

A large scale combined private car and commercial vehicle agent-based traffic simulation [☆],^{☆☆}

J.W. Joubert^{a,b,c,*}, P.J. Fourie^a, K.W. Axhausen^{c,1}

^a*Industrial and Systems Engineering, University of Pretoria, South Africa, 0002*

^b*CSIR Built Environment, PO Box 395, Pretoria, South Africa, 0001*

^c*Institute for Transport Planning and Systems (IVT), ETH Zurich, 8093, Zurich, Switzerland*

Abstract

The number of independent and interdependent freight actors (firms), the complex supply chain structures among them, and the sensitivity of shipment data are but a few reasons why modeling freight traffic is lagging its public and private transit counterparts. In this paper we used an agent-based approach to reconstruct commercial activity chains, and simulated them—along with private vehicles—for a large-scale scenario in Gauteng, South Africa. The simulated activities are compared to the actual observed activities of 5196 vehicles that were inferred from GPS logs covering approximately six months. The results show that the activity chains reconstructed are both spatially and temporally accurate, especially in areas of high activity density. With freight vehicles being a major contributor to traffic congestion and emissions, our contribution is significant in bridging the gap between the person and commercial transport modeling state-of-the-art.

Keywords: MATSim, transport demand planning, agent-based transport simulation, commercial, traffic simulation

1. Introduction

Although freight plays an important role in economic development, freight demand modeling has received much less research attention than its passenger counterpart (Ortúzar and Willumsen, 2001). Possible explanations include the many actors (firms) involved; the complex supply chain structures that result in in- and interdependent logistic networks among the firms (Hesse and Rodrigue, 2004); the numerous (and variety) of commodities concerned; and the confidentiality and reliability issues with regards to data gathering.

Rodrigue (2006) notes that freight and passengers have very different spatial dynamics. Yet, they share the same road space. The number of freight vehicles in urban areas may be comparatively small to total vehicular traffic, but each vehicle has a disproportionate contribution to certain emissions and traffic congestion.

Decision-making about freight is usually viewed in literature from the firm perspective (Crainic and Laporte, 1997): the emphasis is on achieving optimal or near-optimal performance with regards to its economic efficiency and service quality. At a strategic level, for example, decisions are made regarding the location of main facilities (production or distribution) and the procurement of resources such as trucks and rolling stock. Tactical decisions may include the design of the service network through route choice and geographic customer segmentation. Operational decisions are typically highly

[☆] Author: jwjoubert; Revision: 133; Last changed on: 2011-01-26 15:53:47 +0200 (Wed, 26 Jan 2011)

^{☆☆} A formal published version of this paper appears in *Transportation Research Record, Volume 2168*

*Corresponding author

Email addresses: johan.joubert@up.ac.za (J.W. Joubert), pieter.fourie@up.ac.za (P.J. Fourie), axhausen@ivt.baug.ethz.ch (K.W. Axhausen)

dynamic and involves decisions such as the scheduling of drivers and crews, and the dispatching of vehicles. Companies most often make these decisions within constrained environments.

Transport planning for central decision-making stakeholders, such as government, takes on a different and aggregated perspective. Their decisions need to reflect the simultaneous actions of very many transport players. Friedrich et al. (2007) identify senders (consignors), freight forwarders, carriers, drivers and receivers at the different decision-making levels in freight transport. They discuss the diverse process steps and decisions that are taken at the various levels, and show how choice sets differ for nong-haul and regional freight transport.

The problem the authors attempt to address in this paper is simulating disaggregate commercial vehicle chains effectively in both space and time. In Joubert and Axhausen (2009) we laid the foundation for understanding commercial vehicle movement and activities. In this paper we build on that foundation, generating intra-provincial commercial vehicle chains for vehicles performing at least 90% of their activities within the study region in Gauteng, South Africa. Spear et al. (2008) refer to these chains as *internal distribution* trips and distinguish them from *through* trips and *internal-external* trips. We simulate the chains in a *Multi-Agent Transport Simulator* (MATSim) toolkit along with private vehicle traffic obtained from Fourie and Joubert (2009). In the process we accounted for the majority of road users, with the exception being the multitude of small paratransit vehicles and through traffic.

To evaluate the time and space accuracy of the simulated activities, we performed an hour-by-hour comparison with the actual observed activity densities for each of the 457 mesozones as described by the Geospatial Analysis Platform (GAP) developed by CSIR Built Environment (2009).

We show that our approach to generate commercial vehicle activity chains produces very accurate results, especially in areas of high activity density. This paper is significant in a number of respects. From a commercial vehicle perspective we accurately model the detailed movements of commercial vehicles without having to model the complex and interdependent relationships among commercial actors such as production facilities (shippers), carriers and receivers. From a wider transport modeling perspective we have shown significant improvement to the private car models with the introduction of commercial vehicles.

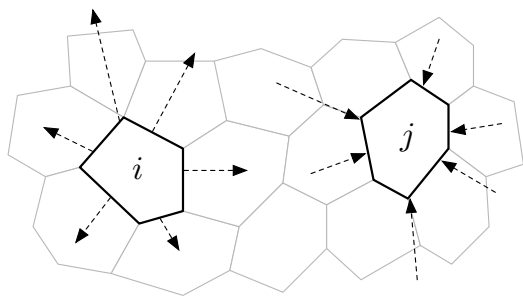
The paper is structured as follows: a brief review of transport modeling and related research in the field is provided in Section 2, along with an introduction to *MATSim*. Initial demand generation, the first step in the simulation process, is discussed in Section 3. The results, presented in Section 4, are discussed mainly from the commercial vehicle point of view. Finally, we end the paper with a conclusion and a brief research agenda.

2. Freight transport modeling

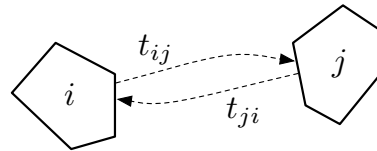
The majority of freight transport models are derived from the classical four-step model originally developed for passenger transport. A detailed review of freight transport models, all derived from the classical approach, can be found in De Jong et al. (2004).

We cast our work against the framework proposed by Tavasszy (2008). The framework is illustrated in Figure 1 and extends the four-step modeling approach to account for decisions and issues relevant to freight modeling.

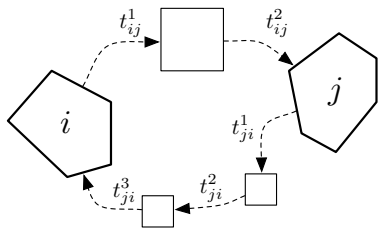
Production and consumption volumes (in tons) are generated for each location, typically a zone, using land-use and transport interaction; trip generation; facility location; and various freight-economy coupling approaches. Tapio (2005) warns that traditional coupling of freight traffic volumes and Gross Domestic Product (GDP) should be considered with caution. Trade values are converted to volume to establish commodity flows, linking locations. Logistics involve choices regarding inventory location and supply chain management. Transportation is challenged with modal choice, intermodal and transshipment decisions. Discrete choice techniques and trip conversion factors are often employed. At this stage, tons are often converted to ton-kilometers, or even to the individual consignment level. Kveiborg and Fosgerau (2007) suggest a distinction between freight traffic, when vehicle-kilometer is the unit of measure, and freight transport for ton-kilometer. The most disaggregate level, network



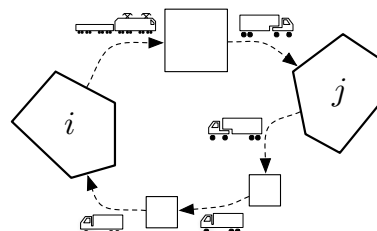
(a) Production and consumption



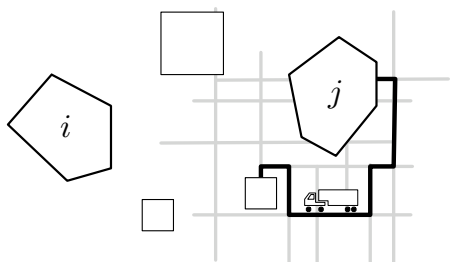
(b) Trade (sales and sourcing)



(c) Logistics



(d) Transportation



(e) Network and routing

Figure 1: Conceptual framework of freight transport modeling (Adapted from Tavasszy, 2008).

and routing, often seeks decision support regarding congestion, tour planning, and city access. Here, chosen techniques include network assignment models and simulation.

Although fairly sophisticated models can be used throughout, many of the approaches only handle flows at an aggregate level (zones) and the detail movement of the freight carriers are not considered (Fernández et al., 2003). Urban tours, where vehicles make multiple stops, are usually completely ignored. Friedrich et al. (2007) report on recent models where activity chain generation is addressed but none of the models, unfortunately, models the behavior at the individual carrier level. Wisetjindawat et al. (2006) consider commodity flow following a top-down approach that explains commodity movement through the interaction among several freight agents in the supply chain. The spatial discrete choice model distinguishes between shippers, receivers and the relationships among them. Their model addresses the commodity movement (generation and distribution), shipment sizing, carrier choice, and the routing and traffic assignment of the carrier vehicles. Hunt and Stefan (2007) present a tour-based microsimulation model of urban commercial movements using data from an extensive survey of 37 000 activity chains in Calgary, Canada.

Agent-based modeling techniques, as opposed to four-step variants, allow one to incorporate and embed the decision-making across the trade, logistics, transportation, and network levels of the framework. Instead of a top-down approach, the individual stakeholders (vehicles, firms, commodities, or industries) have its own autonomous decision-making mandate. All agents are simulated simultaneously, and their interactions with one another and with the environment allows for emergent phenomena that is otherwise lost if a top-down system description approach was followed. Liedtke (2009) notes three advantages that agent-based simulation models present in the freight context: 1) through statistical modeling one is able to represent the heterogeneous nature of freight transport actors and objects; 2) normative optimization engines can be incorporated into the individual agents' behavioral models, allowing observation of agent reactions to policy measures on different time-horizons; and 3) micro model results can be aggregated ex-post as desired to support decisions of various nature.

Liedtke (2009) argues in his agent-based approach that shippers are generally more influential in decision-making than the carriers. Whereas the movement pattern of a person (participating in home, work, education and leisure activities) closely resembles the movement pattern of a (private) vehicle, the situation in freight transportation is more complex. Shippers and recipients make decisions about production, warehousing and ordering. Once orders are placed, carriers (either in-house or outsourced) deal with the tour-planning decisions. Other actors, such as freight forwarders, may also get involved to build transport chains, especially in a multi-national context.

Liedtke's *INTERLOG* model prototype simulates the auction of transport contracts among independent carriers. A simulated population of actors are generated using economic activity, number of employees and location from national statistical descriptors. Shippers and customers are then matched and exchange of goods is set, expressed at microscopic level in metric tons per year. The market interaction module maps how shippers derive shipments and how these shipments are allocated to carriers. Through an experience-based learning process, actors follow behavior-related models to pursue cost minimization, allowing the simulation to settle into a dynamic equilibrium.

Although the autonomous decision-making and learning aspects of Liedtke (2009) is attractive, it requires extensive information that is frequently not available such as detailed sectoral data about companies; inter-sectoral commodity exchanges for different transport markets; and the sensitivity of the transport market to price and distance. Also, the *INTERLOG* model addresses the behavior of only a very small portion—fixed contracts—of total commercial traffic.

Another agent-based approach with experienced-based learning that has until now only been applied to private vehicles, is the Multi-Agent Transport Simulator (*MATSim*) toolkit (*MATSim* development team, 2009). The toolkit has been applied successfully in Gauteng, South Africa, where access to information is limited (Fourie and Joubert, 2009). Since this paper aims to extend the existing agent-based simulation capability of the *MATSim* toolkit to include freight, and extend the existing South African implementation, a brief description of *MATSim* is provided here.

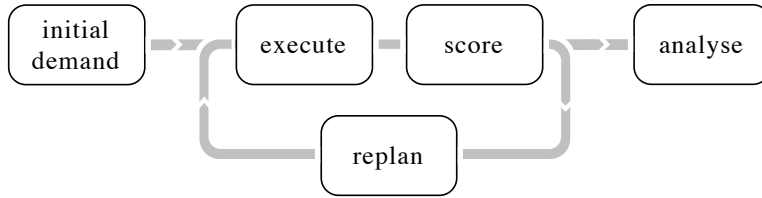


Figure 2: The Multi-Agent Transport Simulator (MATSim) controller (Source: MATSim development team, 2009)

2.1. Multi-agent transport simulation

The *MATSim* toolkit allows for the agent-based simulation of large-scale transport scenarios. Figure 2 illustrates the controlling procedure of the toolkit. The model is initialized with a synthesized population of agents, each having at least one possible plan. A plan consists of a set of sequential activities, each with a location that is associated with a given network. Sequential activities in the plan are connected with a travel leg, described as a route through the network and an intended mode of transport.

When the model is executed, each agent selects a single plan from its set of plans, and all agents execute their chosen plans simultaneously in a mobility simulation. Simulation information is recorded in the form of discrete events such as the start and end of an activity, entering and leaving a link in the network, or waiting to access a link. Simulation events are then interpreted to derive a score for each agent’s selected plan (Charypar and Nagel, 2005). During the replanning step agent plans can be adapted by changing the start time/location of the activities, or the mode/routing of the travel legs. Adapted plans are added to the database of plans, and poor performing plans are discarded. Each agent selects a plan, and the simultaneous execution of plans are repeated (Raney and Nagel, 2006).

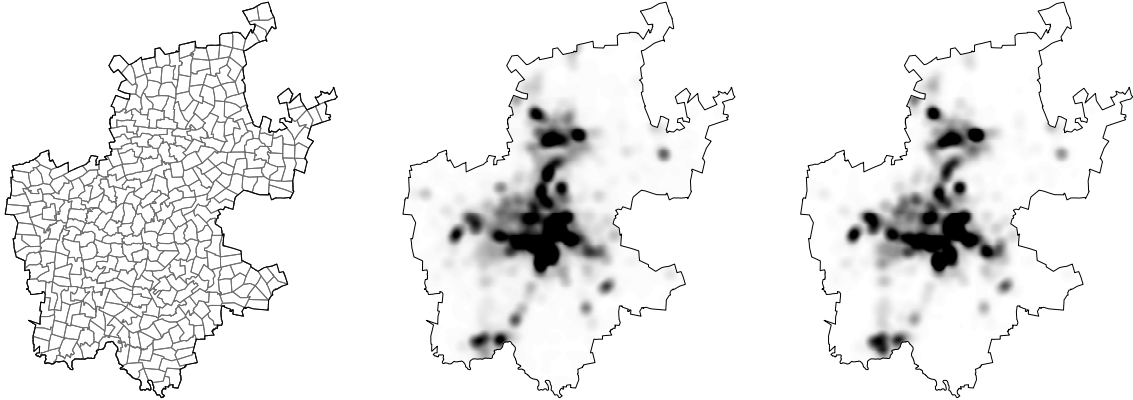
After a specified number of iterations, the simulation terminates and presents statistics such as count comparisons with actual observed counting station data. The higher the number of iterations, the more stable the set of selected plans become.

3. Demand generation

For the purpose of this paper, both the private vehicle plans and the network description was taken from Fourie and Joubert (2009). Private persons’ activity chains are limited to home-work-home to account for the majority of trips during peak periods of the day. In this section we describe the procedure followed to generate an initial demand for commercial vehicles.

From the original 31041 vehicles used by Joubert and Axhausen (2009) to describe the activity chains of commercial vehicles, we were only interested in those that spend 90% or more of their activities within the province of Gauteng. For those 5196 vehicles, activities and activity chains were extracted. Activities were classified as either *major*, lasting in excess of 5 hours, or *minor*, lasting less than 5 hours. Major activities can be thought of as *depots*, or home locations for commercial vehicles. Minor activities include any (and all) activities where vehicles are switched off for periods less than five hours. Once the vehicle is switched on again the activity is considered finished. In the activity extraction, false starts are eliminated, and all consecutive activities at the same location are merged. Examples of minor activities include vehicle refuelling; stopping to perform a service; or loading and offloading, to name but a few. Figures 3b and 3c present the kernel density estimation of the major and minor activities in Gauteng respectively. Conceptually, kernel density estimation fits a smooth surface over the observed activities. For the estimations in Figure 3 a grid size of 500×500 meters and search radius of 5000 meters were used. This was achieved using the *ESRI ArcGIS* implementation of a quadratic kernel function: popularized by Silverman (1986).

With the rich activity set, we were able to build conditional probabilities for four chain parameters: the start location, denoted by M^l ; start time, M^t ; chain duration, M^d ; and the number of activities,



(a) Geospatial analysis platform (GAP) (b) Kernel density estimation of major activities. (c) Kernel density estimation of minor activities.

Figure 3: Demarcation and activity densities for Gauteng, province of South Africa

M^n . To create a plan for a commercial vehicle i , we create the first (major) activity and assign it a location sampled from the kernel density estimation of Figure 3b, M_i^l . Although it is known from Joubert and Axhausen (2009) that, albeit the minority, commercial vehicles may start and end their chains at different locations, we assumed in this paper that chains start and end at the same location. We sampled the chain start time given M_i^l , and we denote the conditional sampling with $M_i^t | M_i^l$; the number of activities, $M_i^n | (M_i^l, M_i^t)$; and the chain duration, $M_i^d | (M_i^l, M_i^t, M_i^n)$.

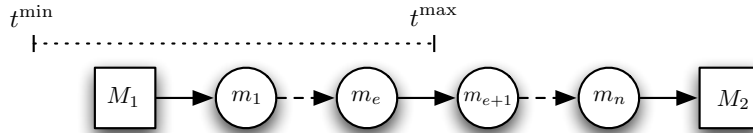
To fill chain i , we sampled M_i^n minor activities from the kernel density estimation shown in Figure 3c, sequenced them randomly, and spaced their start times equally between M_i^t and $M_i^t + M_i^d$. The chain was completed by adding another major activity with the same location as M_i^l , but with a start time of $M_i^t + M_i^d$.

Since, from Joubert and Axhausen (2009), nearly 50% of commercial vehicle activity chains observed has duration in excess of 24 hours, our generated chains had to be adapted to fit within the 24-hour simulation time window imposed on the simulation scenario. Figure 4 depicts the process followed in this paper. The procedure can handle any time window denoted by $[t^{\min}; t^{\max}]$ as is required by other of our projects. For this paper, the time window is specified in seconds for a complete 24-hour day as $[0; 86400]$.

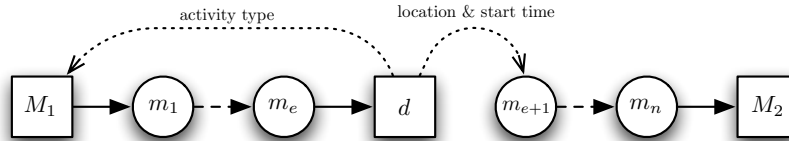
Consider the activity chain in Figure 4a starting with a major activity, M_1 , within the first 24-hour window. The activity chain contains n minor activities, denoted by $m_1 \dots m_n$. The activity chain is completed with major activity M_2 .

Say, for the purpose of this example, the 24-hour window ends somewhere between minor activities m_e and m_{e+1} . We split the activity chain between minor activities m_e and m_{e+1} with the introduction of a dummy *major* activity denoted by d (Figure 4b). The dummy shares its activity type with the first activity, M_1 . The location of d is the same as the location of m_{e+1} . The major activity is added at the end of the first segment of the chain as d with start time 24:00:00 (or 86400 seconds). This has the result that the traveling from activity m_e to activity m_{e+1} is forced to occur at the end of the first segment of the chain. We assume that the impact of whether the traveling occur as the last leg of the first portion, or the first leg of the second portion, is negligible.

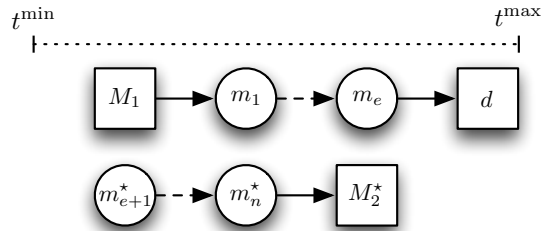
Once the dummy major activity is added, all the activity start times following the dummy activity is adjusted by simply subtracting 24 hours, resulting in the two independent chains illustrated in Figure 4c, with the starred version of the activities denoting the activities with adjusted start and end times. Each of the resulting chains are treated as an independent agent without weighing the chains.



(a) The original activity chain exceeding the specified time window.



(b) Introduction of a dummy major activity at the end of the time window. The dummy shares the activity type of the first portion of the chain; and ending time and location with the first activity of the subsequent portion.



(c) The resulting two activity chains depicting two artificial, and independent, commercial vehicles.

Figure 4: Procedure to artificially *wrap* an activity chain around a given time window, resulting in independent chains, each starting and finishing within the time window.

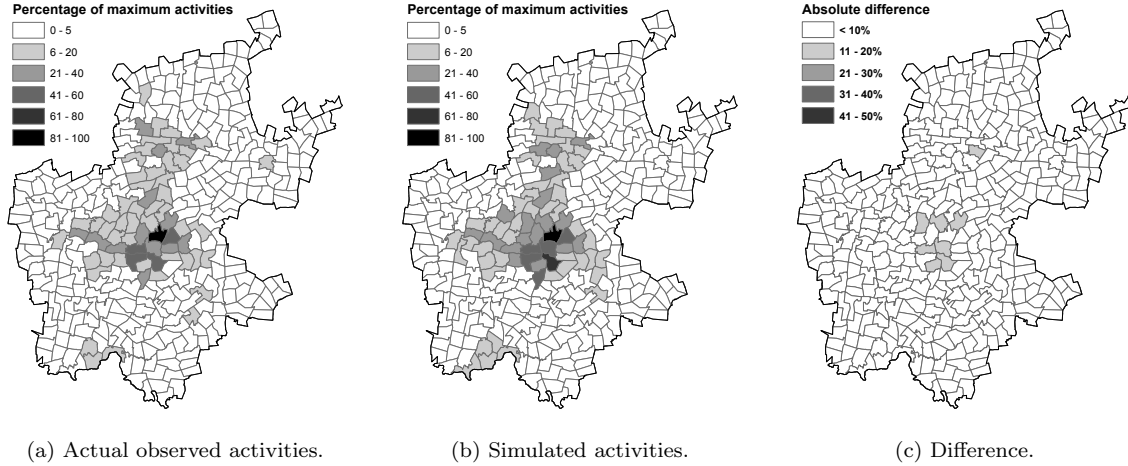


Figure 5: A comparison of the actual vs. simulated activities for the 9th hour. Subfigure (a) shows the actual observed activity densities, normalized to the highest actual activity count over the whole 24-hour period; (b) shows the simulated activity densities, normalized to the highest simulated activity count; and (c) shows the absolute difference between the normalized actual and the normalized simulated activity counts.

Should the remaining portion of the chain still exceed the time window, t^{\max} , the procedure is repeated on the remainder.

4. Simulation results

To account for the variation as a result of the random sampling of activities from the kernel density estimation, we generated 6 populations of freight agents, each with a different random seed. Each population was added to the same population of private car agents to derive 6 unique scenarios. The different scenarios were each simulated for 100 iterations, and in the remainder of the section we report on the simulated results as the average over the six scenarios—they were not significantly different. To measure the repeatability of the six scenarios we compared the trip durations for all agents across thirteen 5-minute intervals for each scenario, calculating the standard deviation during each interval. The average standard deviation across all intervals was then calculated as $\sigma = 1.52\%$.

The private and commercial vehicle plans were combined, and simulated for 100 iterations. To evaluate the activity location and timing of the commercial activities, we analysed the discrete events produced during the 100th iteration. We were interested in activity start times and the location of all minor activities.

To determine if the simulated commercial activities occur at the right geographic location at the right time of day, activities were geographically aggregated to the GAP mesozone level (Figure 3a) so as to be able to compare it with the observed activities reported by Joubert and Axhausen (2009). Temporally, the events were aggregated to the hour of the day in which it occurred. Figure 5 shows, as an example, a geographic comparison of the actual versus simulated activity counts for all activities in the 9th hour, i.e. between 08h00m00 and 08h59m59. The actual observed activities span approximately 180 days, so both the actual and the simulated activity densities were normalized to the highest activity count found in the respective sets. As an example, consider the process of normalizing the simulated data set. The reference is the highest activity count found in any zone at any time of the day. This was found to be 1167 for zone 4917 in the 10th hour. If then, 245 activities were simulated to start in zone 4570 in the 9th hour, the normalized density for zone 4570 in the 9th hour would be $(245/1167) \approx 21\%$.

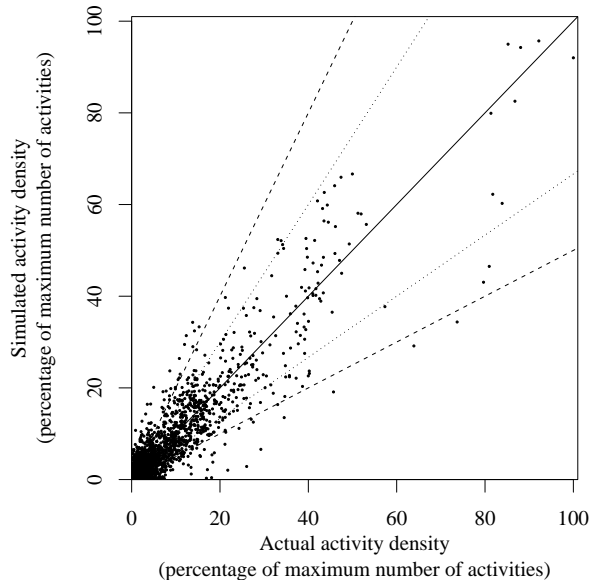


Figure 6: The comparison of simulated vs. actual activity densities for all hours across all GAP zones. From top to bottom, the linear bands represent a 2:1, 1.5:1, 1:1, 1:1.5, and 1:2 ratio of simulated to observed normalized activity densities.

We believe this is a valid approach since GAP zone 4917 had the highest activity counts in both the observed and all the simulated scenarios. Hence, densities were normalized to the same zone.

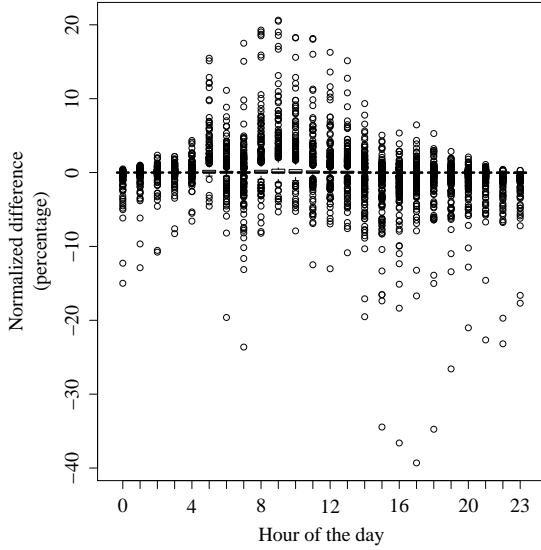
If geographic location is less important, the simulated results may also be presented as we’ve done in Figure 6. Each point represents a unique zone-hour combination. The error seems to be greatest in low-activity zones.

Next we considered the accuracy of the simulated activities for different hours of the day. Figure 7 shows box-and-whisker plots for the normalized difference between the simulated number of activity starts during each hour and the observed number of activity starts. Positive values indicate instances where simulated activity counts were *higher* than the actual observed activity counts, and negative values where they were *lower*. All mesozones are accounted for in Figures 7a and 7c and we hence refer to the comparisons as *unbiased*. Many of the mesozones, however, had no activity starts observed and also non simulated. In itself this is arguably good, but the large number of *zero-observed-zero-simulated* mesozones obscures the results somewhat. Removing all the zero-observed-zero-simulated mesozones yields a result that is *biased* towards simulation differences, which we present in Figures 7b and 7d.

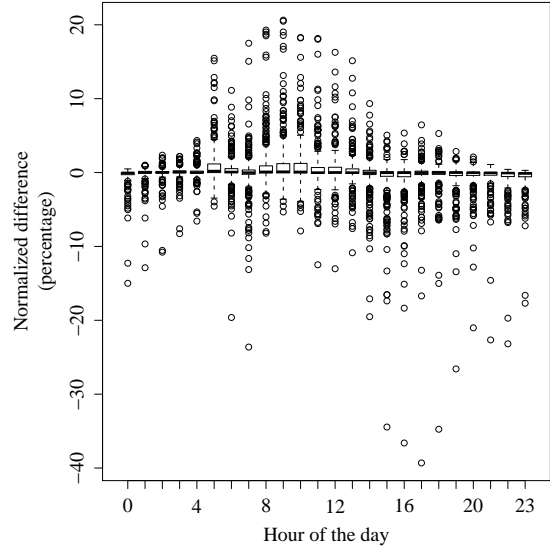
For both the unbiased and biased cases the box widths, the interquartile ranges (IQRs) are very tight, so in Figures 7c and 7d we’ve zoomed in on the range $[-2, 2]$.

In both plots the whiskers ends indicate the lowest datum still within 1.5 times the IQR of the lower quartile, and the highest datum within 1.5 times the IQR of the upper quartile. With all the medians located at (or very near) the zero datum and the very narrow box widths, we argue that the commercial activities are accurately simulated, at least at a time-space granularity of hours and mesozones. Our model over-predicts commercial activities during the period 08:00-12:00, while under-predicting from 15:00 until 00:00. Possible assignable causes for over-estimating include that only home-work-home trips are modeled for private vehicle users, essentially freeing up a large portion of the road capacity during the off-peak hours. Also, the effective road network considered by commercial vehicles is probably more limited than the complete network currently made available to all road users, allowing for more accessibility than what is actually realistic.

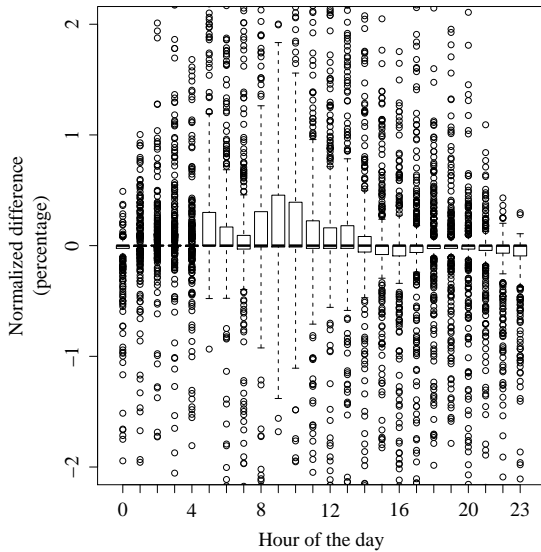
But we were also interested to understand what impact the inclusion of commercial vehicles had on the private cars that were modeled in the simulation. To do that, we followed two approaches.



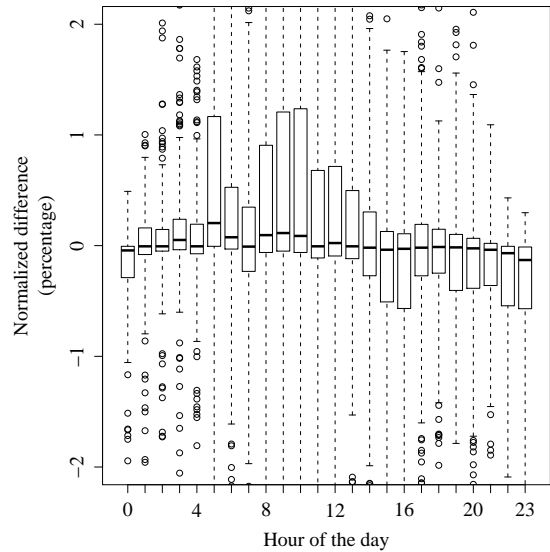
(a) Unbiased comparison including all mesozones.



(b) biased comparison with zero-observed-zero-simulated mesozones removed.



(c) Unbiased comparison, zoomed in.



(d) biased comparison, zoomed in.

Figure 7: Box-and-whisker plot showing the normalized difference between simulated and actual number of activity starts for different hours of the day. A positive normalized difference indicates simulated values higher than actual observed values. Whiskers ends indicate the lowest datum still within 1.5 times the interquartile range (IQR) of the lower quartile, and the highest datum within 1.5 times the IQR of the upper quartile. Circles outside the whisker extents are considered outliers.

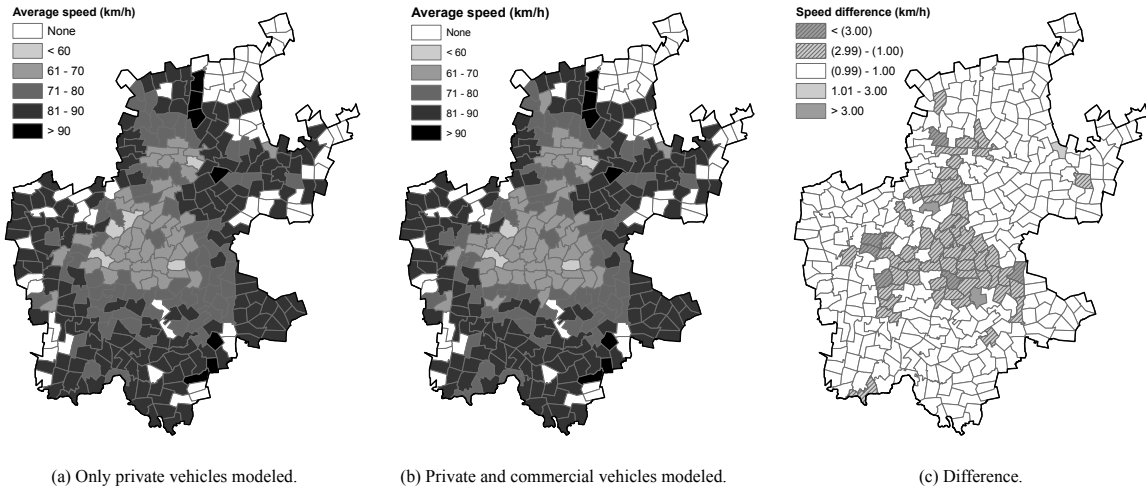


Figure 8: A comparison of the average speed, and speed difference for private vehicles aggregated to mesozone level. Speeds are for the 9th hour.

First, we associated each road link with a mesozone, and calculated the average speed of a mesozone as the average speed of private vehicles across all links associated with that mesozone. If no activity occurred during a specific hour on the link, the link’s free speed was used. Only when no network link was associated with the mesozone was no average speed calculated. In the second approach we measured the total trip time of each private vehicle agent, and calculated the average over all agents. We performed a comparison of these measures when only private vehicles are modeled, against our model where commercial vehicles were included.

The overall impact across the study area is shown in Figure 8. With slower traffic expected in the economically dense areas, the two economic centra (Tshwane in the North, and the larger Johannesburg-Ekurhuleni in the South) are distinguishable with its lighter shades, while the introduction of commercial vehicles shows only a slight decrease in average speeds. The slight change may be attributable to the fact that commercial vehicles, in absolute numbers, only represent about 5% of the total traffic volume.

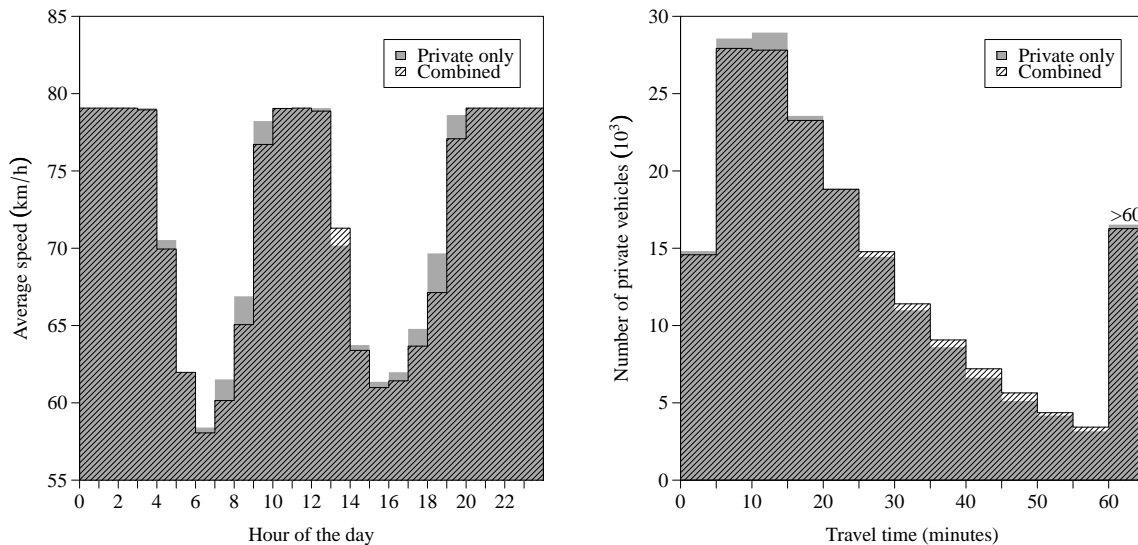
An hour-by-hour comparison for a single mesozone (zone 4917) is shown in Figure 9a. The impact that commercial vehicles have on the average speed of private vehicles is again not very significant. Except between 13:00 and 14:00, slower average speeds are reported for all hours.

A different view of the same analysis is reported in Figure 9b where the overall trip times did not change dramatically either. There are fewer short trips and more trips exceeding 20 minutes. Since only internal distribution trips were simulated, which accounted for only an estimated 35% of Gauteng commercial traffic (Joubert and Axhausen, 2009), we argue that the impact will become higher once through-traffic is incorporated too. Still, the inclusion of commercial vehicles in the agent-based simulation model do have the desired (expected) effect: slowing of traffic, and thus increased travel time.

5. Conclusion

The activity chains generated and simulated in this paper were shown to be a geographically and temporally accurate representation of observed vehicle chains. A number of future extensions to this paper is proposed.

Our likely next step is to generate through traffic and trips for which less than 90% of the activities occur in the study area. Once these trips are accounted for, it will be appropriate to compare the simulated commercial traffic against actual traffic counts. Our contribution shows that we were able



(a) Average speed in a single zone (Zone 4917).

(b) Overall trip travel time.

Figure 9: Comparisons of the speed and travel time of private vehicles when only private vehicles are modeled versus when commercial traffic is combined with private vehicles.

to model commercial vehicle behavior in terms of replanning their route and activity schedules. The impact on policy evaluation should be clear: a base case can be established by modeling the status quo and validating it using actual vehicle counts. If we want to evaluate how commercial vehicle movement change as a result of say a road pricing or lane reservation policy, we can include those policy measures on the network, resimulate the scenario, and compare the vehicle movement (activity densities, activity timing, average link speed, etc) against that of the base case. The simulated commercial vehicles will autonomously replan their routes and activity schedules so as to maximize their expected utilization. In doing so, the simulated agents are giving us a reasonable indication of what we can expect from real autonomous decision-making commercial vehicles in maximizing their utility.

In accordance with Tavasszy (2008), this paper suffers from the empirical challenge of disentangling light goods vehicles from heavier ones. We acknowledge that vehicle size plays a role in congestion impact. Activity chains generated in this paper do not distinguish between vehicle sizes or activity types (commercial freight vs commercial services).

The freight trip chains in this paper were based on and sampled from observed trip chains. New chain structures can hence not *emerge* from the current model. Our approach in this paper to randomly draw attribute values characterizing the commercial chains can be extended. One extension is to replace the random draw with a decision engine capable of simulating and predicting the attribute values based on underlying factors, such as commodity type and the land use of activities. This would require that the underlying factors are adequately described by available data sources.

In the absence of any additional information to the GPS logs, one may be tempted to consider supply chain network structures intractable. Yet, with the inferred knowledge of where major and minor activity locations are, one may be able to analyse the activity chains based on the underlying land use data, and infer the network structures. Contributions such as those made by de Jong and Ben-Akiva (2007), Liedtke (2009) and Wisetjindawat et al. (2006) may prove very useful inputs to infer network and supply chain structures.

It is known that the GPS logs used to infer activities include both service and freight vehicles. We can hence argue that since we sample from all commercial trip chains observed, the model results do account for empty trips, something very often omitted in other modeling approaches. Also,

activities may include collections from production locations, deliveries to receivers, and also activities at hub locations. The latter has been noted by Friedrich et al. (2007) as lacking from network models for freight. Empty trips and hub activities, although accounted for, will unfortunately not be distinguishable within our results.

6. Acknowledgements

The authors are grateful to the South African National Research Foundation (NRF) who was partly funding the research under Grant FA-2007051100019.

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

Crainic, T. G. and Laporte, G. (1997). Planning models for freight transportation. *European Journal of Operational Research*, 97(3):409–438.

CSIR Built Environment (2009). Geospatial Analysis Platform and NSDP Spatial Profiles. Accessed 13 March from <http://www.gapweb.co.za>.

de Jong, G. and Ben-Akiva, M. (2007). A micro-simulation model of shipment size and transport chain choice. *Transportation Research Part B: Methodological*, 41(9):950–965.

De Jong, G., Gunn, H., and Walker, W. (2004). National and international freight transport models: An overview and ideas for future development. *Transport Reviews*, 24(1):103–124.

Fernández, J. E., de Cea, J., and Soto, A. (2003). A multi-modal supply-demand equilibrium model for predicting intercity freight flows. *Transportation Research Part B: Methodological*, 37(7):615–640.

Fourie, P. J. and Joubert, J. W. (2009). Large-scale agent-based transport simulation of South African commuters. Working paper IE-OG-2009-02, University of Pretoria, South Africa.

Friedrich, M., Haupt, T., and Nökel, K. (2007). Freight modelling: Data issues, survey methods, demand and network models. In Axhausen, K. W., editor, *Moving through nets: The physical and social dimensions of travel*, chapter 9, pages 249–277. Elsevier, Amsterdam, The Netherlands.

Hesse, M. and Rodrigue, J. (2004). The transport geography of logistics and freight distribution. *Journal of Transport Geography*, 12(3):171–184.

Hunt, J. and Stefan, K. (2007). Tour-based microsimulation of urban commercial movements. *Transportation Research Part B: Methodological*, 41(9):981–1013.

Joubert, J. W. and Axhausen, K. W. (2009). Inferring commercial vehicle activities in Southern Africa. Technical Report 552, Verkehrs- und Raumplanung, IVT, ETH Zurich.

Kveiborg, O. and Fosgerau, M. (2007). Decomposing the decoupling of danish road freight traffic growth and economic growth. *Transport Policy*, 14(1):39–48.

Liedtke, G. (2009). Principles of micro-behavior commodity transport modeling. Forthcoming in *Transportation Research Part E*, doi: 10.1016/j.tre.2008.07.002.

MATSim development team (2009). The MATSim project. <http://www.matsim.org>. Accessed on June 24th, 2009.

Ortúzar, J. d. D. and Willumsen, L. G. (2001). *Modelling Transport*. Johan Wiley & Sons, Ltd, Chichester, England, 3rd edition.

- Raney, B. and Nagel, K. (2006). An improved framework for large-scale multi-agent simulations of travel behaviour. In Rietveld, P., Jourquin, B., and Westin, K., editors, *Towards better performing European Transportation Systems*. Routledge, London.
- Rodrigue, J. (2006). Transport geography should follow the freight. *Journal of Transport Geography*, 14(5):386–388.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- Spear, B., Giaimo, G., Curlee, T. R., and Neels, K. (2008). Freight modeling: State of the practice. In Hancock, K. L., editor, *Freight Demand Modeling: Tools for Public-Sector Decision Making*, volume 40 of *TRB Conference Proceedings*, pages 21–26.
- Tapio, P. (2005). Towards a theory of decoupling: degrees of decoupling in the EU and the case of road traffic in finland between 1970 and 2001. *Transport Policy*, 12(2):137–151.
- Tavasszy, L. (2008). Freight modeling: An overview of international experiences. In Hancock, K. L., editor, *Freight Demand Modeling: Tools for Public-Sector Decision Making*, volume 40 of *TRB Conference Proceedings*, pages 47–55.
- Wisetjindawat, W., Sano, K., and Matsumoto, S. (2006). Commodity distribution model incorporating spatial interactions for urban freight movement. *Transportation Research Record*, 1966:41–50.