Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 125 (2014) 136 – 146

Procedia
Social and Behavioral Sciences

8th International Conference on City Logistics

Generating Intra and Inter-provincial Commercial Vehicle Activity Chains

Quintin van Heerden[a,b]*, Johan W. Joubert[b]

[a]Transport Systems and Operations, Built Environment, Council for Scientific and Industrial Research,Meiring Naudé Road, PO Box 395, Pretoria, 0001, South Africa
[b]Center of Transport Development, Industrial and Systems Engineering, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa

Abstract

Modelling large-scale traffic flow systems at a disaggregate level can be data intensive as it requires extensive knowledge about the activities and activity chains of vehicles. This paper focuses on activity chain generation for commercial vehicles. We use a large sample of GPS records to extract a complex network and sample chain characteristics from. The paper makes a valuable contribution in both its methodology, and in its focus on intra and inter-provincial vehicle populations simultaneously. The simulated chains are validated in terms of vehicle kilometrekilometres travelled and its spatiotemporal accuracy, comparing favorably in both.

© 2014 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of the Organising Committee of the 8th International Conference on City Logistics.

Keywords: Commercial vehicle modelling; activity chain; freight; urban

1. Introduction

Commercial vehicles may account for only a small portion of the total vehicle population, but they contribute disproportionately to the traffic conditions, the impact on road quality, and the environment through noise and emissions. On the positive side, they contribute to the economic activity and resulting gross value-add of a region.

* Corresponding author. Tel.: +27 12 841 3377; Fax: +27 12 841 4044.
E-mail address: qvheerden@csir.co.za

1877-0428 © 2014 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of the Organising Committee of the 8th International Conference on City Logistics.
doi:10.1016/j.sbspro.2014.01.1462

Winter & von Hirschhausen (2006) use a value of time for commercial vehicles that is three times that of private car commuters. Freight is expensive, though, since transportation costs make up a sizable portion of a country's total logistic cost.

Our interest in this paper is on vehicular movement, and not the commodities transported, as is often the focus in freight flow models. Our understanding of how commercial vehicles behave and move around lags our understanding of how people move around. Many of the freight models are adapted from classical four stage models, for example Marker Jr and Goulias (1998) and Al-Deek, Johnson, Mohamed & El-Maghraby (2000), to name but a few. In practice, many modellers add commercial and freight vehicles as mere background noise. Comparing both aggregate and disaggregate freight models, Hensher & Figliozzi (2007) and Samimi, Mohammadian & Kawamura (2009) highlight the gap in our understanding of freight vehicles' behaviour in order to capitalize on the decision-making benefits that disaggregate freight modelling holds. Chow, Yang & Regan (2010) review the value of disaggregate freight models, noting that good freight demand models should have a strong behavioural foundation. To this extent contributions such as Holguin-Veras & Patil (2005), Ruan, Lin & Kawamura (2012) and Schroeder, Zilske, Liedtke & Nagel (2012) have been very valuable.

Disaggregate modelling approaches can provide more accurate and richer result sets that are time-dependent and allows for aggregation to any level required for better decision-making. But they are more often than not computationally expensive and data intense. Modelling commercial vehicles more accurately and realistically should result in more accurate predictions of travel time as both Gao, Balmer & Miller (2009) and Fourie (2010) show using agent-based models. When testing infrastructure investment decisions, say the expansion of a portion of the road network, improved travel time prediction allows better evaluation of the direct effects (travel time savings) of the investment. Although agent-based models hold promise of improved disaggregate modelling and decision-support, much work remains.

This paper aims to contribute by taking another step in disaggregate commercial vehicle activity chain modelling. We build on an earlier paper by Joubert, Fourie & Axhausen (2010) who show how large-scale scenarios that include both private car and commercial vehicles can be simulated in an agent-based setting using complete activity chains. These chains were the result of passive GPS logging of large samples, and not the consequence of distribution channels (Ruan, Lin & Kawamura, 2012) or complex contract negotiations (Liedtke, 2009; Schroeder, Zilske, Liedtke & Nagel, 2012), the latter two being much more sophisticated, but often unattainable as a first approach for low and middle-income countries, such as South Africa.

The earlier work by Joubert, Fourie & Axhausen (2010) had two main drawbacks, both which we aim to address in this paper. Firstly, the synthetic population of commercial vehicles only accounted for those agents that remain predominantly inside the study area. Ruan, Lin & Kawamura (2012) call these urban commercial vehicle movements while in the context of this paper we will refer to such chains as intra-provincial. The reason is context-specific as we use Gauteng as study area, the smallest of the nine provinces in South Africa and the economic heart of the country (see Fig. 1). It would not make sense to focus on any one of the five metropolitan and district councils within Gauteng as they have morphed into a megacity with the majority of people and goods movements between the metros. In this paper we extend the earlier work and create separate and unique populations for the intra- and inter-provincial commercial vehicles.

Secondly, Joubert, Fourie & Axhausen (2010) sample the commercial vehicle activity locations from a kernel density estimate and randomly sequence them. Although their travel time and spatiotemporal validation affirm their approach, results reveal that the chains significantly overestimated the vehicle kilometrekilometres travelled (VKT). We will employ complex networks to improve on the activity sequencing. In addressing these issues, our paper makes a methodological and substantive contribution. Methodologically, to the best of our knowledge, it is the first integration of GPS records and complex networks to generate activity chains. We demonstrate the use of a large GPS data set (40,000+ vehicles over a six-month period) to generate full-day activity chains. We start by presenting the data and methodology used to extract the activities and activity chains from GPS data, as well as the resulting complex networks. The generation of the activity chains for both intra- and inter-provincial vehicles are explained, after which the simulated chains are validated. We conclude with suggestions for further development.

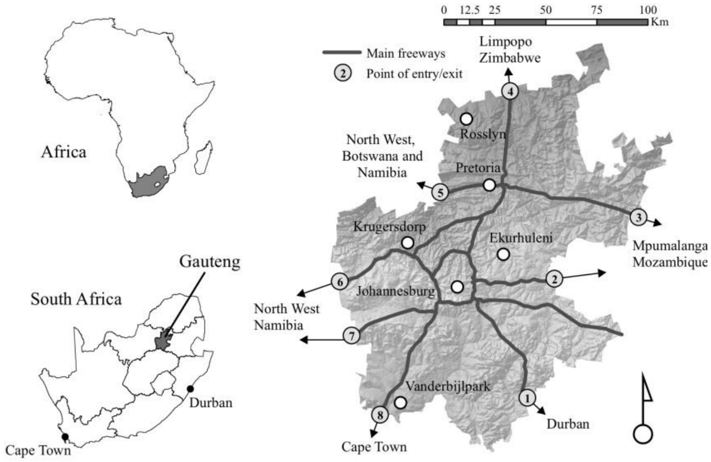


Fig. 1. The landlocked Gauteng, South Africa. The smallest of the nine provinces with the highest economic activity and peculiar omnidirectional connections with surrounding provinces. (Source: Joubert, 2012)

2. Data and Methods

This paper, like Joubert, Fourie & Axhausen (2010), Joubert & Axhausen (2011), Joubert (2012), van Heerden & Joubert (2012), uses the Digicore Fleet Management data set that includes the GPS logs for 41 711 commercial vehicles over the six-month period 1 January 2009 -- 30 June 2009. The cTrack data provide a large sample of various vehicle-related metrics such as engine temperature, opening and closing of doors in an ignition-on state, harsh braking, and over revving. Our point of departure is the activity chains extracted from the GPS records (Joubert & Axhausen, 2011). For each vehicle we have all its complete observed activity chains: each chain starts and ends with what is referred to as a major activity (lasting in excess of 5 hours), and for the rest made up of minor activities. Although we do not know the activity purpose, e.g. loading, delivery or service, we do know the exact location and start and end times of the activity.

Based on the total number of activities that were observed for a vehicle in the given study area, Gauteng, we distinguished between intra-provincial vehicles; those that have at least 60% of their activities inside Gauteng; and inter-provincial vehicles, those with at least one but fewer than 60% of their activities inside Gauteng (Joubert & Axhausen, 2011). A total of 10,428 intra and 11,262 inter-provincial vehicles were identified. The remaining 48% of the vehicles had no activities inside Gauteng and were omitted for this paper.

Since GPS records are noisy, the density-based algorithm proposed by Joubert & Axhausen (2013) is used to estimate the real location of the facilities where activities occur. If at least 15 activities were observed in a 30m radius, a cluster was created. The activity locations of all activities associated with a cluster were then changed to the location of the centroid of the cluster. Since both intra and inter-provincial vehicles cross the boundary of Gauteng, we created eight artificial gateway facilities on the most prominent routes leading into and out of the province (Fig. 1). The portions of the activity chains that fell outside the boundary were subsequently ignored and replaced with an appropriate entry or exit activity.

Using a direct vehicle trip between two clustered facilities as a proxy for a dyad (link) between the two facilities, we extracted a weighted (complex) network of trips. For each trip of each chain of each vehicle we created a link in the network between the origin and destination nodes (facilities), but only if both the nodes were included during the clustering. If not, the trip was ignored. All nodes that coincided with major activities were flagged. The result is a complex network of connectivity between facilities at which many vehicle activities were observed. For a comprehensive introduction to spatial networks the reader is referred to Boccaletti, Latora, Moreno, Chavez & Hwang (2006) and Barthélemy (2011).

We created two networks: one from intra and the other from inter-provincial vehicles' observed activity chains, anticipating that their activity chain structures are different.

3. Activity Chain Generation

Day-to-day variation of the level of activity was studied by van Heerden & Joubert (2012). In line with expectation, Mondays through Fridays were all very similar while weekends differed distinctly. As a result, a Tuesday was chosen as a typical day for our activity chain generation, and in particular 27 January 2009 was randomly chosen as an anchor date. The rest of the section is split into covering the two types of activity chains: intra- and inter-provincial.

3.1. Intra-provincial chains

We considered the 10,428 intra-provincial vehicles and identified 6,052 with activity chains that started on the randomly selected anchor date. A conditional probability matrix was constructed using the observations of each chain's start time, number of activities, and chain duration.

In a spatial (complex) network, the *degree* of a node refers to the number of connections that node has with other nodes in the network. Since we used trips as a proxy for the connections, and multiple trips by the same or different vehicles may occur between two facilities, each connection can be weighted, resulting in a *weighted degree*. We used directed edges since it is fair to say that many direct vehicle trips from *a* to *b* does not imply the same number of direct trips from *b* to *a*.

We started by filtering the 38,264 nodes in the intra-provincial network to those 12,488 that were flagged as being possible locations for major activities. From these we randomly sampled one using its location for the first major activity, weighting each node's probability by the node's weighted degree. From the conditional probability matrix we sampled a start time for the major activity; the number of activities per chain *given* the sampled start time; and the chain duration *given* the start time and the number of activities.

For the minor activity locations making up the activity chain, we sampled from the complex network. Starting with the first major activity, we identified all the nodes connected to it via outgoing connections, weighing the probability to be chosen of each with the connection's weight. Once sampled, we proceed in the same way for the sampled node: randomly sampling from the weighted outgoing connections. We proceeded iteratively until we've reached the sampled number of minor activities. The timing of the activities was evenly spaced over the chain's duration.

The final major location was sampled from the complex network by only considering major candidates from the last activity's location. If none existed, one was sampled in the same way as the first major activity.

If a node had no outgoing edges to sample the next location from, all facilities within a 5-kilometre radius were considered and one was randomly chosen from these as the next activity's location.

3.2. Inter-provincial chains

Only 1,528 of the 11,262 inter-provincial vehicles were identified as having an activity chain starting on the sampled anchor date. One plausible argument is that long haul inter-provincial activity chains typically exceed a single day, and may have already been in progress on the anchor date.

We had to distinguish between two types of inter-provincial activity chains. We first address *in-out* chains: those chains that start outside the province, enter through one of the gateways, perform one or more activities inside the province before leaving again through one of the gateways. Of the 1,528 inter-provincial vehicles, 541 were of the *in-out* type, and we used their observed chains to again populate a conditional probability matrix noting the start time, number of activities per chain and the chain duration. Recall that the portion(s) of a vehicle's chain that falls outside the province has already been replaced with a single entry (or exit) activity, so that the start time was in this case already the time of entry. Additionally we sample a gate pair at which the entry and exit occur. These form the first and last major activities of the chain. The inter-provincial complex network we used to sample the minor activity locations from consisted of 23,908 nodes, and we followed the same sampling procedure as for intra-provincial chains. Minor activities were again evenly spaced over the chain's duration.

The second type of inter-provincial chain is *out-in* chains: those chains that start inside the province; leave through one of the gateways; perform a number of activities outside the province before returning again through one of the gateways. Here we dealt with the first (before leaving) and second (after returning) portion of the chain separately. For the first portion we identified 542 vehicles and created a conditional probability matrix from the start time, number of activities, and the chain duration. The first major activity was sampled from the 5,099 nodes that were flagged as possible major activity locations. The required number of minor activities was then sampled, as before, from the complex network. A gate pair was sampled, and we used the first gate as the final major activity, i.e. exit activity that finished the chain.

For the returning part of the out-in chain we start with the second gate of the sampled pair as the first major activity. Here we created a conditional probability matrix of the start time, number of activities, and duration of 445 identified vehicles observed on the anchor date. As before, the minor activities following the entry were sampled from the complex network, and the final major activity was sampled from the 5,099 nodes flagged as possible major activity locations.

3.3. Chains spanning multiple calendar days

It occasionally happened that a chain spans over multiple calendar days. Since our ultimate goal is to use the synthetic population in an agent-based transport model, which is commonly only set up for a single 24-hour period, we had to deal with the multi-day chains. Following the procedure of Joubert, Fourie & Axhausen (2010) we split those into multiple artificial chains, with each segment fitting within a single 24-hour period.

4. Validation

In this section we validate the synthetic population's activity chains using three metrics. Firstly we check whether the vehicle kilometrekilometres travelled (VKT) within the province compares favorably with the observed VKT. Secondly we ensure that the activity chains of inter-provincial traffic enter and exit at the appropriate gates. Lastly, we check the spatiotemporal accuracy of the activities.

4.1. Vehicle kilometrekilometres travelled

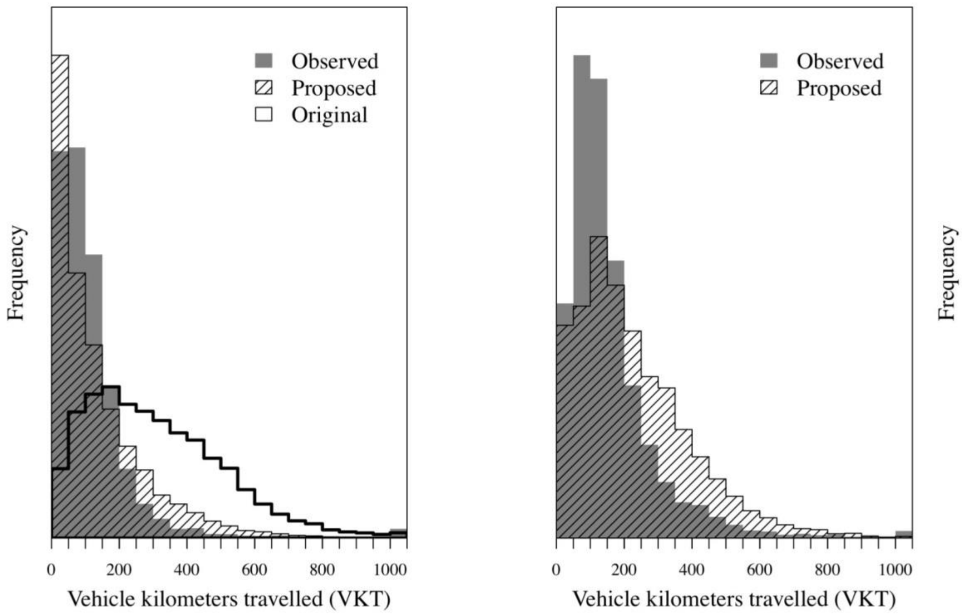
The agent-based plans generated by Joubert, Fourie & Axhausen (2010) had two major shortcomings. Firstly, it only focused on intra-provincial vehicles, inflating the population to also account for the inter-provincial traffic. From van Heerden & Joubert (2012) we know that inter-provincial vehicles' activity chains are quite different and should be addressed independently. Secondly, when simulated in the Multi-Agent Transport Simulation (MATSim) toolkit, the activity chains resulted in a significant over-estimation of the vehicle kilometrekilometres travelled (VKT), even though the population size itself was much smaller than what it ought to have been.

To see if our complex network approach is an improvement, we approximate the VKT for each vehicle in this section. GPS signals from the fleet management data set are often infrequent, as much as 5 minutes apart, especially if no irregular signals like harsh breaking is reported. This means that we cannot do accurate map matching to obtain the true VKT travelled. To estimate the VKT, we use an $A*$ landmark routing algorithm to find the shortest route between consecutive activities in the activity chain (Hart, Nilsson & Raphael, 1968). The histogram shown in Fig. 2 shows the different VKT frequencies for the observed chains, our proposed synthetic population's chains, and the original population's chains as used in Joubert, Fourie & Axhausen (2010). We distinguish again between intra- (2a) and inter-provincial (2b) chains. The overestimation of the original population is quite visible in Fig. 2a. Fitting a two-parameter Weibull distribution to the observed VKT distribution yields a scale parameter λ = 120.7 and shape parameter k = 0.994. The proposed population's VKT yields a Weibull distribution with similar parameters, λ = 131.9 and k = 0.917.

Although the proposed intra-provincial population is more accurate than the original, which had a fitted Weibull distribution with λ = 344.8 and k = 1.578, we still have to reject the null-hypothesis (H_0) that claims the proposed population has the same VKT distribution than the observed VKT distribution. A χ^2 statistic is calculated as 1633, much higher than the critical value $\chi^2_{(1-\alpha)(b-c)} = 32.67$ when using a confidence level $\alpha = 0.05$, $b = 21$ bins and $c = 2$ denoting the parameters for the Weibull distribution. The proposed population compares favorably with the original population in that the latter had a χ^2 value of 35.622.

There is a relation between the VKT and the number of activities a chain has. Chains with many activities are assumed to be longer (in distance). In Fig. 3 we plot the VKT-to-number-of-activity relationships for both the observed and our synthetic activity chains, distinguishing again between intra- (3a) and inter-provincial (3a) traffic. Consistent with expectation, long haul inter-provincial activity chains tend to have fewer activities per chain. Since we calculate the VKT using only the portion of the chain *inside* the province, the scale of the y-axis for intra- and inter-provincial vehicles are the same.

Table 1 provides some summary statistics on the number of activities per chain observed and simulated. The proposed population of commercial vehicle chains again compares favorably.

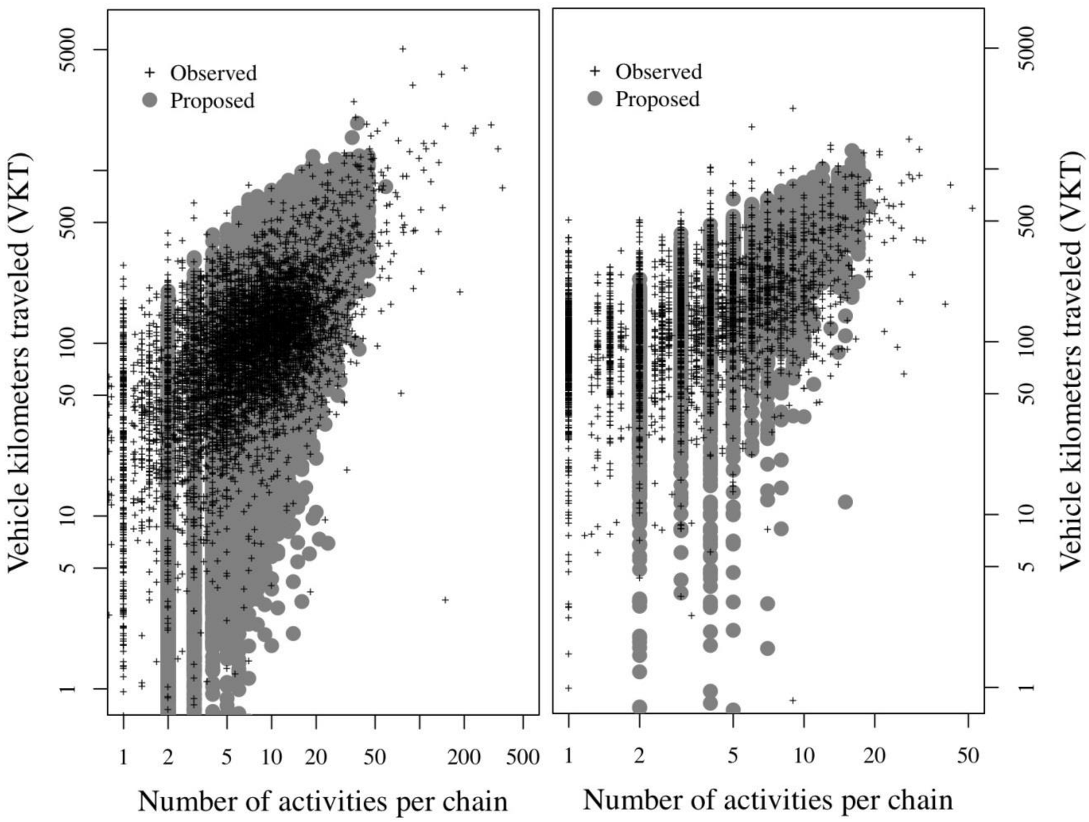


(a) Intra-provincial (b) Inter-provincial

Fig. 2. Comparison of vehicle kilometrekilometres travelled (VKT)

Table 1. Summary statistics for the number of activities per chain

		Percentiles				
		25%	50%	75%	95%	99%
Intra-provincial	**Observed**	3	6	11	23	49
	Simulated	5	9	14	22	31
Inter-provincial	**Observed**	1	3	5	11	18
	Simulated	3	5	7	11	16



(a) Intra-provincial (b) Inter-provincial

Fig. 3. Comparison of vehicle kilometrekilometres travelled (VKT) as a function of the number of activities per chain

4.2. Gateway pairs

To validate whether vehicles from the synthetic population enter and exit through the correct gates, the gate pairs from the synthetic population were compared to the gate pairs from the observed data.

Table 2a shows the observed proportions of in-out chains' gate pairs. The table represents all normal Tuesday chains in the 6-month period. To ease interpretation we shaded the proportions, the darker a cell appears in the row (or column), the higher the proportion of that specific gate pair. Table 2b shows the proportions of in out-chains from the synthetic population. Similarly, Tables 3a and 3b show the observed proportions from the observed and synthetic population for out-in chains. All proportions smaller than 0.05 were omitted from the tables. The row and column totals are expressed as percentages of the total number of entries (and exits).

The proportions obtained from the generated synthetic population correlate well with the observed data. The prominent dark diagonal is again more evident in the simulated out-in chains' table, which also corresponds with the observations.

Table 2. Fractions of gate pairs for in-out chains

(a) Observed in-out pairs

From	To 1	2	3	4	5	6	7	8	Total entries
1	0.565	0.137	–	0.053	0.052	–	0.110	–	0.232
2	0.079	0.658	–	–	–	–	0.118	–	0.197
3	–	0.378	0.396	–	0.074	–	–	–	0.083
4	0.185	0.105	–	0.360	0.089	–	0.176	–	0.062
5	0.073	0.062	0.063	0.056	0.664	–	–	–	0.094
6	0.065	0.106	–	–	–	0.516	0.169	0.080	0.058
7	0.083	0.122	–	–	–	–	0.631	–	0.199
8	0.092	0.089	–	–	0.126	0.110	0.137	0.405	0.074
Total exits	0.195	0.242	0.061	0.062	0.106	0.063	0.211	0.059	

(b) Synthetic population in-out pairs

From	To 1	2	3	4	5	6	7	8	Total entries
1	0.563	0.133	–	0.060	0.057	–	0.111	–	0.236
2	0.080	0.637	–	–	–	–	0.117	–	0.207
3	–	0.363	0.391	–	0.089	–	0.052	–	0.084
4	0.227	0.099	0.055	0.343	0.072	–	0.171	–	0.061
5	0.086	0.076	0.076	0.054	0.629	–	–	–	0.106
6	–	0.099	–	–	–	0.550	0.185	0.060	0.051
7	0.093	0.139	–	0.055	–	–	0.597	–	0.184
8	0.070	0.112	–	–	0.130	0.102	0.130	0.419	0.072
total exits	0.200	0.246	0.064	0.064	0.115	0.052	0.197	0.061	

Table 3. Fractions of out-in gate pairs

(a) Observed out-in pairs

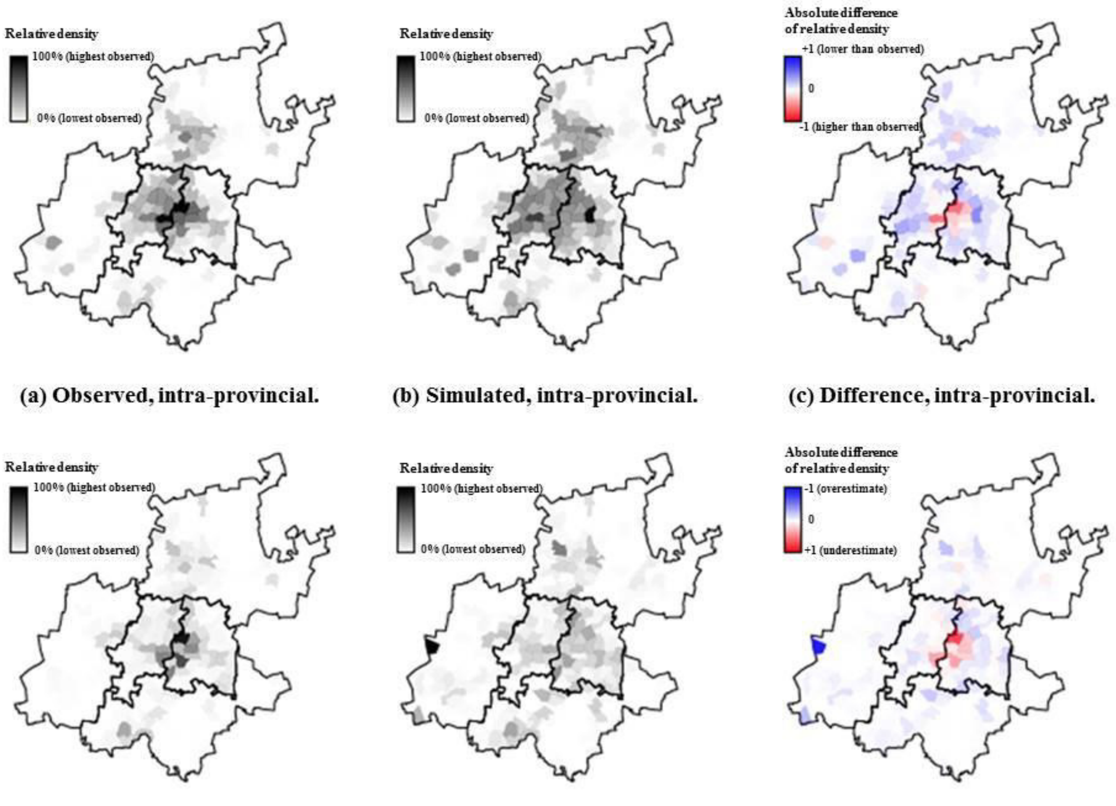
From	To 1	2	3	4	5	6	7	8	Total entries
1	0.855	0.087	–	–	–	–	–	–	0.097
2	–	0.769	0.176	–	–	–	–	–	0.241
3	–	0.209	0.702	0.072	–	–	–	–	0.099
4	–	–	0.074	0.851	–	–	–	–	0.086
5	–	–	–	–	0.802	–	–	0.139	0.170
6	–	–	–	–	–	0.767	0.079	0.127	0.068
7	0.067	–	–	–	–	–	0.862	–	0.160
8	–	–	–	–	0.241	0.063	–	0.644	0.078
total exits	0.106	0.223	0.120	0.087	0.162	0.064	0.152	0.087	

(b) Synthetic population out-in pairs

From	To 1	2	3	4	5	6	7	8	Total entries
1	0.843	0.102	–	–	–	–	–	–	0.102
2	–	0.781	0.161	–	–	–	–	–	0.259
3	–	0.190	0.710	0.081	–	–	–	–	0.091
4	–	–	0.093	0.839	–	–	–	–	0.084
5	–	–	–	–	0.837	–	–	0.105	0.202
6	–	–	–	–	–	0.684	0.137	0.158	0.041
7	0.061	–	–	–	–	–	0.873	–	0.157
8	–	–	–	–	0.309	–	–	0.658	0.065
total exits	0.108	0.235	0.115	0.084	0.199	0.035	0.150	0.073	

4.3. Spatiotemporal comparison

So far we've mainly addressed the spatial aspects of the activity chain validation. In this section we add a temporal dimension. Spatially we aggregate the activity density to the mesozone level provided by the Geospatial Analysis Platform (GAP), a near equal-area demarcation that follows important administrative and physiographic boundaries (CSIR Built Environment, 2012). Temporally the activities are aggregated on an hourly basis. We again distinguish between intra- and inter-provincial activity chains and illustrate the comparison for the tenth hour in Fig. 4. Over the entire 24-hour horizon the simulated chains compare favorably, both spatially and temporally.



(a) Observed, intra-provincial. (b) Simulated, intra-provincial. (c) Difference, intra-provincial.

(d) Observed, inter-provincial. (e) Simulated, inter-provincial. (f) Difference, inter-provincial.

Fig. 4. Spatiotemporal comparison of activity densities during the tenth hour. Observed and simulated densities are expressed relative to the highest daily density. The differences between observed density, o, and simulated density, s, are calculated and expressed as o – s.

Densities for the observed and simulated maps are scaled relative to the highest daily value for that map. Consider for example a zone, say zone 1, for which the observed activity count in the sixth hour was 10. If the highest observed activity count throughout the day was 350 in another zone, then zone 1's relative density will be $10/350 \approx 2.9\%$. It follows that the relative density for any map will have a range between 0% (no activities observed) and 100% (highest number of activities for the day observed).

In any hour h the difference between the relative observed density in zone z, o_z^h, and the relative simulated density in that zone and hour, s_z^h, is a measure of how accurate our synthetic population is in both space and time simultaneously. A positive difference, that is when $o_z^h - s_z^h > 0$, implies that the synthetic population is underestimating the activity density and shows fewer activities taking place in that hour in that particular zone.

Conversely, a negative difference implies an overestimation with more activities simulated in that hour than what was observed in that particular zone.

We chose the tenth hour as it has, for both intra- and inter-provincial chains, the highest activity density. In the case of intra-provincial chains we see a similarity in the general distribution. This is indeed a positive result since the locations were not sampled from a kernel density estimate, but is rather the result from the activity sequences from the complex network.

Recall that the complex network from which we've sampled the activity sequences is the result of a clustering algorithm. The clustering resulted in many activity locations being omitted from the complex network. Intuitively one may expect our synthetic activity chains (Fig. 4b) to then be more concentrated around the same facilities that did make it into the complex network. Yet, from Fig. 4a we see that the observed chains were actually more concentrated. One argument might be that the activity locations that were omitted following the density based clustering, may indeed be centrally located. This is an avenue that could be investigated in future. The clustering parameters evaluated by Joubert & Axhausen (2013) considered a gold standard based on expert opinion. One may wish to choose different parameter values: values that find a balance not with the best match to the standard, but rather a play-off between the computational burden of extracting and sampling from a complex network, and the number of activity locations dropped. In an ideal situation the complex network should contain all the activity locations, but this is indeed not computationally viable (yet).

The intra-provincial chains are dominated by the high number of simulated activities (entries and exits) that appear at the sixth gateway. We could not find any particular reason for the high number of simulated activities.

5. Conclusion

The simulated population created in this paper succeeded in addressing the drawbacks of earlier work. Firstly, it considered both intra- and inter-provincial commercial vehicles and generated independent subpopulations, each representing the original observed subpopulations accurately. Secondly, the use of complex networks to sample activity sequences from yielded more accurately vehicle kilometrekilometres travelled while remaining fairly accurate in both space and time.

Since the simulated population is the result of many random draws, variation will necessarily occur between different simulated populations, and rightfully so. We are continuing with this research to consider the repeatability of the results for different simulated populations, and in doing so establish a confidence interval for the validation results.

Two avenues of further research are noted. In this paper we spread the activities evenly over the duration of the chain, with little cognizance of the distance between consecutive activities, or the time of day at which the trip between them occur. The spatiotemporal validation results may indeed improve further if a suitable heuristic can be implemented. The second avenue would be to simulated the proposed population(s) and evaluate a) how the commercial vehicles adapt their activity chains in the co-evolutionary setting such as MATSim; and b) how well the simulated population compares against actual traffic patterns.

This commercial vehicle population will serve as a base commercial vehicle population, which will be simulated in conjunction with private vehicles. Recent developments in commercial vehicle modelling by Schroeder, Zilske, Liedtke & Nagel (2012) have seen the development of a freight framework in MATSim that provides the capability to model behavior-rich commercial vehicle traffic. Such a population can then be introduced into the model by iteratively replacing activity chains from the base population by activity chains that result from behavioral interactions found in supply chains.

Acknowledgements

We are grateful to Digicore Fleet Management for the continued use of the large GPS sample for our research. This work was partially funded by the South African National Roads Agency Limited (SANRAL). The first author is also grateful to the South African National Research Foundation (NRF) for funding. All graphs and maps were produced using R (R Core Team, 2012).

References

Al-Deek, H. M., Johnson, G., Mohamed, A., & El-Maghraby, A. (2000). Truck trip generation models for seaports with container and trailer operation. *Transportation Research Record*, *1719:*1-9.

Barthélemy, M. (2011). Spatial networks. *Physics Reports*, *499(1-3),* 1-101.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424(4-5),* 175-308.

Chow, J. Y. J., Yang, C. H., & Regan, A. C. (2010). State-of-the art of freight forecast modelling: lessons learned and the road ahead. *Transportation*, *37(6):*1011-1030.

CSIR Built Environment (2012). GAP. Accessed 13 March 2012 from http://www.gap.csir.co.za/.

Fourie, P. J. (2010). Agent-based transport simulation versus equilibrium assignment for private vehicle traffic in Gauteng. The 29th Annual Southern African Transport Conference, pp. 12-23.

Gao, W., Balmer, M., & Miller, E. J. (2009). Comparison of MATSim and EMME/2 on Greater Toronto and Hamilton Area Network, Canada. *Transportation Research Record, 2197,* 118-128.

Hart, P. E., Nilsson, N. J. & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics SSC4*, *4(2),*100-107.

Hensher, D., & Figliozzi, M. A. (2007). Behavioural insights into the modelling of freight transportation and distribution systems. *Transportation Research Part B: Methodological*, *41(9):* 921-923.

Holguin-Veras, J., & Patil, G. (2005). Observed trip chain behavior of commercial vehicles. *Transportation Research Record*, *1906:*74-80.

Joubert, J. W. (2012). Analysing commercial through-traffic. *Procedia - Social and Behavioral Sciences, 39,* 184-194.

Joubert, J. W. & Axhausen, K. W. (2011). Inferring commercial vehicle activities in Gauteng, South Africa. *Journal of Transport Geography*, *19(1),*115-124.

Joubert, J. W. & Axhausen, K.W. (2013). A complex network approach to understand commercial vehicle movement. *Transportation*. Forthcoming, Doi: 10.1007/s11116-012-9439-0.

Joubert, J. W., Fourie, P. J., & Axhausen, K. W. (2010). Large-scale agent-based combined traffic simulation of private cars and commercial vehicles. *Transportation Research Record*, *2168,* 24-32.

Liedtke, G. (2009). Principles of micro-behavior commodity transport modelling. *Transportation Research Part E: Logistics and Transportation Review, 45(5),* 795-809.

Marker Jr, J. T., Goulias, K. G. (1998). Response freight model under different degrees of geographic resolution | geographic information system application in pennsylvania. *Transportation Research Record, 1625:*118-123.

R Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ruan, M., Lin, J., & Kawamura, K. (2012). Modelling urban commercial vehicle daily tour chaining. *Transportation Research Part E: Logistics and Transportation Review, 48(6):*1169-1184.

Samimi, A., Mohammadian, A., & Kawamura, K. (2009). Behavioral freight movement modelling. In Resource Paper, Workshop 2: Behavioral paradigms for Modelling Freight Travel Decision-Making. The 12th International Conference on Travel Behaviour Research, Jaipur, India.

Schroeder, S., Zilske, M., Liedtke, G. & Nagel, K. (2012). Towards a multi-agent logistics and commercial transport model: The transport service provider's view. *Procedia - Social and Behavioral Sciences, 39,* 649-663.

van Heerden, Q. & Joubert, J. W. (2012). Commercial vehicle behaviour: analysing GPS records. In: 42nd Computers and Industrial Engineering (CIE) Conference, Paper 99.

Winter, M., & von Hirschhausen, C. (2006). Environmental HDV road charging for Berlin - Theoretical considerations and empirical estimations. Available online from http://ideas.repec.org/p/cni/wpaper/2006-01.html.