describes the procedure followed to transform household travel survey, census and Geographic Information System (GIS) data into an activity-based transport demand description, executed on network graphs derived from GIS shape files.

We investigate the influence of network resolution on solution quality and simulation time, by preparing a full network representation and a small version, containing no street-level links. Then we compare the accuracy of our data-derived transport demand with a lower bound solution. Finally the simulation is tested for repeatability and convergence.

Comparisons of simulated versus actual traffic counts on important road network links during the morning and afternoon rush hour peaks show a minimum mean relative error of less than 40%. Using the same metric, the small network differs from the full representation by a maximum of 2% during the morning peak hour, but the full network requires three times as much memory to execute, and takes 5.2 times longer to perform a single iteration.

Our census- and travel survey-derived demand performs significantly better than uniformly distributed random pairings of home- and work locations, which we took to be analogous to a lower bound solution. The smallest difference in corresponding mean relative error between the two cases comes to more than 50%.

We introduce a new counts ratio error metric that removes the bias present in traditional counts comparison error metrics. The new metric shows that the spread (standard deviation) of counts comparison values for the random demand is twice to three times as large as that of our reference case.

The simulation proves highly repeatable for different seed values of the pseudo-random number generator. An extended simulation run reveals that full systematic relaxation requires 400 iterations. Departure time histograms show how agents 'learn' to gradually load the network while still complying with activity constraints.

The initial implementation has already sparked further research. Current priorities are improving activity assignment, incorporating commercial traffic and public transport, and the development and implementation of the minibus taxi para-transit mode.

Keywords

MATSim; transport demand planning; agent-based transport simulation; transport microsimulation; traffic simulation; transportation demand modeling; transportation planning; dynamic traffic assignment; travel demand modeling; activity-based analysis

Acknowledgements

I would like to express my sincerest thanks and appreciation to the following persons:

- My supervisor, Dr. Johan W. Joubert, for his excellent and unfailing guidance and support. I am grateful for his selfless offerings of time, effort and insight, and consider him a formidable ally and mentor in my research career.
- Dr. Michael Balmer, Institute for Traffic Planning and Transport Systems, ETH Zürich, for his review of this dissertation and valuable feedback.
- Prof. Dr. Kai Nagel and his team from Transport Systems Planning and Transport Telematics, Institute for Land and Sea Transport Systems, Technische Universität Berlin, where I was received for two months in the European summer to lay the groundwork for this study. Prof. Nagel and his entire team made the visit a most valuable and memorable experience, but I would like to thank Marcel Rieser and Gregor Lämmel in particular, for providing key insights, programming skill and technical support.
- The MATSim development teams for continuing technical support.
- Business Connexion GIS, for their GIS dataset of the Gauteng transport network.
- Mr. Piet Alberts from Statistics South Africa (Stats SA), who provided us with Census 2001 and National Household Travel Survey (NHTS) data.
- My family, for their unwavering love and support from California, England and South Africa.

This work is based upon research supported by the National Research Foundation (NRF) (grant FA2007051100019). The author also acknowledges financial aid from the University of Pretoria and the NRF Knowledge Interchange and Collaboration programme (UID 67539), which made possible the research visit to Berlin during June and July 2008.

Contents

List of Figures

List of Tables

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

List of Acronyms

ABA	Activity-Based Approach
ABS	Agent-Based Simulation
BRT	Bus Rapid Transit
DTA	Dynamic Traffic Assignment
EA	Evolutionary Algorithm
EAPSU	Enumeration Area Primary Sampling Unit
FSM	Four Step Model
GIS	Geographic Information System
ITS	Intelligent Transport Systems
MATSim	Multi-Agent Transport Simulation Toolkit
NHTS	National Household Travel Survey
SANRAL	South African National Roads Agency Ltd.
SP	Sub-Place
Stats SA	Statistics South Africa
TAZ	Traffic Analysis Zone
UTM	Universal Transverse Mercator
WGS84	World Geodetic System 1984

Chapter 1

Introduction

1.1 The need for transport demand planning

Humans live their lives by pursuing goals in the face of obstacles. Each goal can be broken down into sub-goals; these in turn can be broken down into even smaller sub-goals, and so forth. Ultimately, each goal translates into a series of activities that need to be performed, i.e. being somewhere at some time doing something.

It is not always possible to perform all activities in the same place. Whenever two activities occur in sequence in different places, the need for transport arises. It follows then that the demand for transport is derived from the needs of people to perform activities at certain times in particular places.

In the real world, there exist several obstacles to efficient movement between activity locations. These include:

Spatial obstacles The distance between points of interest and the topography of the landscape.

Time availability Not just one's own time, but also those of others, as their presence might be required to perform an activity.

Physical resource availability We require access to vehicles and proper surfaces to travel between points of interest. We are therefore constrained by our material wealth.

The needs of others We are surrounded by other people pursuing their goals. Having similar goals as others can work to our advantage; we increase efficiency by pooling our resources and synchronising our activities. But the needs of others can also work to our detriment; other people take up physical capacity of the transport network, which causes traffic congestion, thus affecting the speed and efficiency of movement.

Past decisions Consider the narrow and labyrinthine medieval roads of old European city centres, and their impact on modern vehicle traffic. The transport network infrastructure of one era is inherited by the next; these permanent features from the past might become an impediment for future generations.

Overcoming these obstacles requires rigorous planning on the part of the commuters who create the demand for transport, as well as the governing institutions who supply access to transport resources. On the commuter side, for instance, an individual can decide on a different schedule of activities, or she might decide to take a different route to work and avoid congestion. Commuters' plans are relatively short-term and translate into changes in their behaviour. On the institution's side, planning is done for the medium to long term, and plans translate into policy changes and network changes or expansions. For example, a policy change might be the introduction of a congestion charge, which could deter people from using private vehicle transport in a city centre. A network change, on the other hand, can have a similar effect. For example, improving public transport service levels in the city centre might encourage people to leave their cars behind. In both cases then, institutional planning requires anticipating the effect of interventions on the behaviour of the commuter population.

These examples suggest the difference in complexity and precision involved in the planning of the commuter versus that of the institution. The planning performed by the individual can take a trial-and-error form, as she should be able to absorb the financial impact of her actions. By contrast, the transport system provider's decisions can have a massive capital expenditure impact, as well as a long-term financial impact on, literally, *generations* of commuters.

The discipline of *transport demand planning* evolved to deal with this massive societal impact and to inform complex transport policy and infrastructure decisions in a precise, systematic and scientific manner. Transport demand planning recognises that the demand for transport is derived from people's need to perform activities at certain times in particular places. The discipline aims to achieve efficient, prompt and affordable movement of goods and individuals in a sustainable manner. It does so by anticipating the effect of interventions before implementation, i.e. by testing hypotheses on a representation, or model, of the transport system.

1.2 Transport demand planning technologies

Several transport demand models exist. Their predictive power has increased over the years due to cheaper, faster computing power and improved modelling of the individual decision-making and behaviour from which larger scale effects emerge. There is much resistance in industry to adopt newer technologies, however, as implementation is time-consuming and requires many hours of expert labour. Therefore, traditional modelling methods are still pervasive.

1.2.1 Traditional transport demand planning

The so-called Four Step Model (FSM), illustrated in Figure 1.1, forms the basis of most operational transport demand planning methods around the world. The model evolved during the 1960s in response to the United States' 1962 Federal-Aid Highway Act, which incentivised massive transport capacity expansions in urban areas during that time (the interested reader is referred to Weiner (1997) for an historical overview of the development of the FSM and its role in U.S. urban transportation planning).

Central to the modelling paradigm is the *trip* rather than the *activity*: residential areas are seen as trip producers, whereas centres of economic activity act as trip attractors. The FSM's basic operation is simple (Meyer and Miller, 2001). The study area is broken up into Traffic Analysis Zones (TAZs). These TAZs demarcate areas of similar population density, socio-economic status and land-use.

1. A variety of algorithms interpret land-use and demographic data to estimate the number of trips that each area will produce or attract.

2. The model assigns trips from producers to attractors, to predict the flow of trips from each TAZ to every other TAZ, such that all trip-producing origins and all trip-attracting destinations are accounted for.

3. Each trip flow is subdivided amongst the different transport modes available between two TAZs.

4. An iterative process then assigns each flow to a route through its modal network, re-routing flows after each iteration until the load on the network is balanced in such a way that no further re-routing will improve the expected travel time of any trip.

The FSM produces time-independent, average traffic flows on network links. Consider, however, the following scenario: commuter X gets caught in the traffic flow between some trip producer and trip attractor. She *decides* that she will drive to work half an hour later tomorrow because she got stuck in traffic too long on her way there today. If a number of commuters made a similar decision, a large-scale result would emerge that the FSM would be unable to predict — the *average* flow on that particular set of network links might remain the same, but the detailed time-dependent behaviour has changed dramatically. And this change is not due to some overarching mechanism, but rather emerges from the *decision-making and behaviour* of the system's constituent entities. Therefore, because of its lack of behavioural grounding, the FSM is becoming increasingly less useful in a context where we are interested in making the most of available network capacity by changing and predicting individuals' travel *behaviour*.

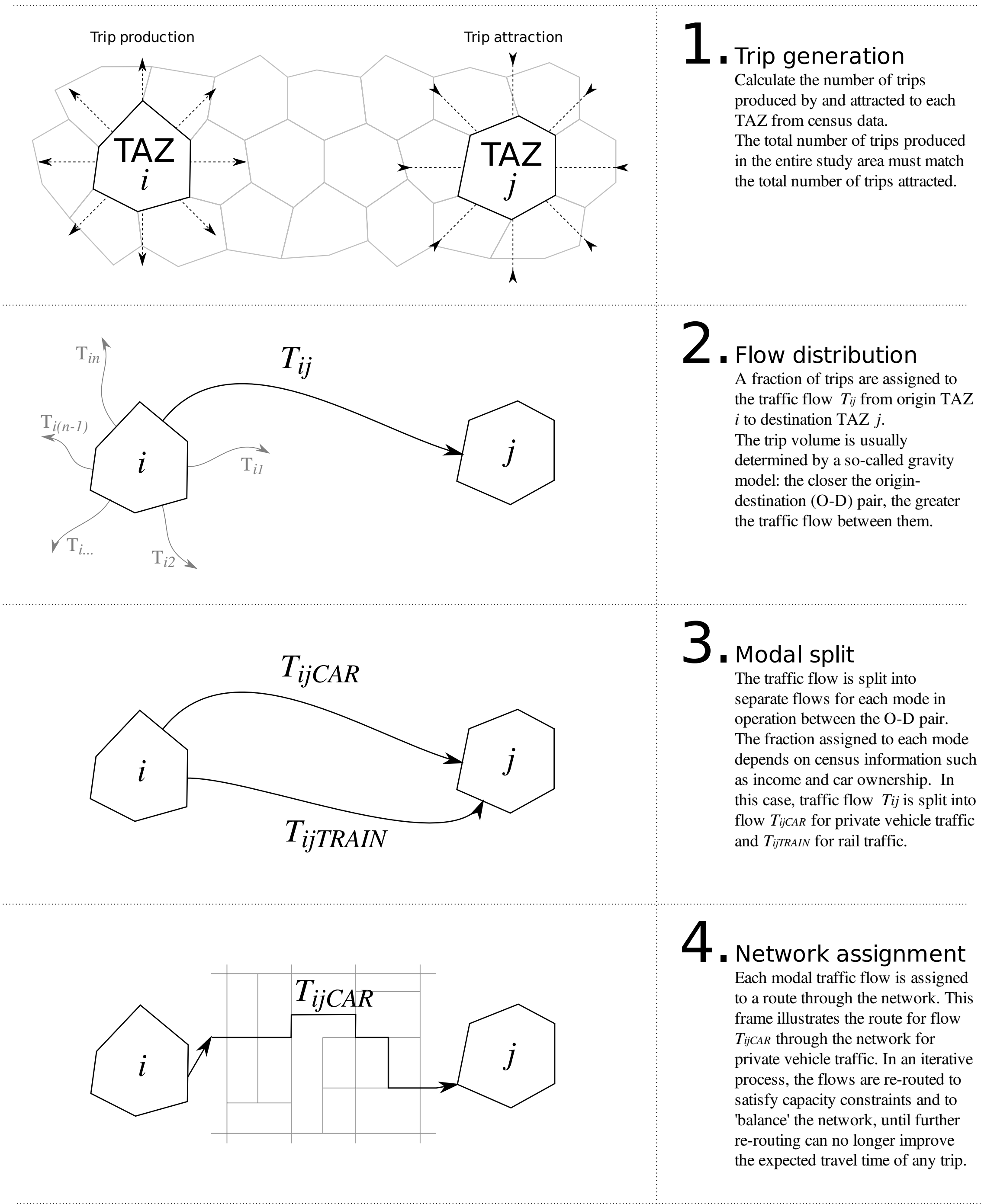


Figure 1.1: Schematic diagram illustrating the four steps involved in traditional transport demand planning. Adapted from Meyer and Miller (2001).

1.2.2 Modern transport demand planning: disaggregation and microsimulation

Modern technologies aim to more closely simulate the individual decision processes employed by commuters, and to shift the focus of transport demand planning back towards the actual activities that generate transport demand (McNally, 2000; Bowman, 1998; Ettema, 1996). These technologies examine the trade-offs individuals make when trying to gain utility from performing activities in certain places versus losing utility in trying to get to an activity location, or getting there too late or too early, or having had to pay a toll to get there, to name but a few examples.

In order to arrive at individuals and their activities, modern models disaggregate the available census data to arrive at a synthetic population. Based on their attributes, these individuals are then assigned a sequence of activities to perform in certain places at certain times. Finally, the routes and transport modes for each connecting trip are calculated to arrive at a full day plan for each individual.

Besides considering their decision processes, the actual behaviour and interaction of individuals in the network are also important. The individual's experience on the road in the presence of others acting out their day plans will affect her assessment of the quality of her own plan, and her future decision-making processes. Modern models therefore perform Dynamic Traffic Assignment (DTA) to produce time-dependent traffic volumes.

Of these modern transport demand planning models, the most ideal candidates are those that can retain individuals' attributes from start to end, as well as capture the transient effects that arise from commuter behaviour and interaction in the network. These models are known as Agent-Based Simulations (ABSs). In ABS, the modeller describes the constituent entities, or agents, that compose a system, and then simulates the behaviour of agents in each other's presence. The assertion is that, through ABS, system-wide behaviour can be "grown" from the decision-making, interaction, behaviour and learning of the constituent agents. Stated differently, ABS can capture *emergent effects*: complex system-wide phenomena that arise from the underlying agent dynamics (Bonabeau, 2002). ABS therefore should not only be able to predict the influence of capacity expansion decisions, but also that of changes to transport policy.

1.3 The South African transport environment: a case for agent-based transport simulation

The transport environment of the South African province of Gauteng presents an excellent case for the application of ABS to transport demand. From Figure 1.2 can be seen that Gauteng is South Africa's smallest province, occupying only 1.4% of the country's land area. But, according to the 2001 census, it is home to 8.8 million people and contributes 33.9% to the country's GDP. Accord-

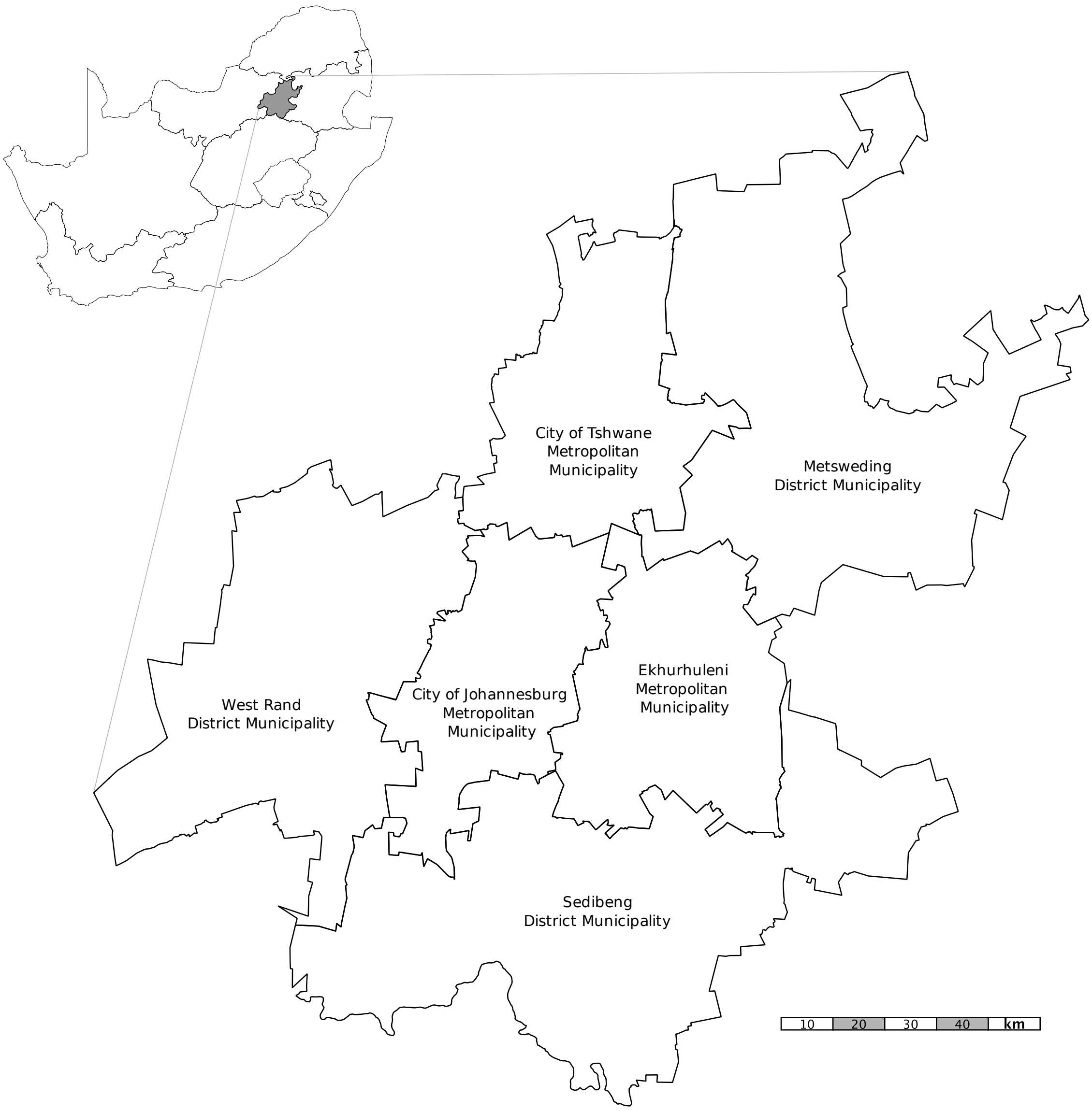


Figure 1.2: Gauteng Province — its relative location in South Africa, and its major district municipalities.

ing to the Gauteng Economic Development Agency (http://www.geda.co.za), the province generates 10% of the GDP of the entire African continent. As the most urbanised and populous province, Gauteng generates a massive demand for private, public and commercial transport. The following sections give a short overview of some of the most pertinent arguments for developing a microsimulation approach to Gauteng's transport demand planning.

1.3.1 Changing needs, obsolete methods

Transport demand planning in South Africa has suffered from serious neglect during the past 30 years. Obsolete procedures are used in strategic planning of future transportation networks, network loading constraints and road flow volumes (Diedericks and Joubert, 2006). Meanwhile, the demands from decision-makers have become more stringent; see for instance the South African Transport Master Plan requirements (DOT, 1996). The focus is shifting towards management of existing infrastructure, through interventions such as an Intelligent Transport System and the establishment of a Traffic Management Centre. Furthermore, the transport system is becoming increasingly complex as additions are planned to an already eclectic mix of First and Third World technologies. For instance, in Gauteng, the addition of bus rapid transit and the high-speed Gautrain modes will have a massive and, as of yet, unpredictable impact on commuter behaviour.

1.3.2 Apartheid-era design

The transport network layout is at odds with the needs of a commuter population for which it wasn't designed. The major transport infrastructure decisions were made during the apartheid era, when the changing needs of an increasingly integrated society were never considered. As a consequence, certain areas of the network infrastructure operate at or near capacity, where the predictive capability of traditional methods break down. This is due to dynamic, non-steady-state effects such as traffic spill-back from overloaded links onto the rest of the network, and individuals changing their travel behaviour in order to avoid traffic congestion.

1.3.3 Para-transit: the minibus taxi system of public transport

The relocation of black people during the apartheid years to the periphery of metropolitan areas gave rise to the birth of the minibus taxi industry. This dynamic transport mode is largely unregulated, de-centralised, driven by the free market principle, and rapidly adjusts to changing demand, making it responsible for the majority of public transport trips.

Minibus taxis are 15–25 seater vehicles, operating on semi-fixed routes, with no fixed time schedule. They stop anywhere and everywhere to pick up or drop off passengers but, in contrast with ordinary taxi services, charge fixed fares based on the length of their assigned route. Users flag down taxis on demand, a complex interaction that is impossible to translate into analytical terms and fed into a traditional demand planning model.

The larger scale dynamics of this para-transit mode clearly emerges from the interactions of its constituents. ABS is therefore the only paradigm capable of modelling the paratransit mode's current state, as well as predicting the effect any proposed governing policies might have on its behaviour.

1.3.4 Non-integrated transport modes

The transport system is multi-modal, with very little or no integration of schedules and routes between different modes. In the Gauteng area, for example, public transport is provided by several different bus operators, minibus taxi operators and a commuter rail system. These modes operate according to their own schedule, on their own routes, with no coordinated *inter-modal* planning.

Intra-modal planning involves the assignment of routes to service operators, such as the negotiations that take place within taxi associations. However, the focus of such negotiations are not focused toward improving efficiency and integration for the sake of the commuter; rather, the aim is to avoid violent conflict between different operators imposing upon each other's territories.

The complexities arising from such a lack of coordination are nearly impossible to translate into analytical terms to be fed into a traditional transport demand model. Only ABS is capable of modelling the necessary interactions between commuters and various modes of transport, as well as the decision-making involved on the individual level in selecting a series of interconnecting modes to get from one point to the next.

1.3.5 Freight traffic

South Africa has a large proportion of its road freight distributed during the day, when commuter transport demand is highest. As public transport is usually unavailable at night, and the working class lives on the outskirts of town, deliveries cannot be handled after hours. Freight, therefore, has a marked influence on day-time traffic patterns.

In Gauteng, an area with uneven topography and steep road gradients, the dynamic effects of sluggish freight vehicles cannot be ignored, and they significantly affect the effective flow capacity of a road. An ABS modeling approach is probably well-suited towards capturing the dynamic effects emerging from the presence of freight vehicles on the road.

1.4 Selecting an appropriate simulation platform

The arguments presented in the previous section serve as an outline of the requirements for a suitable South African transport demand planning technology. In summary, an ideal modelling technology should be able to:

- predict the traffic arising from commuters trying to realise their activity schedules;
- capture the dynamics of commuter behaviour and interaction such as mode-change, congestion spill-over and congestion avoidance, traffic signals, information feedback from intelligent traffic systems;
- simulate commuter learning;
- predict commuter reaction to policy changes, such as public transport modal integration or the introduction of a congestion toll;
- be adaptable to accommodate South African peculiarities, such as the para-transit minibus taxi system and freight transport;
- be able to work with a variety of input data;
- provide tools for detailed analysis of results; and
- be able to simulate a large-scale network and commuter population in a reasonable time on affordable hardware.

Lawson (2006) provides a review of a number of traffic microsimulation technologies. These platforms were investigated to find the best fit with the requirements listed above.

1.5 MATSim

Of the available ABS technologies, the Multi-Agent Transport Simulation Toolkit (MATSim) (MATSim Development Team, 2008) was found to be the most viable candidate. MATSim is in constant development, with expert teams in Berlin and Zürich working together to expand and refine the simulation platform. MATSim is an open-source project, and its source-code is freely distributable under the GNU Public License. Most importantly, MATSim has a modular structure, and the simulation package can be modified to suit South African requirements by modifying existing module code, or adding new modules to the package.

MATSim is currently used to simulate the private vehicle traffic for the Kanton Zürich in Switzerland, as well as for the German state of Berlin-Brandenburg (Balmer, 2007). In his study Balmer compared the performance of MATSim versus that of VISUM (`http://www.ptv.de`), a leading commercial transport demand planning technology based on the FSM. Both models were fed the same input data: a 24 hour origin-destination traffic flow matrix of the Kanton Zürich. When simulated traffic counts were compared to actual traffic count data, MATSim delivered a better prediction of the change in traffic volume over the course of the day.

The Berlin-Brandenburg and Kanton Zürich scenarios were remarkably different from each other in terms of their input data. MATSim proved flexible enough to produce accurate results, regardless of the difference between the two applications. These results suggest that MATSim might be successfully adapted to Gauteng's transport scenario.

1.6 Research question

From the reasoning outlined above, our research team arrived at the objective of this study: to develop and test an initial implementation of MATSim for Gauteng. The main research question can then be stated as follows:

> Given the network, land-use, census and individual travel-pattern data available for Gauteng, will MATSim be able to predict the private vehicle traffic observed on some reference date for a selection of key network links?

1.7 Research design and methodology

The main research deliverable is a working initial implementation of MATSim for Gauteng, which will be referred to as MATSim-SA. This initial implementation only simulates private vehicle transport demand. As with the initial studies by Balmer (2007) for Berlin and Zürich, we assumed a very simple sequence of activities: all commuters leave home in the morning to travel to work, spend some time there and return home. The final results compare simulated travel patterns to those observed in reality, for a number of strategic links in the Gauteng road transport network.

MATSim-SA will form the basis of further research, with the aim of continuous expansion and improvement. A disciplined development approach was required to facilitate a process of continuous improvement while simultaneously resulting in worthwhile research deliverables. To arrive at an initial implementation, we followed the design research methodology outlined by Manson (2006), which is shown schematically in Figure 1.3. He summarises the design research process as follows:

Awareness of the problem The researcher becomes aware of a problem and constructs a formal proposal to begin a new research effort.

Suggestion The researcher comes up with one or more tentative designs, which are intimately connected to the proposal.

Development The researcher builds one or more artefacts. For MATSim-SA, the artefacts would be the initial MATSim data set for private vehicle traffic, as well as the simulation results produced by MATSim from the data set.

Evaluation Once constructed, the artefact is evaluated against the criteria that are either implicitly or explicitly contained in the proposal. In the case of MATSim-SA, the main criterium is the research question posed above. Deviations from expectations must be tentatively explained. Before and during construction, researchers will posit hypotheses about how the artefact will behave. New suggestions are abducted and the design is modified, and the process repeats.

Conclusion At some point when the effort is considered "good enough", or some previously specified performance measurement is attained, results are consolidated and written up. Knowledge is classified as firm (repeatable facts learnt) or loose ends (anomalies, the subject of future research).

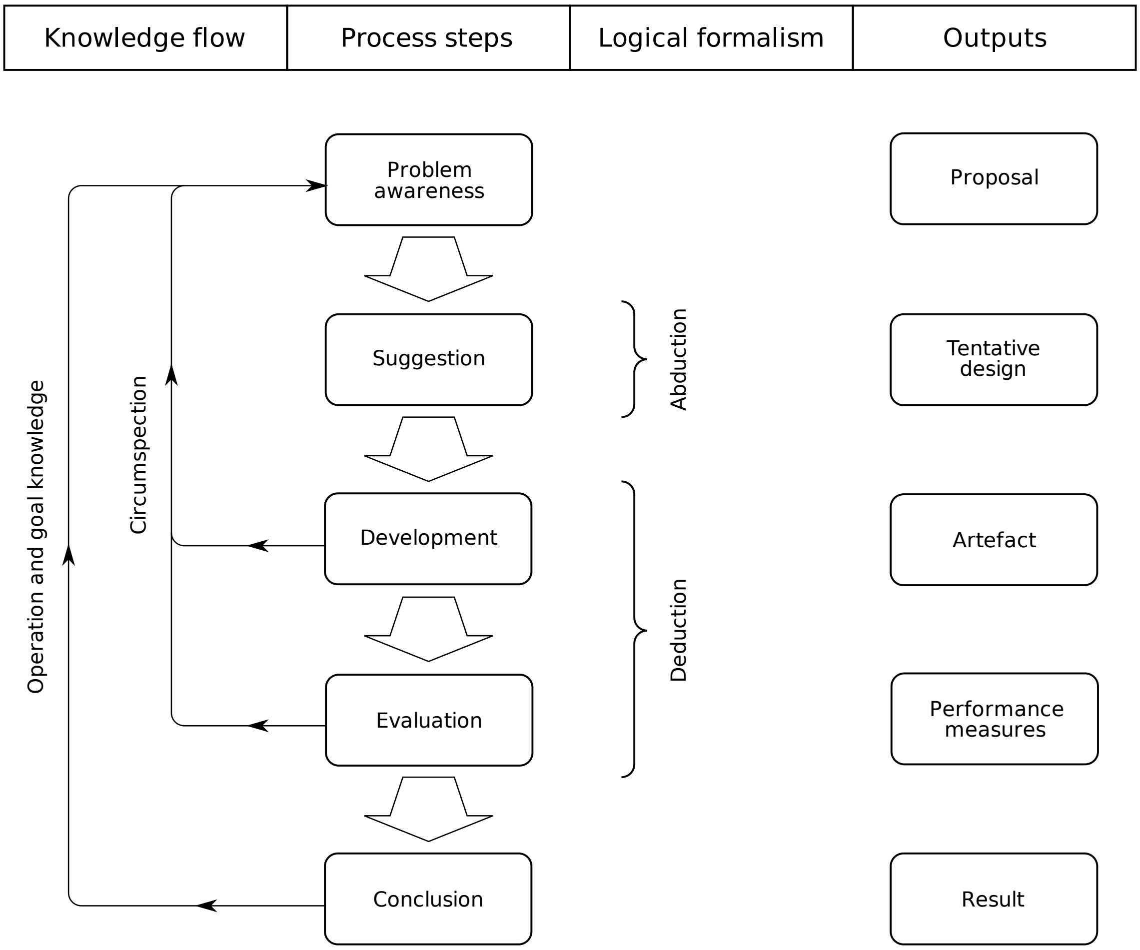


Figure 1.3: The general methodology of design research. Adapted from Manson (2006) and Vaishnavi and Kuechler (2005).

The design research methodology meets the requirements of producing a transferable artefact that can form the basis of further improvement, while generating worthwhile knowledge in the process. The method fits well with the research and development approach followed by the core MATSim teams in Zürich and Berlin.

1.8 Outline of dissertation

This chapter outlined the need in South Africa for an advanced transport demand planning technology and put forward the main arguments why the MATSim platform is the strongest available candidate to fulfil our current and future demand planning needs. The main research objectives and deliverables were identified, along with the development methodology.

The following chapter provides greater detail from the literature on the relevant developments in modern transport demand planning. Chapter 3 describes the processes of network generation, initial demand generation and the simulation process itself. Chapter 4 presents an analysis of the results, and compares them against real-world data from traffic count stations in Gauteng. Chapter 5 summarises our findings and concludes with an outlook on current and future developments.

Chapter 2

Literature review

A review of the literature on transport demand planning reveals two themes to be central to the current state of the art:

1. Recognising the *activity*, not the *trip*, as the primary unit of analysis. Transport is a derived demand, and the so-called Activity-Based Approach (ABA) to transport demand modelling improves on the crude concepts of trip production and attraction to arrive at a far more precise description of transport demand, both in space and time.

2. Dynamic traffic assignment. The Four Step Model (FSM) cannot calculate when a vehicle is on a particular link. In a sense, it assumes that the vehicle is simultaneously on every link in its chosen path (Meyer and Miller, 2001). Modern techniques discard this unrealistic assumption and predict link volumes over the course of a day.

This chapter gives an account of recent advances in these two fields of study, which form the basis of MATSim's principles of operation.

2.1 The ABA to transport demand planning

In the 1970s, traditional demand planning models based on the FSM were increasingly criticised, due to general unreliability and a failure to assess the effects of policy measures correctly. Ever since, transport demand modelling research has become increasingly oriented towards developing models that predict complex responses in individual activity and travel patterns to new policy measures.

The core perspective of the ABA is that, by predicting which activities are performed at particular destinations and times, one has implicitly forecasted the trips and timings necessary to bring about those activity patterns. Activity-based models therefore aim to disentangle, identify and subsequently mimic the decision-making processes that lead to individual activity schedules (Ettema, 1996).

2.1.1 Aims of the ABA

The ABA attempts to predict the answers to four questions on each individual in a transport study area (Bowman, 1998; Ettema, 1996; Meyer and Miller, 2001):

Who? Individuals tend to make activity decisions that are influenced by their defining attributes. The activity schedule of a married, middle-aged, middle-class working mother differs from that of a low income, single male student. Demographics affect all aspects of an activity schedule: home location, primary activity type and location, travel modes, how easily activities may be re-scheduled, whether the person travels with others to activity locations and so forth.

The ABA therefore requires a synthetic population of individuals with defining attributes as its point of departure. It is for this reason that ABA transport demand models are frequently referred to as being "disaggregate".

What? Based on her attributes, an individual is expected to perform a selection of activities from a likely set of choices. The elements of this choice set are determined by travel surveys, and a variety of algorithms are used to assign a set of activities to each individual.

When? The sequencing, timing and duration of activities are important attributes that are influenced and constrained by factors such as facility operating hours, expected travel times, availability of transport, the presence of other individuals and precedence (i.e. if you are at work from 09h00, you must have left home before then).

Where? If you have many opportunities for gainful employment close to home, you are unlikely to travel very far to work. You might decide to take the bus or walk if you live and work close to the city centre. If employment opportunities are only available further away or there is an incentive to make more money in the neighboring town, you may be forced to commute. These activity location decisions are influenced by several factors and, in turn, can affect such decisions as mode choice, activity sequence and, in the longer term perspective of overall urban dynamics, even the home location.

Once the individual is synthesised (the answer to the "who?" question), the ABA attempts to answer the remainder of the questions "simultaneously"; i.e. it recognises that all activity-related decisions are interdependent and that, in reality, individuals integrate across a large set of choice decisions and constraints to come up with their final activity schedule. In general, this integrative approach relies on feedback between different model components when constructing the activity schedules, as well as repeatedly executing and evaluating schedules in a network assignment step.

2.1.2 Methods and techniques

The ABA to travel demand modelling does not refer to a single technique. Rather, the term groups together a variety of models with different conceptual frameworks, but all of which are means to the same end: to arrive at a realistic schedule of activities for each individual in the study area. It remains the responsibility of the transport planner to decide on the combination of techniques appropriate to their planning requirements and available data.

Ettema (1996) provides a much-cited review of the historical development of the ABA and discusses the underlying conceptual frameworks of the most prominent models to have emerged. In general, he classifies ABA models as follows :

Space-time geography. These models systematically identify the set of feasible activity patterns, given land-use patterns, time constraints and available transport options. They provide a good evaluation of the possibilities open to the commuting population, but provide little information on the behavioural response, i.e. the set of patterns that will be chosen for a specific set of policies.

Discrete-choice theory. These models derive a set of activity patterns from those observed in reality, thus incorporating the interdependencies between different decisions that real individuals integrate into their overall activity pattern. As this choice set of activity patterns is limited to those observed in reality, not all relevant alternatives may be included in the model. Stated differently, the model might be able to provide a good representation of the status quo, but prove to be too inflexible to predict the behavioural response to radical changes in policy.

Microeconomic theory. The so-called utility-maximising models are useful in describing individuals' activity time expenditures, but not the sequencing of those activities, nor the resultant travel behaviour in terms of the destination and mode choice required to describe the trips that link activities.

Artificial intelligence and cognitive science techniques. These techniques describe activity selection and scheduling in terms of human reasoning processes. However, they tend to be qualitative, making them difficult to calibrate, as well as specific to individual data gathered in a study area, making them difficult to transfer to other individuals or contexts.

The interested reader is referred to Ettema (1996) for his account of the underlying theory of activity scheduling behaviour and resulting activity patterns. In an earlier review, Axhausen and Gärling (1992) critically examine the behavioural assumptions of conceptualisations and models of activity scheduling, and highlight areas in need of future research. More recently, Pas (1996) provides an overview of advances in ABA to travel demand modelling, both in terms of the methodologies being used and the phenomena being modelled.

2.2 Dynamic traffic assignment

The activity schedules of the study population represent an initial demand for transport that the transportation infrastructure has to accommodate. The trips that connect activities need to be routed through the network in order to assure the timely arrival at activity locations, while respecting the capacity constraints of the network. In the same way activity schedules go beyond the traditional practice of trip generation and attraction to produce a time-dependent demand for transport, the process of Dynamic Traffic Assignment (DTA) goes beyond the static assignment approach of the FSM to produce time-dependent link volumes.

DTA has evolved significantly in the past 30 years, originally to fill the shortcomings of traditional demand planning methods, but more recently to aid in the evaluation and operational deployment of Intelligent Transport Systems (ITS) technologies (Peeta and Ziliaskopoulos, 2001). DTA refers to a broad spectrum of problems in the field of mathematical programming, and has been tested for tractability in various solution approaches.

In general, mathematical tractability is traded off against realism in terms of the traffic phenomena and driver behaviour being modelled. Based on the demands of the application, a number of solution approaches can be employed, ranging from analytical to simulation-based procedures.

In their review, Peeta and Ziliaskopoulos (2001) evaluate a number of solution approaches and classify them into four broad methodological groups: mathematical programming, optimal control, variational inequality and simulation. The first three groups are labelled as analytical approaches. They found the greatest limitation with the analytical models to be the trade-off between mathematical tractability and traffic realism. On the other end of the spectrum, a simulator achieves a higher degree of realism and solution accuracy at the expense of computational efficiency and the ability to analytically derive useful theoretical insights. Simulation models are, however, much easier to implement for general networks and are easily modifiable and extensible. These attributes make simulation modelling an attractive method for transport demand planning, where the planning horizon is far greater than that of ITS, and the costly objectives of capacity planning and governing policy require a greater degree of model detail and accuracy.

Analytical techniques seek to establish a user equilibrium for the network such that no user can perform better by travelling a different route through the network for any given trip. Simulation techniques pursue an analogous solution state through the iterative process of systematic relaxation (Nagel, 1995). Balmer (2007) summarises this process as follows:

1. Start with an initial guess for the routes.
2. Perform network loading by executing all routes simultaneously in a traffic flow simulation.
3. Adjust some or all of the routes based on the results of the network loading step. As the link travel times for the current iteration are known, this feedback step simulates commuter

learning: re-routing aims to optimise travel time, so the adjusted set of routes avoid congested links where possible.

4. Repeat from the second step until the average travel time shows no significant improvement from one iteration to the next.

2.3 Agent-based transport demand simulation

The ABA delivers individual activity schedules for a synthesised population. If the process of DTA can maintain individual integrity, a far richer set of results can be obtained: instead of just link travel times, it becomes possible to evaluate individual plan performance following each traffic assignment step. This information can then be fed back into the activity generator, allowing for the adjustment of individual activity schedules that goes beyond the re-routing of trips. As the analytical approach to DTA already proves intractable dealing with *anonymous* time-dependent trip flows through the network, this individualised, *multi-agent* approach is limited exclusively to the domain of simulation.

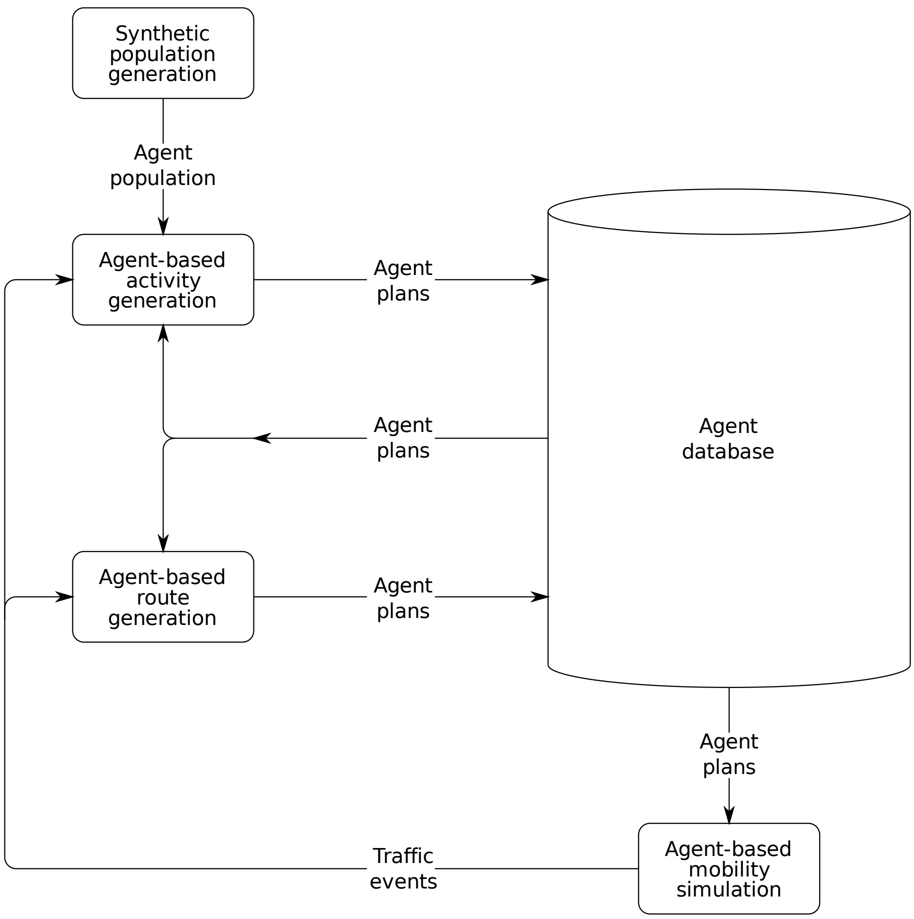


Figure 2.1: An agent-based transport demand modelling framework. Adapted from Balmer et al. (2006).

Balmer et al. (2006) propose a framework for Agent-Based Simulation (ABS) of transport demand. Shown in Figure 2.1, this framework integrates the state of the art of activity-based demand modelling and DTA through ABS, with the microsimulation step providing feedback both to activity generation and route calculation. It forms the basis of operation for the Multi-Agent Transport Simulation Toolkit (MATSim) package, the open-source transport demand ABS which is currently being developed and implemented in Zürich and Berlin (MATSim Development Team, 2008). The following section examines the actual operation of MATSim in more detail.

2.4 Simulation procedure

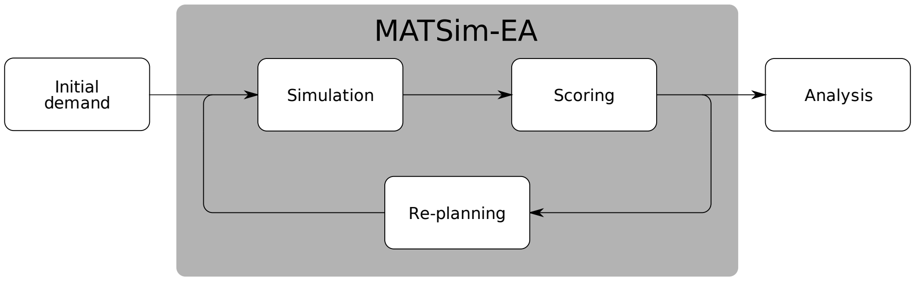


Figure 2.2: Schematic illustration of MATSim's transport demand modelling process. The shaded area indicates the evolutionary "engine" of MATSim, which simulates system learning and adaptation. Adapted from Rieser (2007).

Figure 2.2 shows a simplified view of MATSim's operation. Initial plans are fed into the mobility simulation, executed simultaneously and then evaluated to arrive at a score for each executed plan. The replanning step provides the feedback that allows for the evolution of commuter plans. "Evolution" is an apt term in this context: MATSim's iterative process of mobilty simulation, scoring and replanning is based on the concept of a so-called Evolutionary Algorithm (EA). EAs solve optimisation problems by generating and improving "populations" of candidate solutions using mechanisms inspired by biological evolution such as mutation, recombination and selection (Engelbrecht, 2002).

In MATSim's context, "population" refers not to the commuter population; the simulation process ends with exactly the same "people" it started out with. Instead, it refers to the collective "memory" of the commuters; a collection of plans associated with each commuter that changes with increasing iterations. From an EA perspective, it is this "population" of plans that evolves over time.

2.4.1 Simulation

Each individual starts out with only one plan, specified in the process of initial demand generation. This plan is a simple schedule of activities, their locations in the study area and a preferred mode of transport to link those activities. During the first iteration, each plan is routed through the network using a Dijkstra algorithm that attempts to minimise travel time (Balmer, 2007). All commuter plans are then executed simultaneously in the mobility simulator. During the simulation, events are registered for each individual, such as travel time, activity departure and arrival times.

2.4.2 Scoring

In the next step, a score value is calculated for each executed plan. From an EA perspective, the score is a measure of fitness for each plan. MATSim uses a simple utility-based approach to calculate a plan score (Balmer, 2007) :

$$U_{total} = \sum_{i=1}^{n} U_{perf,i} + \sum_{i=1}^{n} U_{late,i} + \sum_{i=1}^{n} U_{travel,i} \tag{2.1}$$

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 one of these utility values are time-dependent functions of arbitrary complexity (the interested reader is referred to Charypar and Nagel (2005) for a detailed description of these utility functions). Regardless of the constituent function complexity, it is evident from Equation 2.1 that greatest utility and, consequently, plan fitness, derives from more time spent performing activities, while avoiding travel and arriving late at activity locations.

2.4.3 Replanning

Once executed plans have been scored, the next step in MATSim's iterative demand simulation process executes. The so-called replanning step refers to a sequence of configurable algorithms, and is analogous to the mutation and selection processes found in EAs. In his thesis, Balmer (2007) lists a number of algorithms used to select from and expand the "population" of plans.

These algorithms ensure that the "population" of plans gradually adapts to the transport environment. From an EA perspective, the replanning step is analogous to the mechanisms of mutation and selection — activity times get adjusted, routes are optimised based on recent system performance and poorly performing plans get discarded. This process of adaptation is reflected in the improvement of average plan scores and travel times with increasing iterations, until a form of the Nash-equilibrium is reached, whereby changes in commuter plans do not produce any further improvement in total utility.

A core requirement in the MATSim development process is modularity. The algorithms listed above, as with all components of the MATSim framework, are selectable and configurable to suit the requirements of the transport modeller's application. But while the components that make up any particular demand simulation process are changeable, the MATSim framework, as shown in Figure 2.1 and principle of operation in Figure 2.2, remain unchanged. Adherence to the framework, operational principles and well-defined interfaces of the MATSim project make it possible for a distributed team of user-developers to collaborate and thus continuously improve the application, while ensuring the capability to customise it to their requirements.

This chapter provided an overview of recent advances in transport demand modelling, which form the background and motivation for our own implementation of MATSim for Gauteng. In the following chapters, details of the development and performance of the South African implementation unfold.

Chapter 3

Data preparation

The aim of the initial implementation of Multi-Agent Transport Simulation Toolkit (MATSim) for Gauteng was to model the home–work–home activity chain for the census year of 2001. We assume that this primary activity chain is responsible for the majority of traffic during the morning and afternoon rush hour peaks.

The minimum requirement for a MATSim simulation is a set of initial plans, a road network and a simulation configuration (`config.xml`) file. Initial plans refer to a set of activity schedules for each individual, while the road network is a directed graph of nodes and links. The `config.xml` file specifies the relevant modules to invoke during mobility simulation, replanning and post-simulation analysis, and also contains simulation parameters such as the projection coordinate system, network attribute scaling factors and utility values to use for activities and travel when calculating plan performance.

Figure 3.1 provides an overview of the process followed to transform our input data to MATSim specifications. The sections to follow describe each step in detail.

3.1 Network development

The bottom section of Figure 3.1 shows the steps involved in transforming our Geographic Information System (GIS) shapefile data into a MATSim `network.xml` file. In MATSim, a valid network description is a weighted digraph of nodes and links that should satisfy the following two conditions:

1. Each link must specify its flow capacity, usually in terms of vehicles per hour, along with its length in meters, free speed in meters per second, and number of lanes.

2. Each node in the network should be reachable from any other node, to prevent vehicles from becoming trapped.

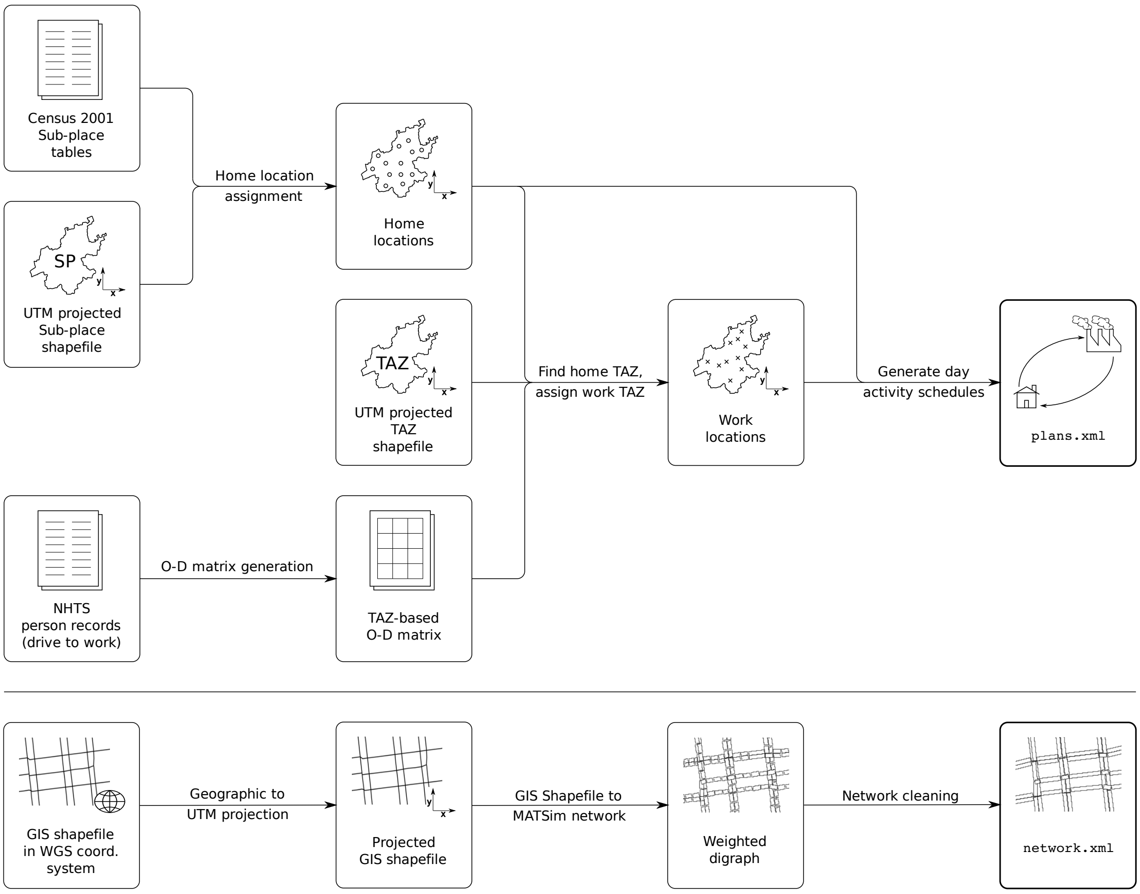


Figure 3.1: Overview of the process followed in preparing Gauteng input data for use with MATSim.

3.1.1 Source data

Our research group obtained a GIS shapefile depicting the Gauteng road transport network from Business Connexion GIS, a Gauteng-based business solutions provider. The GIS dataset forms part of their business intelligence product range, and is known to be valid for 2007. Each line shape in the file is tagged with attributes, including the type of road, speed limit and measured length.

3.1.2 Projection

As with the other GIS data used in this project, the network shapefile uses a geographic coordinate system, whereby coordinates are recorded as latitude and longitude decimal degree values on an ellipsoid, relative to some datum point. In this case, the ellipsoid is the World Geodetic System 1984 (WGS84) ellipsoid, commonly known as WGS84, and the datum point is the ITRF91 (epoch 1994.0) coordinates of the Hartbeeshoek Radio Astronomy Telescope.

MATSim, however, requires Cartesian coordinates for all spatial input data, which means that geographic coordinates need to be transformed. We transformed our network and other relevant geospatial data to the Universal Transverse Mercator (UTM) coordinate system (USGS, 2001). The transformation proved valid as the measured Euclidean distance of the converted network line shapes match their real-world values to within 0.1%.

3.1.3 Network graph preparation

The full Gauteng road network, as recorded in the GIS shapefile mentioned above, comprises approximately 250,000 road segments (referred to as *links*), meeting in approximately 185,000 junctions, intersections or other discontinuities (referred to as *nodes*). Describing each link completely, according to the specification given in the introduction to this section, would have been a near impossible task to perform manually in the allocated time for this study. Instead we used the information contained in the GIS shapefile to specify and infer link attributes.

We are fortunate enough to have a GIS dataset is hierarchically organised by road type, making it possible to define the network at more than one resolution. It was therefore possible to discard low-capacity, low speed suburban street-level links from our final network description, and only describe a network consisting of roads classified to be main roads, dual carriage ways, national roads and highways. Eliminating the street-level links from the network description greatly increases the computational efficiency of the simulation, possibly at the expense of simulation accuracy.

We therefore compiled two network descriptions: one using the full GIS dataset, and another containing only the main road classification and higher. If the smaller network gives similar counts on major arterials as the full network description, it becomes a valuable proxy for the full network, allowing one to rapidly simulate various scenarios.

The two networks derived in this process are shown in Figure 3.2 and Figure 3.3 respectively.

3.1.4 Network cleaning

The network graphs derived from the GIS shapefile contain a number of features that do not comply with the second requirement of a valid network, namely that all nodes be reachable from any other node. These features were removed with the `NetworkCleaner` class, a Java utility that forms part of the MATSim toolbox. `NetworkCleaner` identifies the largest contiguous cluster of links in the network, and then proceeds to remove all links that cannot be reached from this main cluster. It also rids the main cluster of nodes from which there is no escape (sinks) and nodes that are unreachable (sources).

Class `NetworkCleaner` produced a valid network, but close inspection of the network graph shows a lot of very short links, as can be seen in the first two images of the sequence shown in Figure 3.4. These short links are due to the structure of the original poly-line in the GIS shapefile from which the network was derived. Each line segment from the GIS shapefile is converted into

Figure 3.2: Full-sized representation of Gauteng's road network, derived from GIS shapefile.

Figure 3.3: Reduced or small network representation derived from GIS shapefile, depicting only main road and higher level network links.

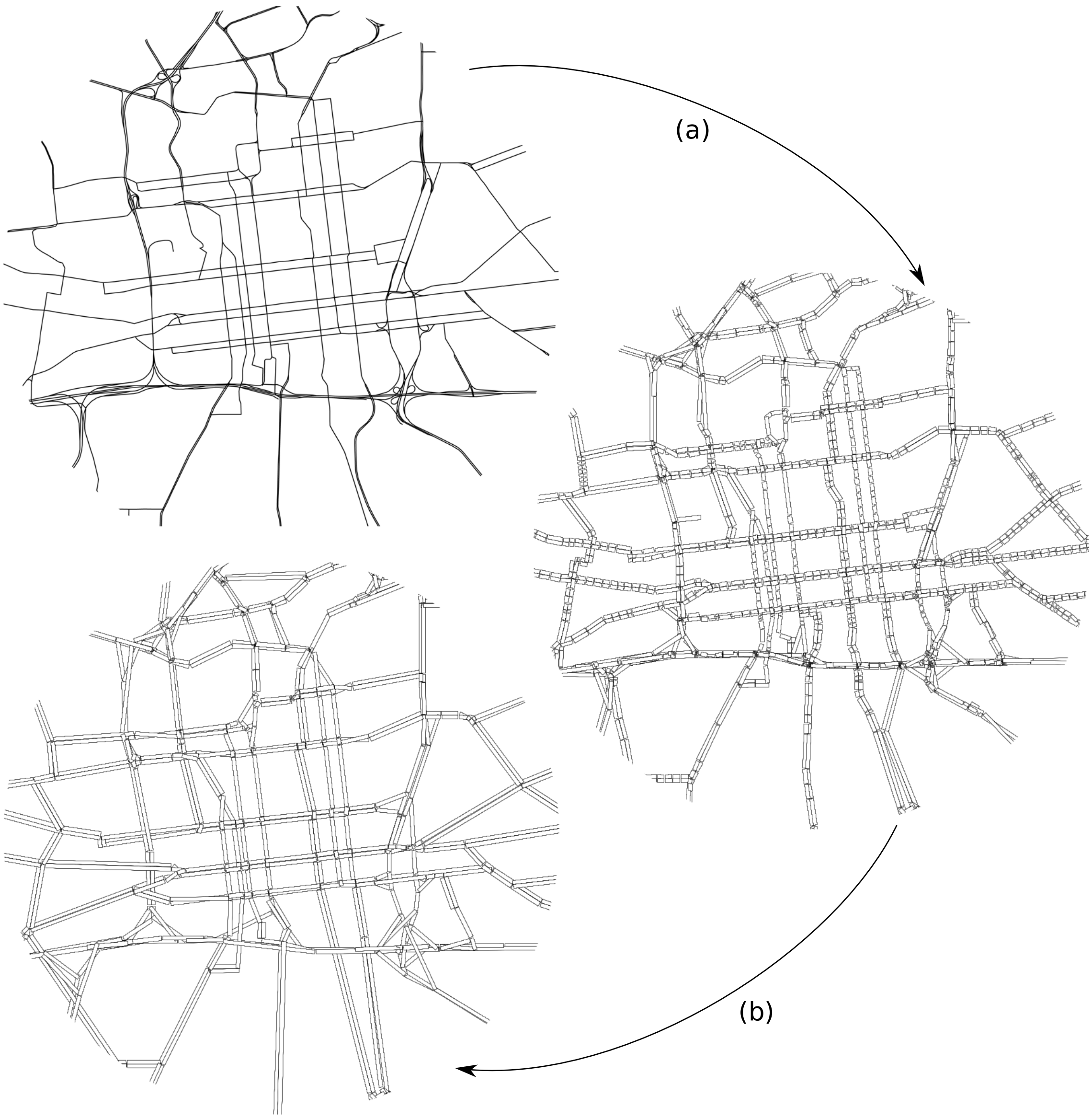


Figure 3.4: A comparison of a section of the Gauteng road network (the Johannesburg city centre), as it is represented in the original GIS shapefile, versus its initial conversion to a network graph (a), and after redundant links have been consolidated (b).

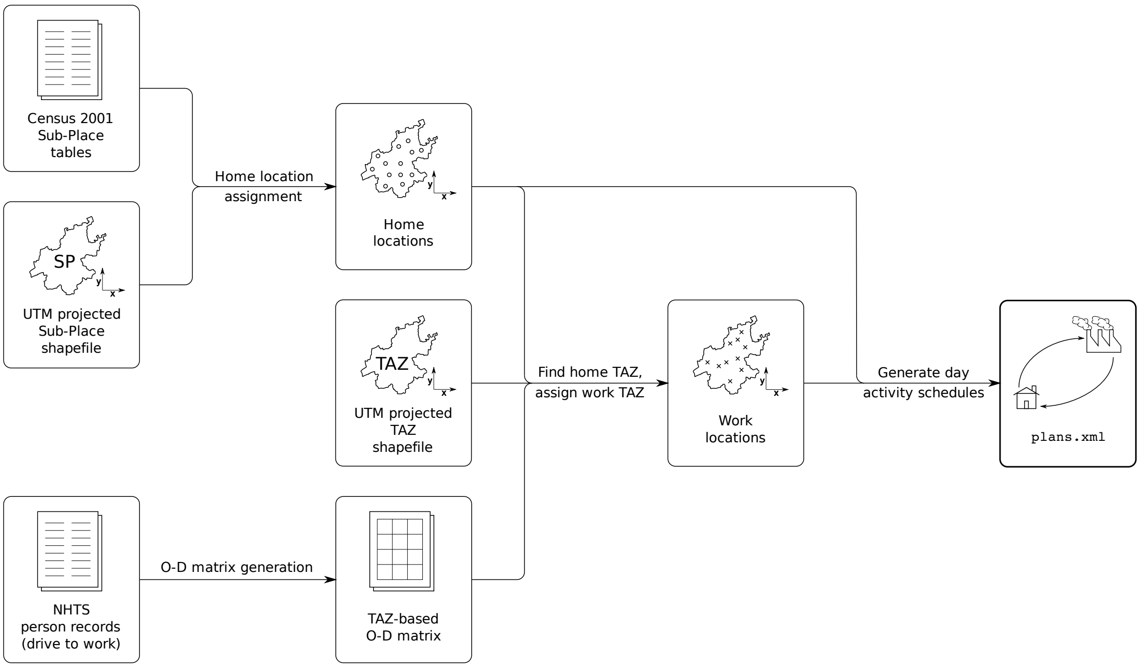


Figure 3.5: The method followed to derive our initial transport demand from census and travel survey data.

a link on the network graph, resulting in many more links than are necessary to mathematically describe the network. Such a redundant network description results in calculational inefficiencies during simulation, but produces a network graph that more closely resembles the actual road network in a visualisation. For the purposes of this study, it was decided to rid the network of redundant links by fusing them, the result of which is illustrated in Figure 3.4b.

3.2 Initial demand preparation

Figure 3.5 summarises the process followed to derive our initial demand. In the following sections, we describe each step in more detail.

MATSim requires the initial day activity plan for each individual to contain at least the following information:

1. The person's activity locations, each given as a set of coordinates. These coordinates are used to associate each location with a link in the network.
2. Preferred departure and arrival times for each activity, or a departure time for the first activity and preferred durations for subsequent activities.
3. A travel leg connecting each activity, specifying the mode of travel to be used.

Our initial implementation only models the private vehicle transport mode. For the purposes of this initial implementation, we therefore needed to identify the population segment that drives to work, determine where they live and where they are likely to work. To generate home locations, we used processed 2001 census data, provided by Statistics South Africa (Stats SA). We derived likely work locations from the 2003 National Household Travel Survey (NHTS). The following sections discuss the initial demand generation process in detail.

3.2.1 Home location assignment

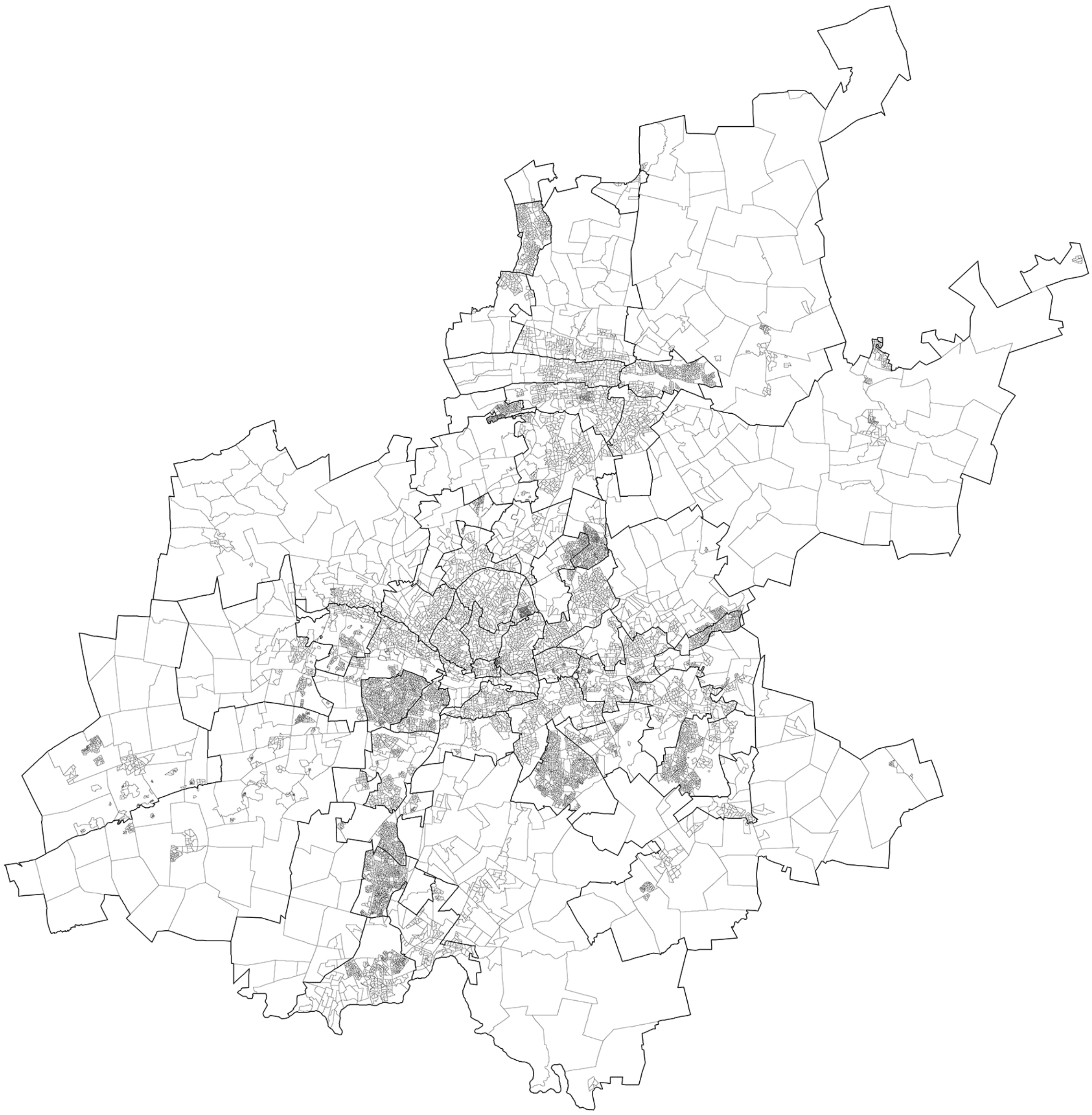


Figure 3.6: The EAPSUs used during the 2001 census. The darker lines illustrate the grouping of EAPSUs into TAZs, used during the 2003 NHTS. Source: Stats SA.

In preparation for the 2001 census, Stats SA produced a mapping of South Africa whereby the entire country was demarcated into a total of 80,787 non-overlapping EAPSUs. Each EAPSU is a small land unit of manageable population and area allocated to a single person to enumerate during the census count (Stats SA, 2001). Figure 3.6 shows the EAPSUs for Gauteng.

The optimal strategy for generating our baseline synthetic population would be to create a synthetic individual with the exact attributes recorded in the person records of the census, and then to locate each individual at a likely location within their particular EAPSU. A likely location here refers to some domestic structure, such as a house or an apartment block. The larger the structure, the more likely it is that an individual will be located there. This method of assigning surface location is called dasymetric mapping (Mennis, 2003).

Unfortunately, in order to protect the identity and information of the population, Stats SA only released anonymous records of a 10% sample of the 2001 census. The smallest identifiable geographic unit in these person records is the municipality, and therefore is far too coarsely grained to serve as input to a highly detailed transport demand analysis.

Stats SA provides another anonymous source of information, a collection of so-called Sub-Place (SP) tables, which ensures individual anonymity through aggregation, with a far higher degree of spatial resolution than the municipality level. A SP refers to a grouping of EAPSUs, and is the first level in a hierarchy of such groupings that was defined by Stats SA to allow for the systematic aggregation of census data. SP tables provide aggregate figures on parameters of interest from the census questionnaire, such as the number of males/females in each SP, or the SP's income distribution. For this study, we used the number of people who reported to drive to work in each SP to arrive at a study population size of 924,680 individuals. We assumed these individuals to be responsible for the bulk of private vehicle traffic in our study area, and proceeded to generate a 10% synthetic population for each SP.

In order to locate individuals inside an SP, we employed the following strategy. Each SP is composed of a number of EAPSUs; these are assumed to have approximately the same population density. Therefore, if a SP has a total of x people who drive to work, and is composed of y EAPSUs, our synthetic individuals were randomly distributed within the boundaries of the SP in such a way that each composing EAPSU contains, on average, $\frac{x}{y}$ individuals. The result of this process is shown in Figure 3.7.

3.2.2 Work location assignment

A person's primary activity location is the result of a very complex function, depending on such variables as who the person is, where she lives, car ownership, her level of education and experience, and the number of opportunities that exist for a person with her attributes, to list but a few. These variables also show interrelation amongst themselves. A person's work location might be dependent on where she lives if there are many equal opportunities available, making it possible to opt for the job closest to home, in which case home location largely determines job location. On the other hand, less plentiful job opportunities or earning a high salary in a remote area could make it worthwhile to move closer to where there is work, in which case job location

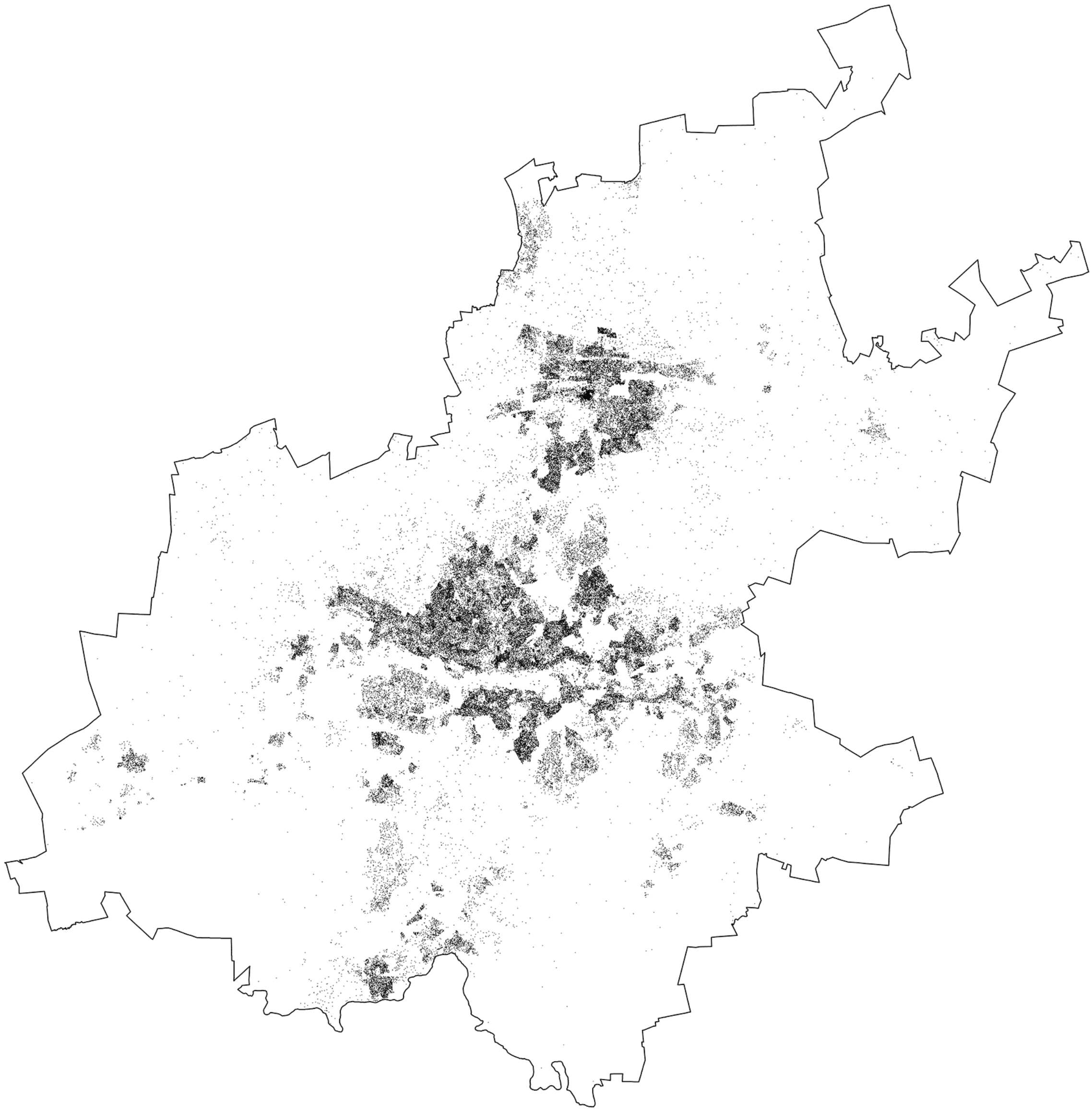


Figure 3.7: Home locations for the 10% synthetic population of private vehicle drivers used in this study.

largely determines home location.

Models of activity location assignment attempt to capture this complex interaction between decision variables and system constraints to varying degrees, using census information and travel surveys for calibration and validation. An ideal model would be able to predict changes in primary activity location as a function of some system change, by having the location assignment arise from the agent's simulated decision-making.

For lack of a model of such sophistication, we based our work location assignment on an extrapolation of travel survey data. As the 2001 census did not record the locations of each individual's primary activity, we used data from the 2003 NHTS to assign a work location to each synthetic individual, such that a survey of our synthetic population would return the results of the original NHTS.

The NHTS is a survey of the detailed travel behaviour of a sample of 45,000 households throughout South Africa. The survey was developed and conducted by Stats SA for the Department of Transport in 2003. In Gauteng, 7,839 individuals participated in the survey, of which 3,038 reported to drive to work. Assuming no significant change from the census date, this sample represents only 0.33% of the segment of the population who reported in 2001 that they drive to work. For lack of further information, it was decided to proceed with NHTS data to derive a first approximation of work location.

In the NHTS, the smallest discernible geographic unit for work location is called a Traffic Analysis Zone (TAZ). Gauteng contains a total of 58 TAZs, each composed of a number of adjacent EAPSUs from the 2001 census, as is illustrated by the darker lines in Figure 3.6.

Each person record in our NHTS data set collection records home and work location as a reference to one of these 58 TAZs. From these records, a so-called origin-destination matrix was compiled, which shows how many people coming from a particular TAZ work in every TAZ in the study area. Normalising this matrix by dividing each element by its row total gives an estimated probability of a person's work location TAZ based on their home location TAZ.

This normalised O-D matrix was then used to probabilistically assign a work location TAZ to each individual. It was assumed that an individual assigned to work in a particular TAZ will be equally likely to work in any EAPSU that the TAZ is composed of. Therefore, if a TAZ has a total of x work activities assigned to it, and is composed of y EAPSUs, work activity locations were randomly distributed within the boundaries of the TAZ in such a way that each composing EAPSU contains, on average, $\frac{x}{y}$ work activity locations. The result of this process is shown in Figure 3.8.

Once home and work locations have been established, all that remained was to construct the activity schedules that make up a `plans.xml` file. Each individual requires an initial estimate of the time they depart for work. In this application, initial departure time from home is a uniformly distributed random variable ranging in value between 05h00 and 07h00. It was then assumed that

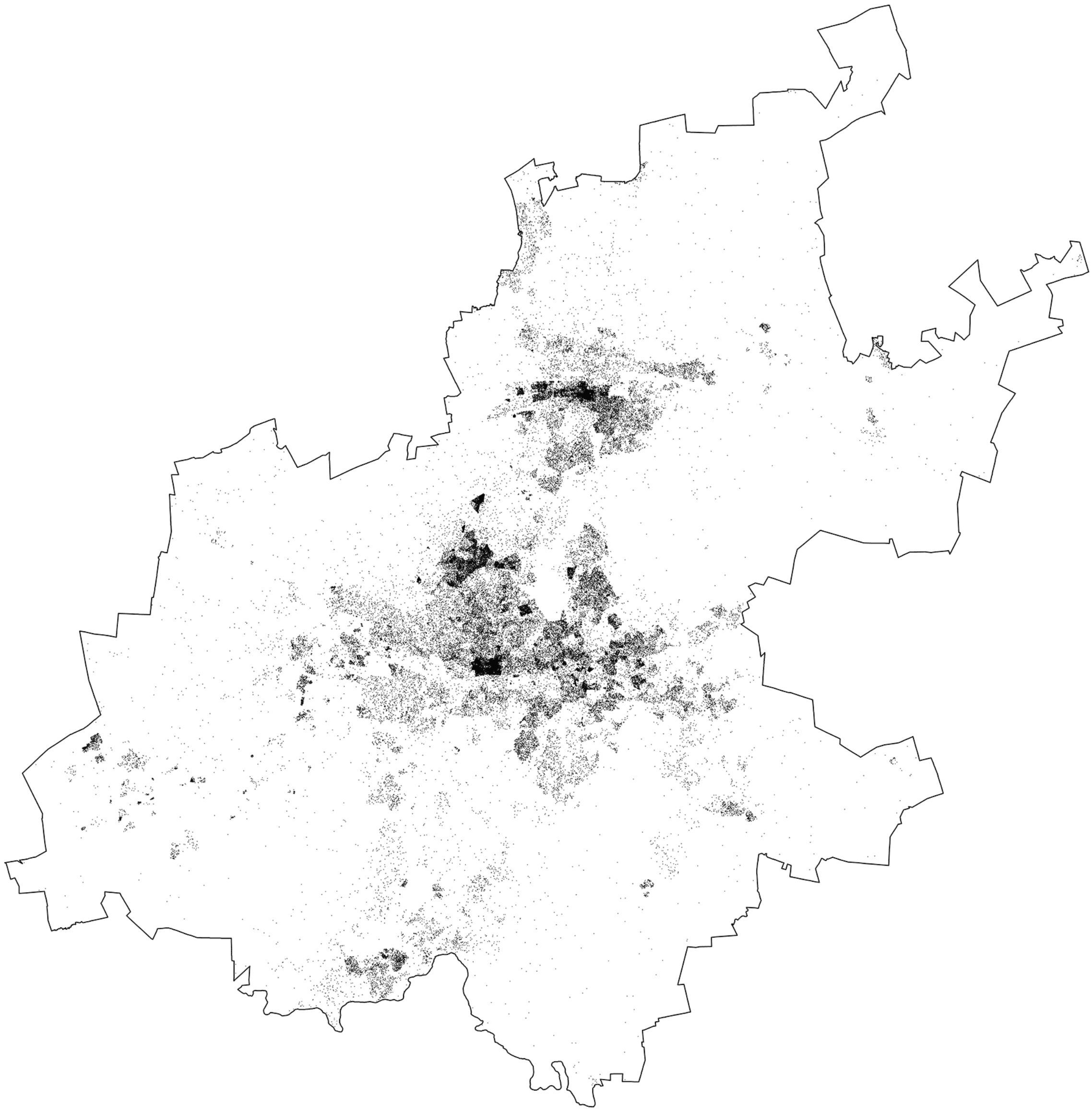


Figure 3.8: Work locations for the 10% synthetic population of private vehicle drivers used in this study.

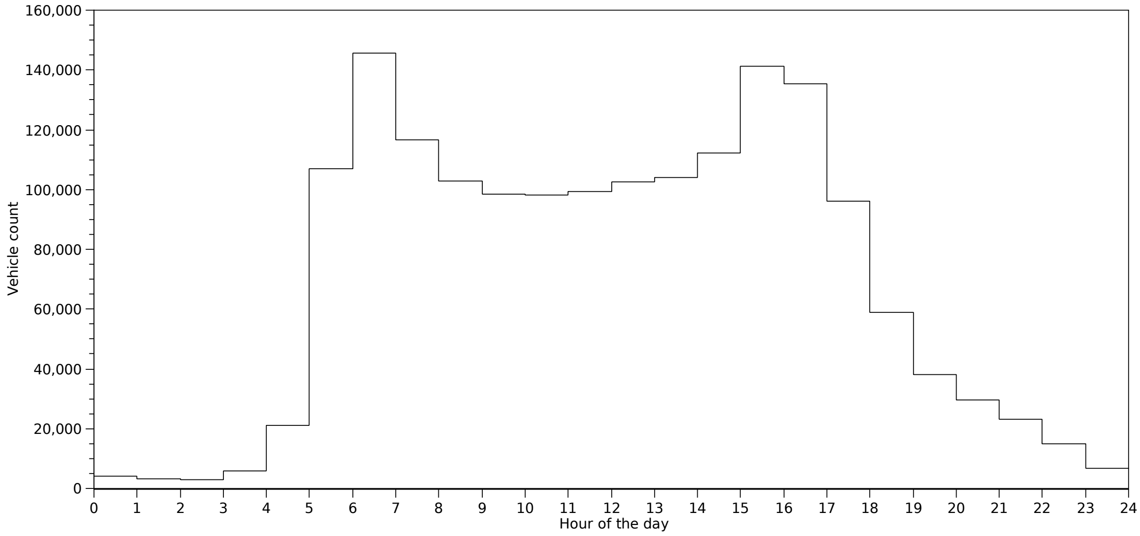


Figure 3.9: Total traffic counts for the 41 selected SANRAL counting stations used in this study, showing peaks for hours 7 and 16, thus suggesting a nine hour workday.

agents will remain at work for nine hours before returning home, as South African labour law prescribes a workday of eight hours plus one hour for lunch. This assumption is confirmed by the South African National Roads Agency Ltd. (SANRAL) traffic counts data used for validation, as can be seen in Figure 3.9. Finally, the activity schedules were written to a `plans.xml` file to complete the initial demand generation process.

3.3 Simulator configuration

Once the network and initial demand were prepared, MATSim was configured to perform the simulation. A `config.xml` file was prepared, which is a list of simulation parameters formatted to the appropriate MATSim XML specification. These simulation parameters include references to input data, scaling factors for network capacity (in order to perform simulations on population samples of varying size), parameters for the utility scoring functions, and a list of replanning modules to invoke after each iteration.

For the initial implementation, the simulation was given the same configuration parameters used by Balmer (2007) in his study. The only parameter that was set to a different value, was the typical duration for work, which is used in the utility function calculation. In Balmer's case, this value is set to eight hours, compared to the nine hour workday for South Africa.

In the next chapter, we will reveal and discuss the results produced from these data.

Chapter 4

Results and discussion

Our primary measure of quality of simulation results is how accurately they compare with reality. In this case, we compare simulated traffic counts against the actual 2001 South African National Roads Agency Ltd. (SANRAL) counting station data from a selection of 20 pairs of network links in the Gauteng area. These stations are shown in Figure 4.1. They are mostly situated near important intersections on the main arterial routes in Gauteng. We consider results from the morning and afternoon traffic peaks, which were shown to occur between 05h00–08h00 and 14h00–17h00.

SANRAL traffic counts data vary with the day of the week. We assumed our demand to represent a "typical" workday, where the influence of weekend behaviour is minimal. We therefore compared simulated counts against the average hourly counts for a Wednesday, as Wednesdays lie exactly in the middle of the week, and the influence of weekend behaviour is arguably at a minimum. Besides SANRAL counts, we also examine vehicle departure and arrival times, trip durations and utility scores to compare the solution quality of various alternatives.

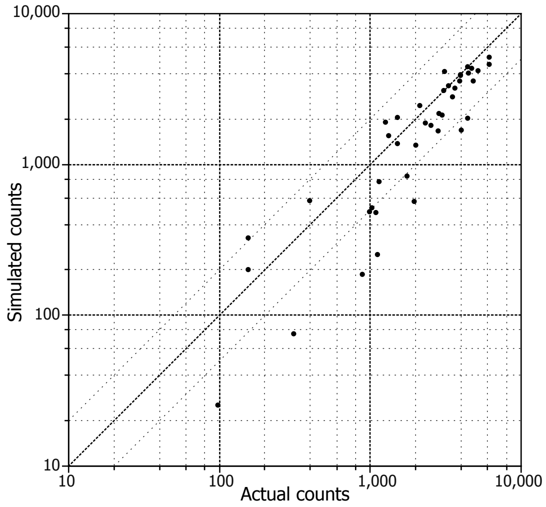
Results are organised as follows: we start by comparing the full network with the small network, in order to see if the small network, with its reduced computational footprint, is a suitable proxy for Gauteng's full road network. We then examine the repeatability and convergence of simulation results. The chapter concludes with an investigation into the quality of our initial demand.

4.1 The influence of network resolution

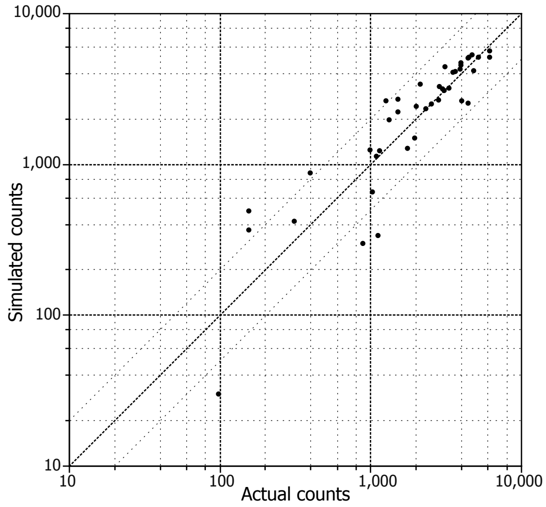
Figure 4.2 compares morning traffic with actual traffic counts during the peak hour of 06h00–07h00, as well as an hour before and after the peak period. The diagonal bands in each scatter plot indicate simulated versus actual counts ratios of 2:1, 1:1 and 1:2, starting from the topmost line. Counts ratios within these bounds are, for now, considered acceptable. Comparing the two network resolutions in Figure 4.2 leads to the following observations.

In general, the relative counting error becomes greater with decreasing link capacity. This

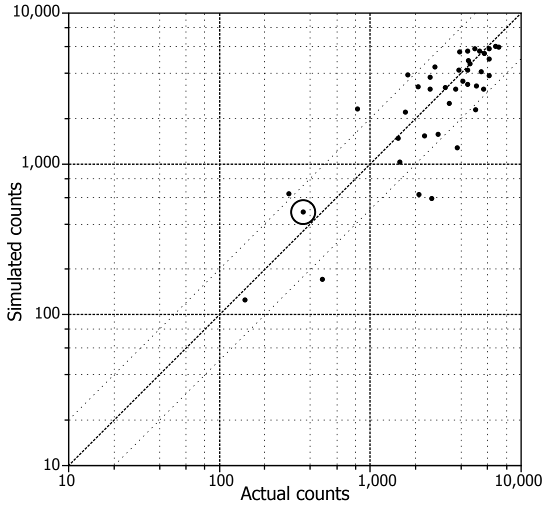
Figure 4.1: SANRAL counting stations selected for this study.



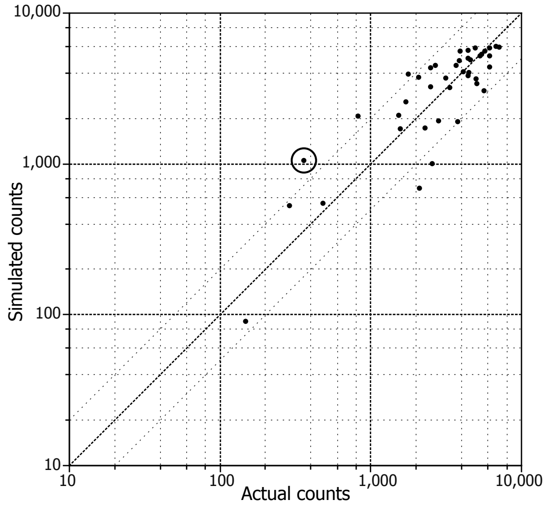
(a) Full network — 05h00–06h00



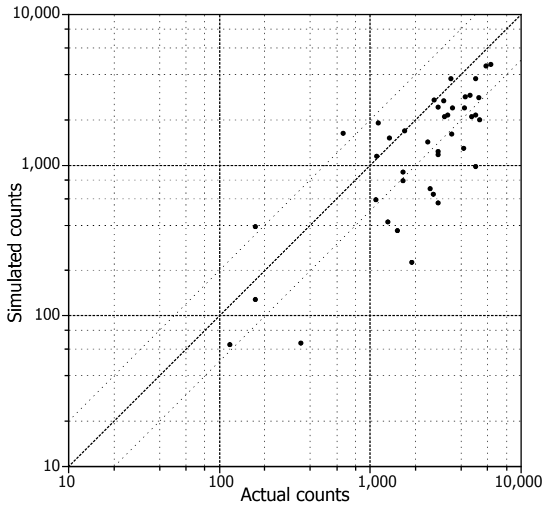
(b) Small network — 05h00–06h00



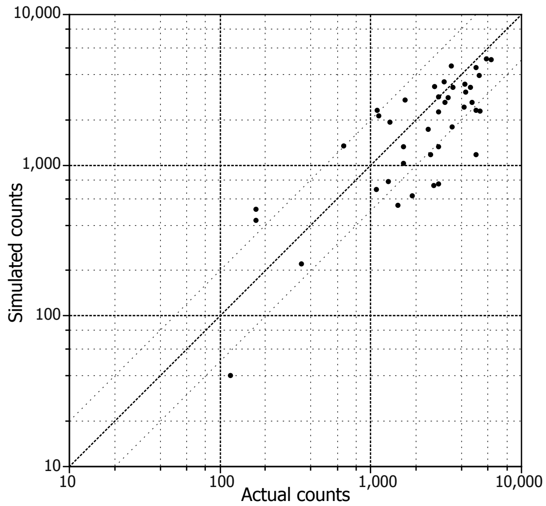
(c) Full network — 06h00–07h00



(d) Small network — 06h00–07h00

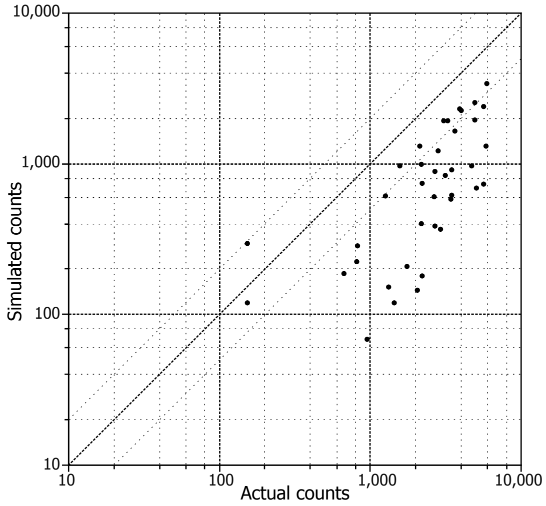


(e) Full network — 07h00–08h00

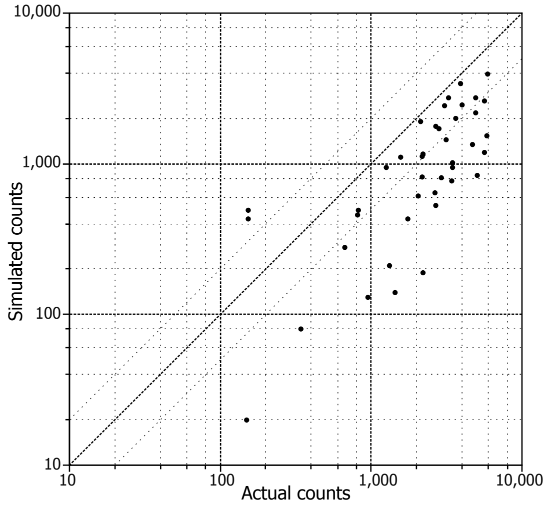


(f) Small network — 07h00–08h00

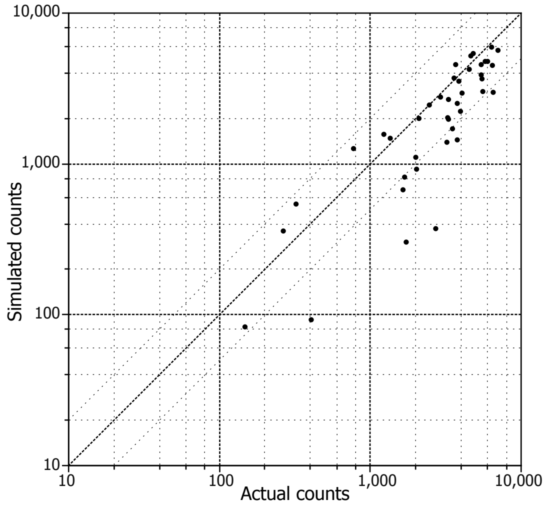
Figure 4.2: Counts comparison for the morning peak, 05h00–08h00. The left-hand column shows results from the full network, while the right-hand column is for the small network.



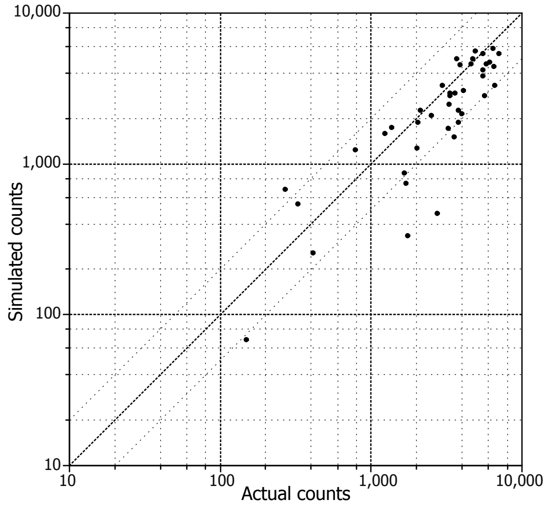
(a) Full network — 14h00–15h00



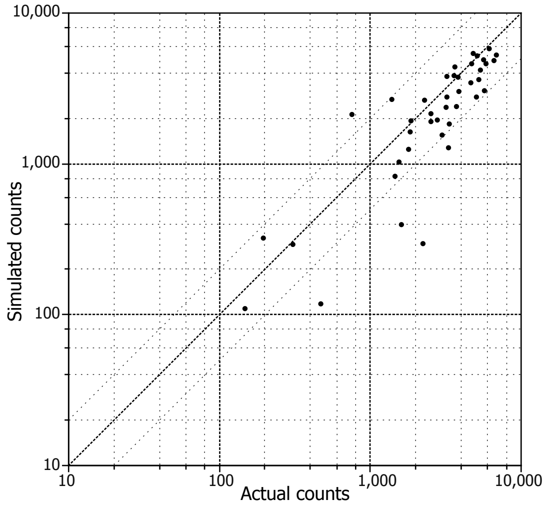
(b) Small network — 14h00–15h00



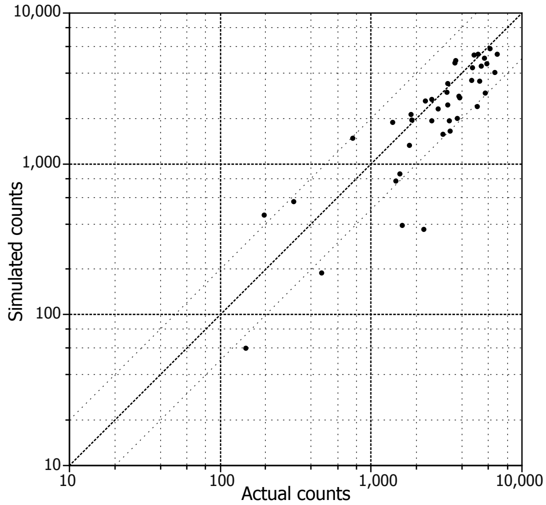
(c) Full network — 15h00–16h00



(d) Small network — 15h00–16h00



(e) Full network — 16h00–17h00



(f) Small network — 16h00–17h00

Figure 4.3: Counts comparison for the afternoon peak, 14h00–17h00. The left-hand column shows results from the full network, while the right-hand column is for the small network.

is not surprising, as the major arterials become saturated during the peak hour and reach full capacity in reality and simulation, thus yielding counts ratios close to 1:1. It should be noted that no link in either the full or small network representation has a capacity in excess of 6,000 vehicles per hour (2,000 vehicles per lane, up to a maximum of three lanes). In reality, however, a number of links show vehicle counts in excess of 6,000. These links are usually close to large interchanges, where the number of lanes increases due to off- and onramps merging with main roads.

Insofar as outliers are concerned, the over-counts for both the full network and the small network were found to occur at count stations 4, 6 and 10 in Figure 4.1. A number of factors were identified to be possible causes for these over-counts. All these counting stations are found on network links classified in the original Geographic Information System (GIS) file to be national highways, each with three lanes in either direction and a speed limit of 120 km/h. During network construction, a capacity of 6,000 vehicles per hour was assigned to national highway links. The majority of roads that feed into highways are classified to be major roads in the original GIS file, and were assigned a capacity of 2,000 vehicles per hour, and two lanes in either direction. In reality, these national highways were found to only have two lanes in either direction, and a large proportion of their feeder routes are ordinary roads with one lane in either direction, with an 80 km/h speed limit and an expected flow capacity of 1,000 vehicles per hour. These parts of the virtual networks therefore advertise contiguous stretches of road with far more capacity than what is available in reality.

The circled points in Figure 4.2c and Figure 4.2d represent the counting station traffic in an easterly direction at location number 4 in Figure 4.1. The full network shows a large reduction in error for this station, possibly due to a larger density of streets to absorb traffic in that area.

4.1.1 An alternative measure of counting station error

The usual measures of counting station error are defined as follows:

$$\text{Mean relative bias (\%)} \quad \bar{e} = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{4.1}$$

$$\text{where} \quad e_i = 100 \times \left(\frac{x_{sim,i} - x_{real,i}}{x_{real,i}}\right) \quad \text{for counting stations } i = 1, 2 \ldots n \tag{4.2}$$

$$\text{Mean relative error (\%)} \quad \overline{|e|} = \frac{1}{n}\sum_{i=1}^{n} |e_i| \tag{4.3}$$

with $x_{sim,i}$ and $x_{real,i}$ denoting simulated and real traffic counts for an hourly interval, recorded at counting station i.

Table 4.1 shows that the magnitude and spread of the relative counting error are smaller for the full network than for the small network through the entire morning peak. This trend is perhaps due to the fact that the full network consistently yields much smaller simulated counts on our observed links for the peak hours than the small network; a fact that logically follows from the higher density of links available to agents in the full network. If a network consistently yields

Table 4.1: Summary error statistics for traffic counts comparisons: morning peak, full network vs. small network.

	$\bar{e}$ (%)[a]		$\overline{\lvert e \rvert}$ (%)[b]		$\hat{\sigma}_e$ (%)[c]	
Time	Full	Small	Full	Small	Full	Small
05h00–06h00	-17	+17	34	37	26	44
06h00–07h00	+1	+15	37	39	37	41
07h00–08h00	-31	-8	49	49	31	39

[a] Mean relative bias. See (4.1).

[b] Mean relative error. See (4.3).

[c] Sample standard deviation for mean relative error. See (**??**).

under-counts across all links, then its mean relative error is, by definition, bound to a maximum of 100%, as no link can yield a negative simulated count. But over-counts are only bounded by the link's capacity, and therefore it is possible to find relative counting errors in excess of 100% for links that are under-utilised in reality.

Mean relative error, as expressed in (4.3) is therefore a skewed representation of error: it exaggerates the influence of a relative over-count compared to an under-count. The diagonal lines of constant ratio in the scatter plots of Figure 4.2 suggest that a simulated versus actual counts ratio of 1:2 is of equal importance than a ratio of 2:1. But, expressed as a relative counting error, the first ratio gives an error of 50%, compared to 100% for the second. (Note that it can be argued that, for our particular case, a relative over-count is a strong indication of a systematic error, compared to a relative under-count, as one would expect under-counts when modelling only a single activity chain and transport mode. It can be argued, therefore, that over-counts should make a greater contribution towards our overall measures of simulation error.)

Consider Figure 4.2e in comparison with Figure 4.2f. Our intuition tells us that the smaller network yielded results that are more consistent with actual vehicle counts because, on average, its points lie closer to the middle diagonal than the full network. The large number of under-counts recorded for the full network are reflected to some extent in its mean relative bias value in Table 4.1. But it shows a mean relative error of equal magnitude to the small network, and a *smaller* spread of error values, in stark contrast with the observed spread in the counts comparison scatter plot. We therefore suggest an additional error measure that is ratio-based, and indicates the degree of deviation from perfect agreement of simulated versus actual counts.

We define the counts ratio error, r_c, and its summary statistics as follows:

$$r_{c,i} = \begin{cases} \frac{x_{sim,i}}{x_{real,i}} - 1 & \text{for} \quad x_{sim,i} \geq x_{real,i} \\ -\frac{x_{real,i}}{x_{sim,i}} + 1 & \text{for} \quad x_{real,i} > x_{sim,i} \end{cases} \tag{4.4}$$

$$\overline{r_c} = \frac{1}{n} \sum_{i=1}^{n} r_{c,i} \tag{4.5}$$

$$\hat{\sigma}_{r_c} = \sqrt{\frac{1}{n} \sum (r_{c,i} - \overline{r_c})^2} \tag{4.6}$$

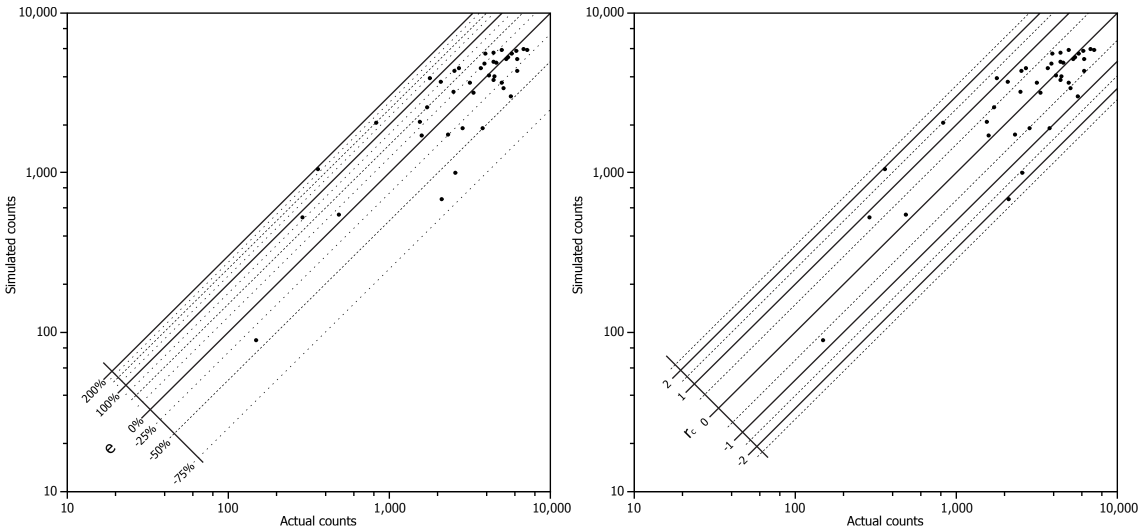
where $i \in \{1 \ldots n\}$ denotes the set of counting stations for which $x_{sim,i} > 0$ and $x_{real,i} > 0$.

Table 4.2: Summary counts ratio error statistics: morning peak, full network vs. small network.

	$\overline{r_c}$ [a]		$\hat{\sigma}_{r_c}$ [b]	
Time	Full	Small	Full	Small
05h00–06h00	-0.59	+0.04	1.09	0.82
06h00–07h00	-0.21	+0.05	0.92	0.72
07h00–08h00	-1.15	-0.46	1.64	1.08

[a] Mean counts ratio error. See (4.5).

[b] Sample standard deviation for counts ratio error. See (4.6).



(a) Relative bias scale.

(b) Counts ratio error scale.

Figure 4.4: Comparison of the scale of relative bias with that of our new counts ratio error metric (the scattered points are our previously stated results for the small network, 06h00–07h00).

Figure 4.4 compares the scales of relative bias with that of our new error metric. Clearly, the counts ratio error metric removes the bias present in the traditional error measure. Table 4.2

compares the statistics of the morning peak for the two network resolutions based on this new metric. As with the counts comparison scatter plots, we only consider counting stations that recorded simulated and actual count values larger than 10. Compare the values for 07h00-08h00: the average magnitude of the counts ratio error for the full network is -1.15 versus -0.46 for the small network. The sample standard deviation also shows a significantly larger spread for the full network, of 1.64 versus 1.08. Recall that the corresponding sample standard deviation for relative error in Table 4.1 actually showed a smaller spread of errors of 31% for the full network, compared with 39% for the small network, contrary to what was observed in that hour's counts comparison scatter plots. The counts ratio error metric therefore extends our insight beyond that of the mean relative error metric.

Table 4.3: Summary error statistics for traffic counts comparisons: afternoon peak, full network vs. small network.

	$\bar{e}$ (%)[a]		$\overline{\lvert e\rvert}$ (%)[b]		$\overline{r_c}$[c]		$\hat{\sigma}_{r_c}$[d]	
Time	Full	Small	Full	Small	Full	Small	Full	Small
14h00–15h00	-65	-45	69	64	-3.68	-2.21	3.65	2.61
15h00–16h00	-22	-12	35	35	-0.71	-0.44	1.33	1.07
16h00–17h00	-13	-13	34	36	-0.49	-0.44	1.27	1.05

[a] Mean relative bias. See (4.1).

[b] Mean relative error. See (4.3).

[c] Mean counts ratio error. See (4.5).

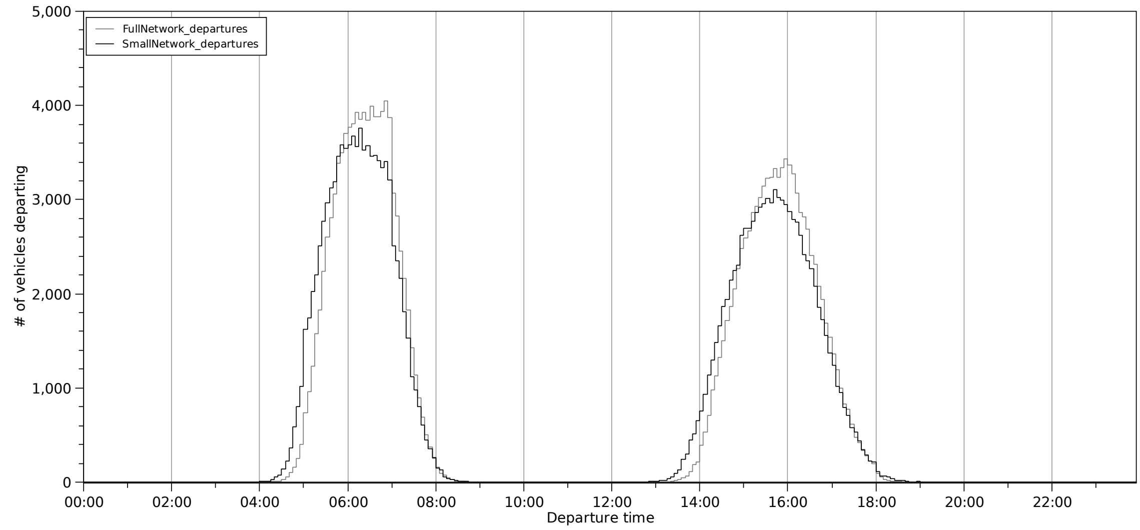
[d] Sample standard deviation for counts ratio error. See (4.6).

Table 4.3 shows the summary statistics of the afternoon peak for the two network resolutions. Both network resolutions produce under-counts, with larger under-counts in general occurring on the full network. In the next section, we examine the travel behaviour for both network resolutions in more detail, in order to understand how the differences in traffic counts arise.

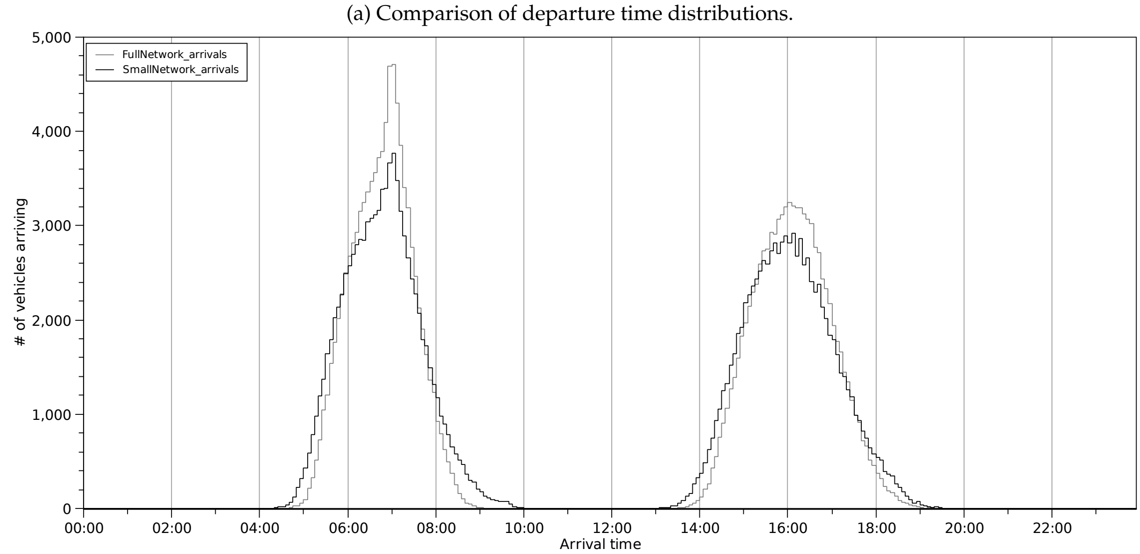
4.1.2 Travel behaviour and utility

Figure 4.5 compares the departure and arrival times during the morning and afternoon peak for the two network resolutions, as well as the time spent en route to home or work. From this series of charts can be seen that agents in the full network depart approximately 20 minutes later in the morning and afternoon. The full network also shows a spike of people arriving at work around 07h00, which is the opening time of the work activity. They also appear to spend less time travelling, as can be seen from the histogram showing the number of vehicles en route.

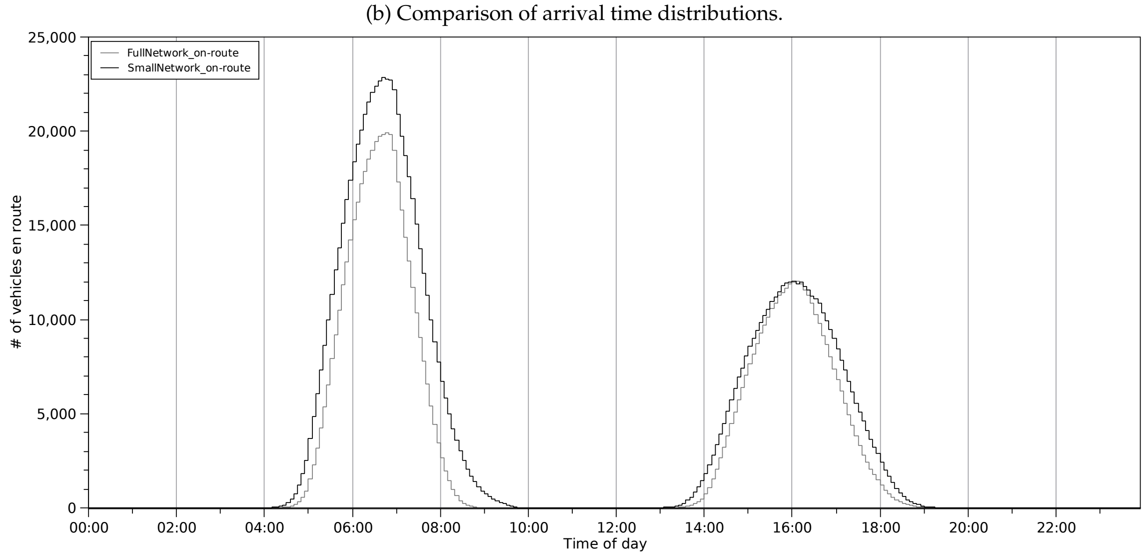
The larger difference between the two network resolutions in number of vehicles en route



(a) Comparison of departure time distributions.



(b) Comparison of arrival time distributions.



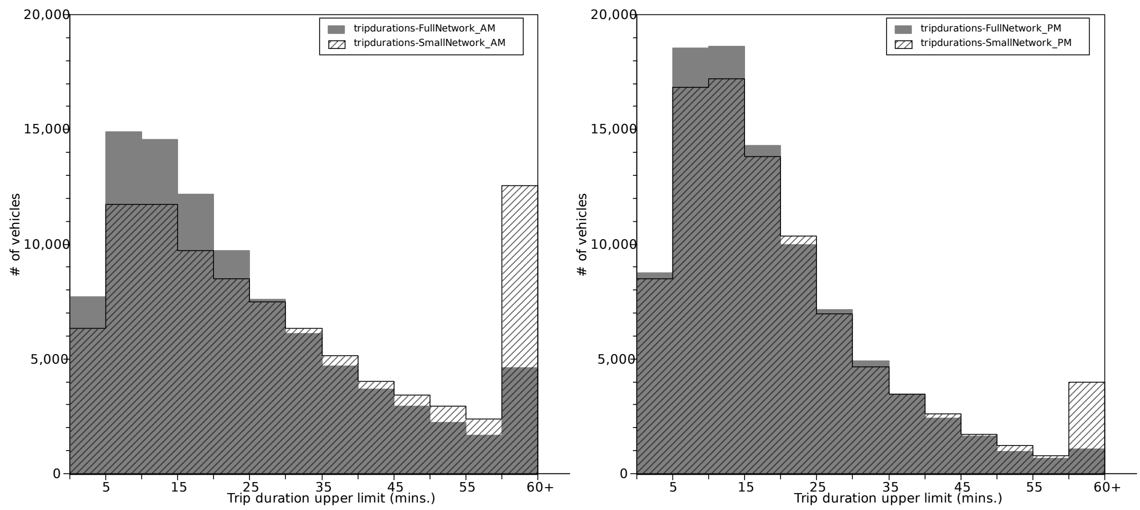
(c) Comparison of number of vehicles en route.

Figure 4.5: Comparison of departure and arrival times and time spent en route for the two network resolutions. Results from the full network are shown in grey, and those from the small network are in black.

for the morning peak accounts for the larger difference in counts comparison error metrics, as discussed in the previous section. During the afternoon, the two networks show similar numbers of vehicles en route. Consequently, their measures of counting error in Table 4.3 are of the same sign and magnitude.

One might speculate why the morning traffic peak is narrower and higher than the afternoon's. If we compare home locations with work locations in Figure 3.7 and Figure 3.8, we notice that home locations are more spread out through the province, while work locations tend to be concentrated in a number of areas. Therefore, as people converge on those areas of concentration, they tend to encounter dwindling network capacity. On the other hand, when traffic diffuses out from centres of economic activity in the afternoon, road users tend to encounter increasing link capacity the further they travel.

Another possible cause for the difference between the two peaks is that there is a penalty associated with arriving late at work in the morning, but none for arriving late at home in the afternoon. The only penalty for arriving late at home is the foregone utility of not being there. Consequently, agents need to make more constrained trade-offs between gaining utility from being at home and losing it for arriving late at work in the morning. Arguably, one might want to investigate the influence of a late penalty for the evening home activity, to simulate the influence of the various pressures, expectations and responsibilities of home life.

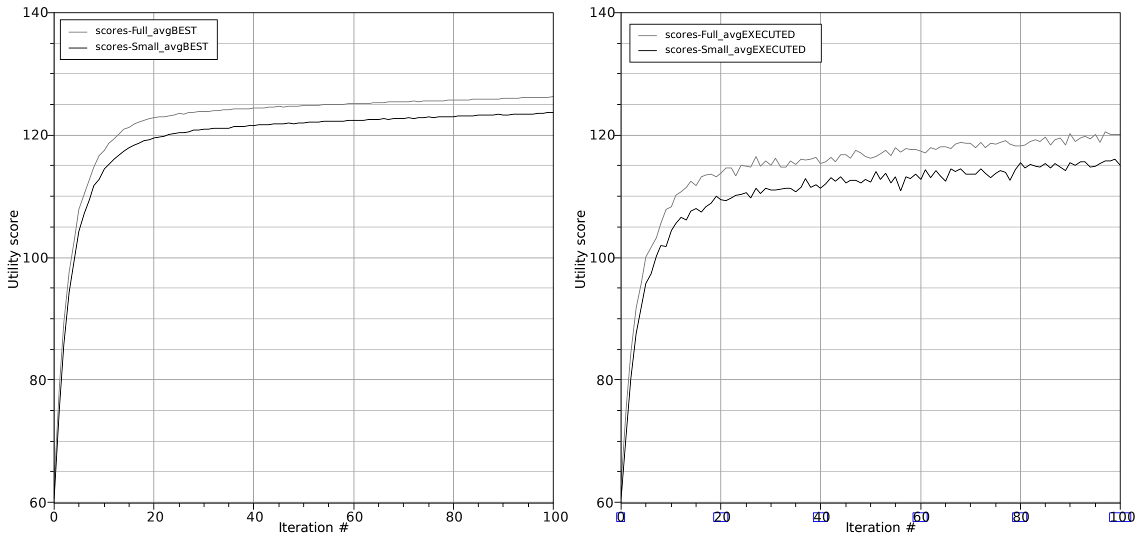


(a) Home to work (AM peak).

(b) Work to home (PM peak).

Figure 4.6: Comparison of trip durations for the two network resolutions. Results from the full network are shown in grey, while those yielded by the small network are shaded in black diagonal lines.

Figure 4.6 compares trip durations for the morning and afternoon peaks on both network resolutions. As expected, people generally take longer to travel in the morning. The full network representation also gives rise to shorter trip times. This is not surprising, as the available capacity density of the full network representation is higher by definition.



(a) Average best score vs. number of iterations.

(b) Average executed score vs. number of iterations.

Figure 4.7: Comparison of the evolution of the average best score and average executed score for the two network resolutions. Results from the full network are shown in grey and those from the small network are in black.

Finally, we compare the average utility values generated by the two network resolutions for a hundred iteration run in Figure 4.7. For both cases, the average best score and average executed score show a slightly upward trend. The full network representation gives an average improvement of approximately five units over the small network for executed plans and only about half as much for the average best score in memory. In both cases, the difference between the two networks amounts to less than five percent.

In all the results considered so far, the two network resolutions delivered similar results, such that all corresponding metrics are within the same order of magnitude throughout the morning and afternoon peaks. In fact, using the traditional metric of mean relative error, the small network differs by a maximum of only 2% for the morning peak hour of 06h00–07h00, and less than 1% for the 15h00–16h00 peak. However, as far as computational performance is concerned, the difference is immense. The full network requires at least 7.5 GB of memory to execute our 10% private vehicle driver population on a four core system at an average of 15′09″ per iteration. In comparison, the small network representation only requires 2 GB memory on the same system to execute the same sample at an average of 2′55″ per iteration. We therefore consider the small network to be a good enough description for cases where we need to perform exploratory work, such as in the further sections of this chapter. Use of the full network is reserved for decision support, when larger real monetary investments are considered.

4.2 Quality of transport demand

The difference in results from the previous section suggest that network topology influences the self-organisation of the simulated transport system. For instance, our commuter population departs, on average, at 06h12 in the morning when travelling on the small network, compared with 06h20 for the full network. We also saw that our results are similar to what is observed in reality; our mean relative error for the counting stations considered in this study comes to less than 40%. The question we try to answer in this section is: to what extent is the observed similarity of the system due to self-organising characteristics of the network, versus the demand for transport that gets imposed on it? Contemplation resolves this question to an even simpler one: just how accurate are our pairings of activity locations?

We have already identified counting station data as an upper bound to solution quality — if we simulate reality perfectly, then we should achieve exactly the same average vehicle counts as in reality. But the idea of a lower bound on solution quality is not so clear-cut. We therefore apply the following reasoning to test the quality of our transport demand.

Thus far, all our results evolved from a transport demand based on census and survey data. If we substitute our assigned home and work locations with uniformly distributed random locations within the province for each agent but keep all other parameters and simulation settings the same as for our reference case, we argue that the results represent a worst case scenario, or the equivalent of a lower bound to our solution.

Table 4.4: Summary error statistics for traffic counts comparisons: morning and afternoon peaks, NHTS- and census-derived demand ("Ref.") vs. random demand ("Rnd.").

	$\bar{e}$ (%)[a]		$\overline{\lvert e \rvert}$ (%)[b]		$\overline{r_c}$[c]		$\hat{\sigma}_{r_c}$[d]	
Time	Ref.	Rnd.	Ref.	Rnd.	Ref.	Rnd.	Ref.	Rnd.
06h00–07h00	15	112	39	126	0.05	1.07	0.72	2.72
15h00–16h00	-12	61	35	98	-0.44	0.41	1.07	1.97

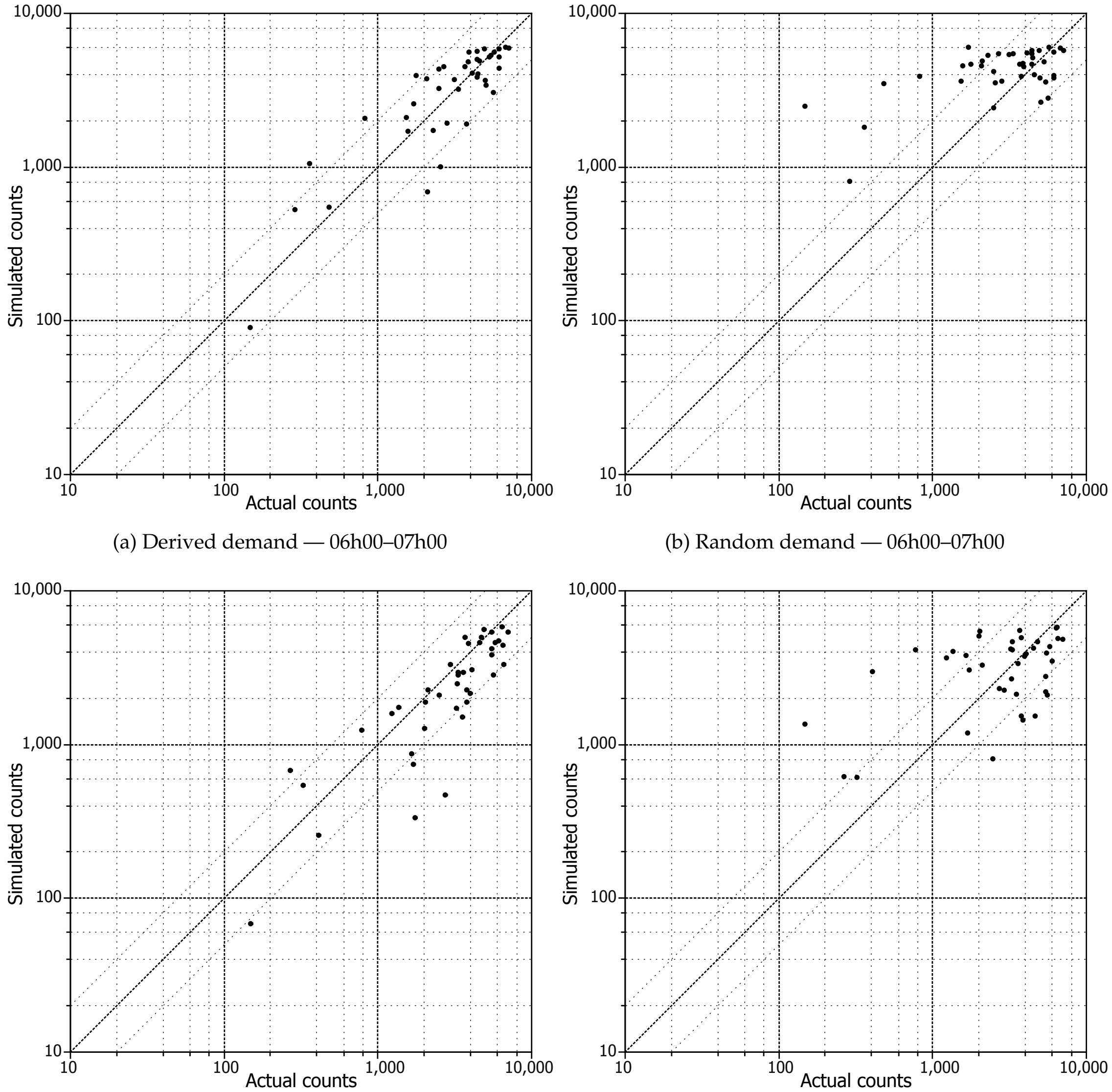
[a] Mean relative bias. See (4.1).

[b] Mean relative error. See (4.3).

[c] Mean counts ratio error. See (4.5).

[d] Sample standard deviation for counts ratio error. See (4.6).

For this purpose, a synthetic population of the same sample size was constructed, and their home and work locations uniformly distributed across the province. The results from this random demand are compared with those of our census- and survey-based demand in Figures 4.8 to 4.10 and Table 4.4. Figure 4.11 compares the error for 06h00–07h00 using the traditional mean relative bias measure, and our new counts ratio error metric, demonstrating the larger spread and shift



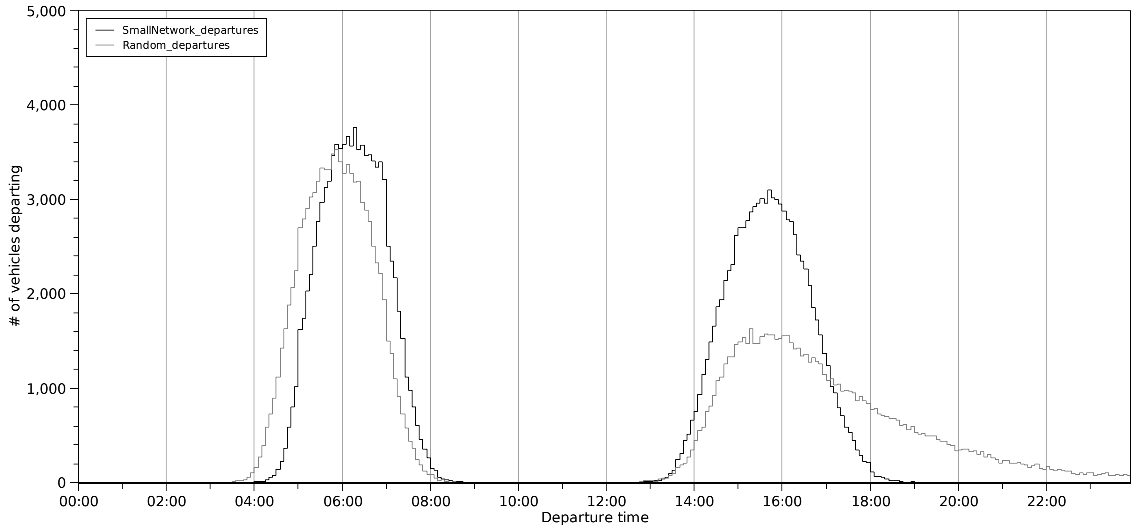
(a) Derived demand — 06h00–07h00

(b) Random demand — 06h00–07h00

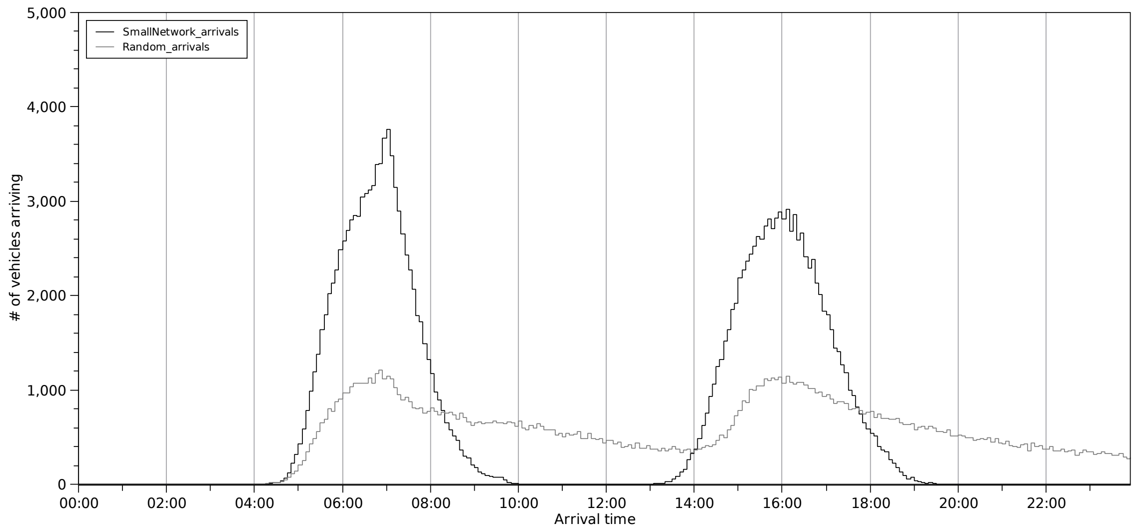
(c) Derived demand — 15h00–16h00

(d) Random demand — 15h00–16h00

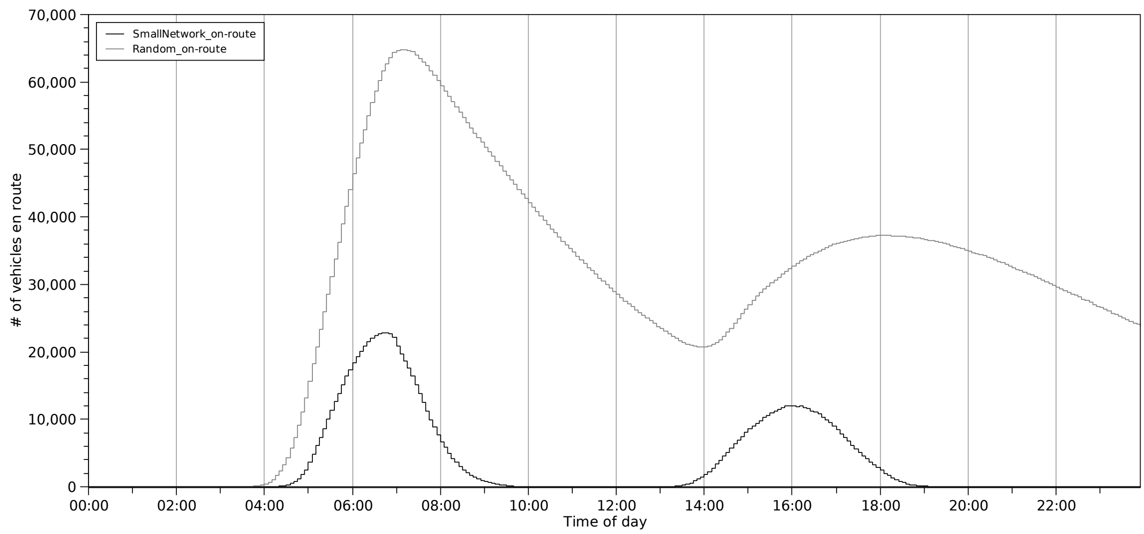
Figure 4.8: Counts comparison for the morning and afternoon peaks on the small network. The left-hand column shows results for our census- and NHTS-derived demand and the right-hand column those for a random demand.



(a) Comparison of departure time distributions.



(b) Comparison of arrival time distributions.



(c) Comparison of number of vehicles en route.

Figure 4.9: Comparison of departure and arrival times and time spent en route for the census- and NHTS-derived demand versus a random demand.

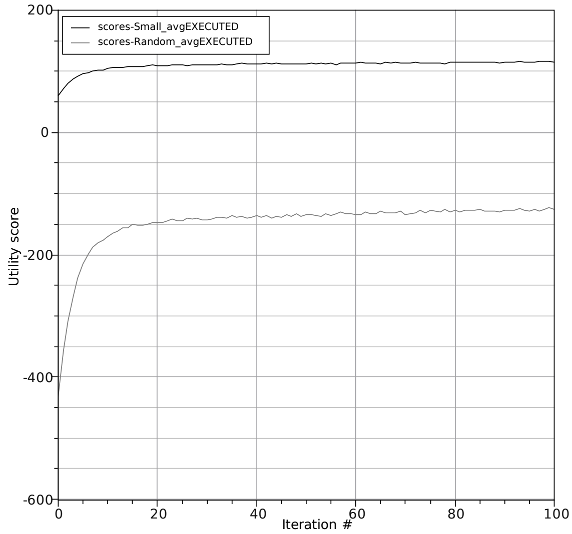
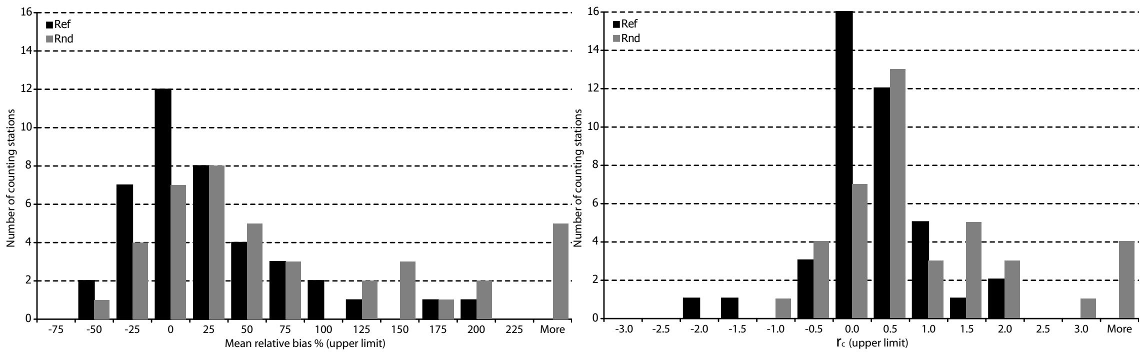


Figure 4.10: Comparison of evolution of utility scores for the census- and NHTS-derived demand (black) versus a random demand (grey).

towards over-counts present in the random sample.



(a) Relative bias comparison: Ref. vs Rnd.

(b) Counts ratio error comparison: Ref. vs Rnd.

Figure 4.11: The two histograms compare the relative bias and counts ratio error distributions of our census- and survey-derived demand (Ref.) with the random demand (Rnd.). Note how the counts ratio error distribution removes the bias towards under-counts that is present in the relative bias comparison.

These results are obviously considerably worse than those from our census- and survey-based demand. We can therefore conclude that our model presents a reasonable first approximation of the primary activity-driven transport demand of the private vehicle driver population of 2001.

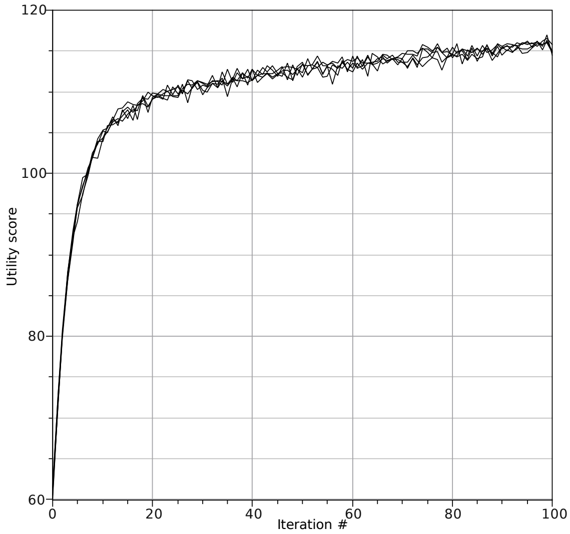


Figure 4.12: Comparison of utility scores for five simulation runs, each with a unique random seed used in population synthesis, and another unique seed used during simulation. All runs executed on the small network.

4.3 Repeatability

Figure 4.12 shows the utility score evolution from five repeat runs of 100 iterations each, each with unique seed values for the Java pseudo-random number generator used in population synthesis and simulation execution. The simulation is clearly highly repeatable as all values lie within 1% from one another. No appreciable distinction could be seen in counts comparisons, departure and arrival histograms or trip duration comparisons.

4.4 Convergence

The continuing upward trend in utility score graphs, such as Figure 4.7 and Figure 4.12, suggests that the process of systematic relaxation is still continuing. In this section, we ran our simulation for a further 400 iterations and compared results with those from 100 iterations, to see if any significant differences show up.

Figure 4.13 shows the utility score results for this extended run. The system appears to converge after 400 iterations, an observation that is confirmed by all other result reports. Even though utility scores only improve by approximately 10 units over 300 further iterations, the influence on system behaviour is dramatic. Figure 4.15 compares peak traffic counts for 100 iterations with those after 500, and Table 4.5 summarises the deviation from actual counts. The table shows that, both for the morning and afternoon peaks, the long simulation run produces large under-counts.

The longer simulation run delivers shorter trip durations, especially during the morning peak, as can be seen from Figure 4.14. This change in trip duration distribution is probably due to trip

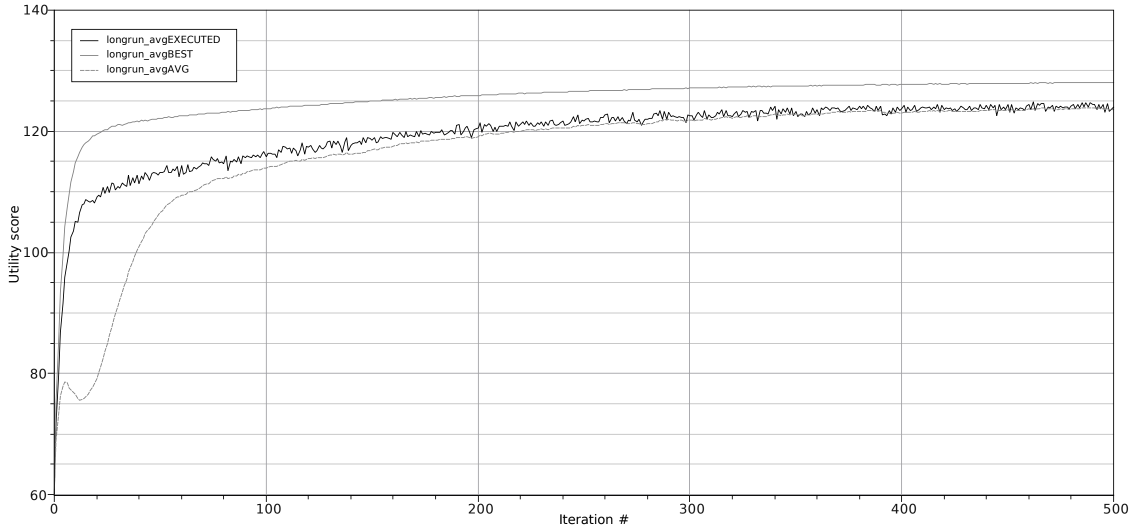
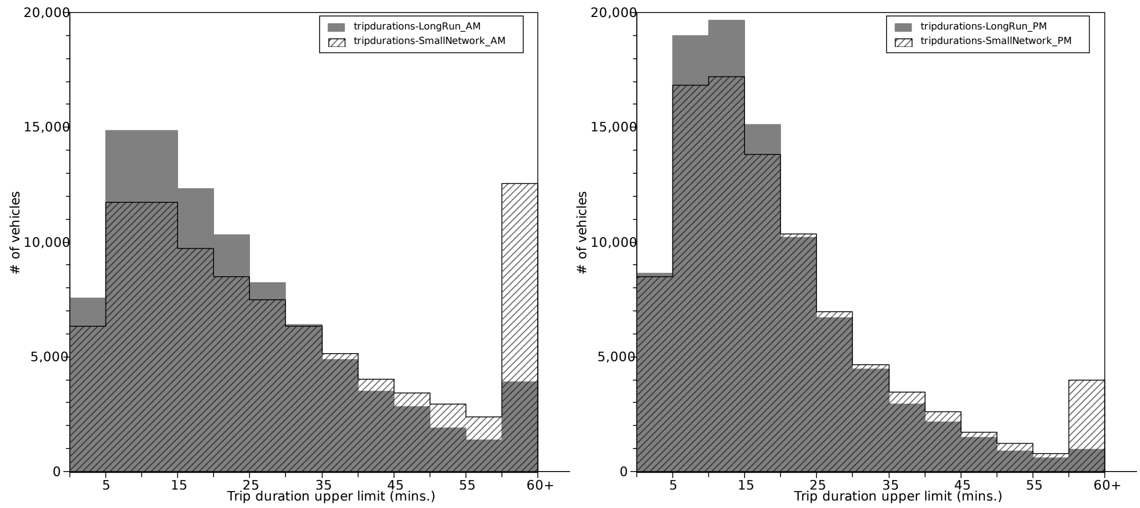


Figure 4.13: Utility score evolution: 500 iteration run.

timings, shown in Figure 4.16. Here we can see that individuals have adjusted their departure times in such a way that they arrive at work en masse between 07h00 and 09h00. These are, respectively, the opening time and latest starting time for the work activity. Before 07h00, agents don't gain utility from being at work, and after 09h00, they are severely penalised for being late. In the afternoon, network loading proceeeds gradually up to 18h00, after which agents don't gain any more utility from being at work. As a consequence, the afternoon traffic peak moves two hours later.

These results are a dramatic illustration of emergence and self-organisation as only an agent-based simulation can replicate. Purely by observing a number of simple rules, the entire system has organised itself into a highly efficient state, without any of the constituent entities "consciously" aiming for system optimisation. The opening and closing times of the work activity are set at 07h00 and 18h00 respectively, and the simulation clearly optimises itself until these times become limiting constraints. In reality, these times are variable quantities, and a certain amount of peak-spreading will be due to flexible operating hours.

Whether the ruthless optimisation we observe in these results is attainable in reality is debatable, as we are modelling only a segment of the population travelling on a network where absolutely nothing changes over the course of 400 days, except for the activity timing of other agents in the system. In reality, we find much more variablity in the transport system. Furthermore, the agent population shows no variability in their behaviour and goal-seeking imperative, unlike human beings that have variable personalities, preferences, responsibilities and other dissimilarities. Therefore we conclude that a simulation run of 100 iterations compares better with actual system behaviour than a 400+ iteration run.



(a) Home to work (AM peak).

(b) Work to home (PM peak).

Figure 4.14: Comparison of trip durations for the morning and afternoon peaks. Results from the 500 iteration run are shown in grey, while those yielded after 100 iterations are shaded in black diagonal lines.

Table 4.5: Summary error statistics for traffic counts comparisons: morning and afternoon peak hours, 100 vs. 500 iterations.

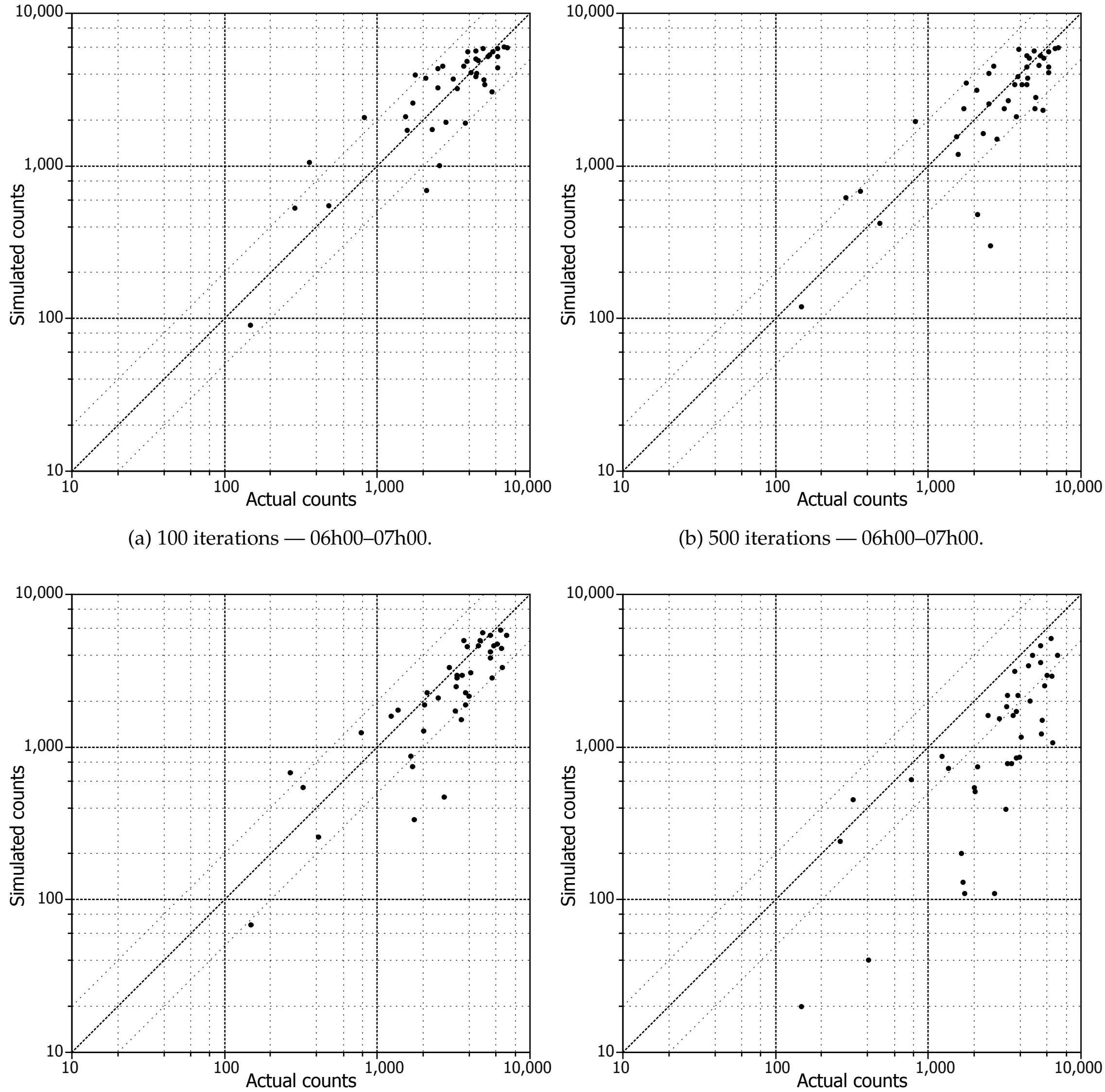
	$\bar{e}$ (%)[a]		$\overline{\lvert e \rvert}$ (%)[b]		$\overline{r_c}$[c]		$\hat{\sigma}_{r_c}$[d]	
Time	100	500	100	500	100	500	100	500
05h00–06h00	17	-51	37	51	+0.04	-2.84	0.82	4.46
06h00–07h00	15	0	39	36	+0.05	-0.3	0.72	1.37
07h00–08h00	-8	33	49	60	-0.46	+0.18	1.08	1.16
14h00–15h00	-45	-80	64	80	-2.21	-14.39	2.61	19.88
15h00–16h00	-12	-54	35	56	-0.44	-3.18	1.07	4.62
16h00–17h00	-13	-21	36	39	-0.44	-0.74	1.05	1.59

[a] Mean relative bias. See (4.1).

[b] Mean relative error. See (4.3).

[c] Mean counts ratio error. See (4.5).

[d] Sample standard deviation for counts ratio error. See (4.6).



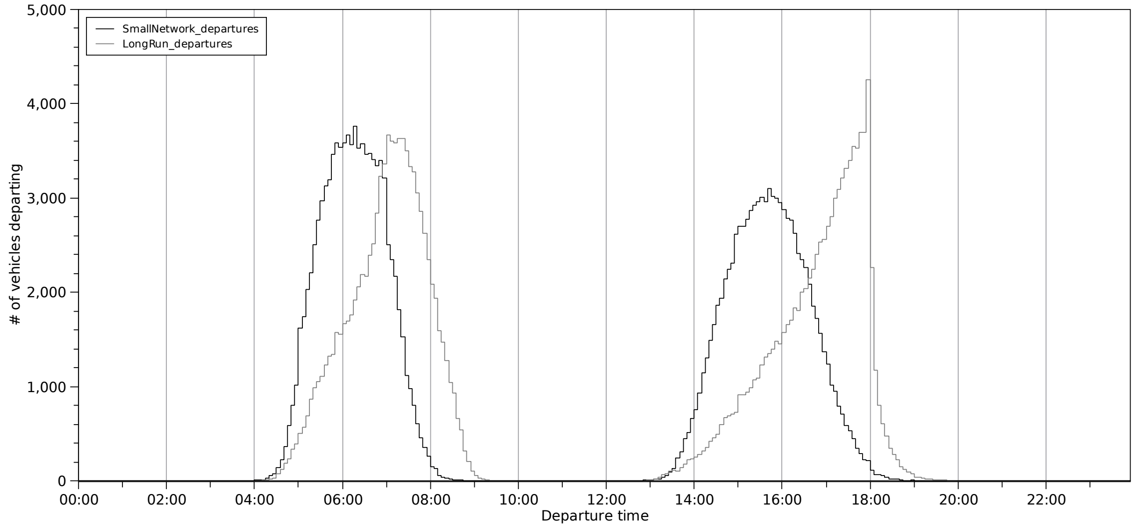
(a) 100 iterations — 06h00–07h00.

(b) 500 iterations — 06h00–07h00.

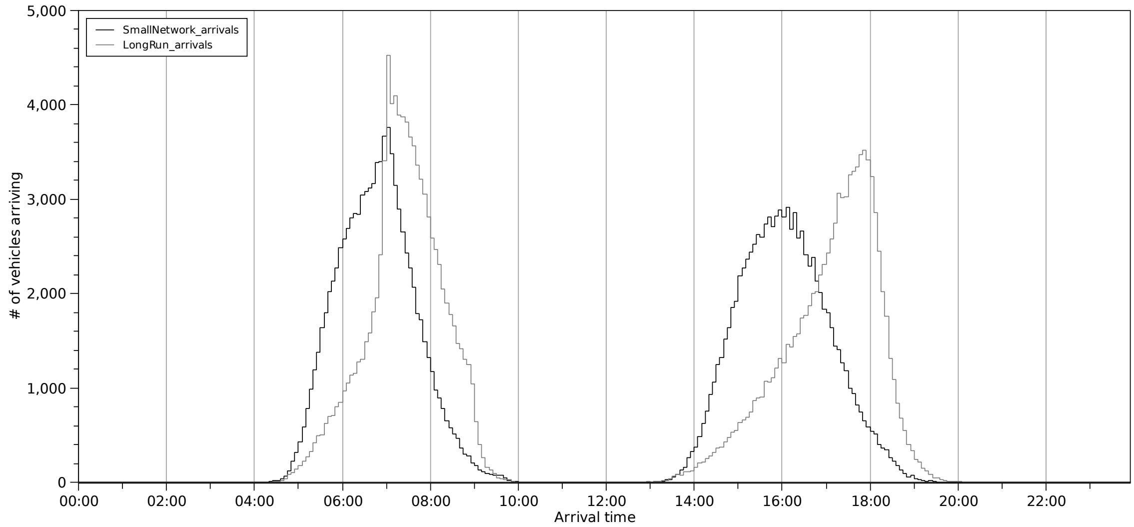
(c) 100 iterations — 15h00–16h00.

(d) 500 iterations — 15h00–16h00.

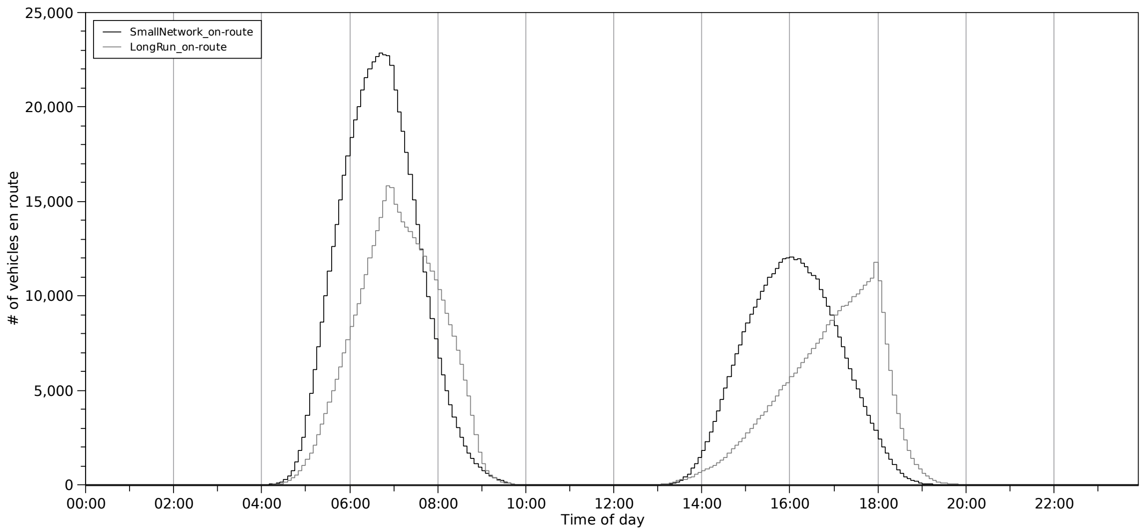
Figure 4.15: Counts comparison for the morning and afternoon peak on the small network. The left-hand column shows results after 100 iterations and the right-hand column those after 500 iterations.



(a) Comparison of departure time distributions.



(b) Comparison of arrival time distributions.



(c) Comparison of number of vehicles en route.

Figure 4.16: Comparison of departure and arrival times and time spent en route after 100 iterations (black) and 500 iterations (grey).

Chapter 5

Conclusion

In this chapter we attempt to answer the main research question from Section 1.6, and summarise our findings.

Our initial implementation of the Multi-Agent Transport Simulation Toolkit (MATSim), with a transport demand generated from the private vehicle driver population of the 2001 census and 2003 National Household Travel Survey (NHTS), and network representation derived from Geographic Information System (GIS) data, manages to predict the peak traffic flow patterns on a typical Wednesday in 2001 with a minimum average error of less than 40%. This error is significantly less than the minimum average error of 97% for a completely random home-work-home demand of the same population size.

In general, our implementation is biased toward under-counts, with larger under-counts registered for a full network representation than one representing only major routes. The magnitude of network under-utilisation was shown to increase with increasing number of iterations, with no significant change observable after 400 iterations. The simulation also proved to be highly repeatable and not sensitive to minor changes to initial conditions due to stochastic effects.

5.1 Sources of error and bias

As our initial implementation only models a particular segment of the total transport user community, and only a single activity chain, a significant degree of under-utilisation was expected across the entire network. However, counts comparisons for both the full and small network representations showed large over-counts for a number of counting stations.

These outlier results proved diagnostic in identifying errors in our network representation. Some feeder routes to major freeways were found to have larger capacities than in reality, making them attractive as alternative routes for the virtual commuter population. Consequently, the affected freeways registered larger simulated counts than in reality. These network errors suggest that capacity improvements to feeder routes might improve utilisation of freeways in reality, and divert some of the load from over-utilised network links.

Major under-count outliers were found in areas of high network density and, consequently, available capacity, especially during the afternoon peak, when commuters move away from centres of economic activity.

5.2 Outlook

Our initial implementation of MATSim for Gauteng has already given rise to a number of further studies in our research group. These studies aim to improve the initial implementation by addressing some of the shortcomings identified in this study.

5.2.1 Improved initial demand

The allocation of home- and work location in this study is an extrapolation of census and travel survey data. Such an assignment is unrealistic, because it depends on the poor resolution of the Traffic Analysis Zones (TAZs) used during the travel survey as well as a very small, outdated sample. The model also fails to produce variation due to personal demographic characteristics relating to the primary activity location decision. A master's degree study conducted by J. Niesing will aim to overcome these shortcomings, as well as introduce secondary activities into our daily activity schedules for the private vehicle commuter population. As the next census is only scheduled for 2011, Ms Niesing will also attempt to update the synthetic population's characteristics using annual household survey data.

5.2.2 Commercial traffic

Commercial traffic contributes significantly towards Gauteng's road transport demand, especially during off-peak hours. But modelling the behaviour of commercial traffic is no trivial task, as the drivers for this transport demand are exactly the same as those that contribute to the fickleness and unpredictability of world stock markets. A PhD study, conducted by M. Botha, will be a novel effort towards an activity-based model of commercial transport demand for South Africa. Ultimately, we aim to adapt MATSim to execute the commercial demand alongside that of the commuting population, and capture all possible interactions that might arise.

5.2.3 Public transport and para-transit modes

Public transport and the para-transit minibus taxi mode contribute significantly towards network utilisation. The author's proposed PhD study will be an investigation into the behaviour of the unscheduled minibus taxi mode, both from the perspective of user and provider. The MATSim core will be adapted to simulate this mode. Our implementation for Gauteng will be extended to include minibus para-transit, as well as scheduled public transport modes, such as Bus Rapid Transit (BRT), the various bus operators in Gauteng, and the Gautrain rapid rail transit system. Such an implementation will require that agents select from, and optimise across multiple modes. M. Moyo in Berlin is currently working on extending MATSim's core functionality to accommodate public transport and travel legs using multiple modes. Ultimately, it is envisioned that agents be able to change and improve their mode choice during the course of a day.

Besides these active projects, a number of further studies have been identified that are expected to improve our initial implementation.

5.2.4 Improved network representation

Results from this study showed that errors in our network representation are a probable cause of counting station error. It was suggested that a Computer Science Masters degree study be scoped to derive a more detailed and accurate network representation from aerial photography and GIS data. Our GIS shapefiles record the centre lines of all roads in the Gauteng road network. These centre lines can serve as guides to an image recognition program, designed to interpret aerial photography data and correctly assign network link capacities.

5.2.5 Improved mapping of land-use

In order to improve our activity-based travel demand, and extend our model to include secondary activities, we require detailed land-use information. Modern techniques such as dasymetric mapping can be applied to establish activity locations and number of activity opportunities based on physical attributes of the urban environment, such as building type and size. These attributes will have to be identified by a suitably developed image recognition application, capable of interpreting aerial and satellite photography.

5.2.6 Extending the scope beyond Gauteng

Recently acquired satellite tracking data from approximately 30,000 commercial and freight vehicles suggest that a large proportion of Gauteng's commercial traffic is destined for, and originates from outside the province's boundaries. This study also only considers private vehicle commuter demand that is completely contained inside the province, an assumption that is refuted by census and travel survey data. Our model should be extended to include demand for all modes, originating from and destined for the rest of South Africa and, possibly, the greater sub-continent.

5.2.7 The influence of road pricing

Results from this study showed significant over-counts for both traffic peaks for the counting stations at location 4 in Figure 4.1. Although excess capacity on feeder routes was identified as a possible cause of an increased utilisation of this national highway, it should also be noted that this network link, like several others in the province, is tolled. Therefore, toll avoidance could be a contributing factor towards its under-utilisation in practice.

The Gauteng roads authorities have announced an aggressive distance-based road pricing strategy on almost all major routes in the province, scheduled for roll-out in 2009-10. Modelling the influence of toll is already part of MATSim's core functionality, and it is expected that an implementation of this module for Gauteng will contribute significantly towards improving our current model, as well as aid in prediciting the effects of the proposed strategy.

We do not purport this list of improvements to be comprehensive. Continued investigation and improvement of this first implementation and our involvement with MATSim's active development community is constantly giving rise to interesting and worthwhile research opportunities. As an open-source initiative, anybody can participate, and user-developers are encouraged to explore and modify the code. This 'bottom-up' approach of constant tinkering sometimes leads to unexpected applications and new directions of enquiry. Therefore MATSim not only models emergent phenomena — it is an emergent phenomenon itself.

Bibliography

Axhausen, K. W. and Gärling, T. (1992). Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews*, 12:323–341.

Balmer, M. (2007). *Travel Demand Modeling for Multi-Agent Transport Simulations: Algorithms and Systems*. PhD thesis, ETH Zürich.

Balmer, M., Axhausen, K., and Nagel, K. (2006). Agent-based demand-modeling framework for large-scale microsimulations. *Transportation Research Record*, pages 125–134.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(90003):7280.

Bowman, J. (1998). *The day activity schedule approach to travel demand analysis*. PhD thesis, MIT.

Charypar, D. and Nagel, K. (2005). Generating complete all-day activity plans with genetic algorithms. *Transportation*, 32(4):369–397.

Diedericks, D. and Joubert, J. (2006). Towards transportation system integration in the City of Tshwane Metropolitan Municipality. In Brebbia, C. and Dolezel, V., editors, *Urban Transport XII: Urban Transport and the Environment in the 21st Century*, pages 795–804. WIT Press.

DOT (1996). White Paper On National Transport Policy. Technical report, Department of Transport.

Engelbrecht, A. (2002). *Computational Intelligence: An Introduction*. J. Wiley & Sons.

Ettema, D. (1996). *Activity-Based Travel Demand Modeling*. PhD thesis, Technische Universiteit Eindhoven.

Lawson, C. (2006). Microsimulation for urban transportation planning: Miracle or mirage? *Journal of Urban Technology*, 13:55–80.

Manson, N. (2006). Is operations research really research? *ORiON*, 22(2):155–180.

MATSim Development Team (2008). MATSim (Multi-Agent Transport Simulation Toolkit).

McNally, M. (2000). The Activity-Based Approach.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55:31–42.

Meyer, M. and Miller, E. (2001). *Urban Transportation Planning - A Decision-Oriented Approach*. MCGraw-Hill.

Nagel, K. (1995). *High-speed microsimulations of traffic flow*. PhD thesis, Universityof Cologne, Koln, Germany.

Pas, E. I. (1996). Recent advances in activity-based travel demand modeling. *Proceedings of Activity-Based Travel Forecasting Conference*.

Peeta, S. and Ziliaskopoulos, A. K. (2001). Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, 1:233–265.

Rieser, M. (2007). Simulating road pricing/tolls with MATSim.

Stats SA (2001). Census 2001 metadata: Geography hierarchy and attributes. Technical Report 03-02-25 (2001), Statistics South Africa.

USGS (2001). Universal Transverse Mercator Grid, Fact Sheet 077-10.

Vaishnavi, V. and Kuechler, B. (2005). Design Research in Information Systems.

Weiner, E. (1997). *Urban Transportation Planning in the United States: An Historical Overview*. U.S. Department of Transport.