Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 184 (2021) 739–744

Procedia Computer Science

www.elsevier.com/locate/procedia

The 10th International Workshop on Agent-based Mobility, Traffic and Transportation Models,(ABMTRANS) March 23 - 26, 2021, Warsaw, Poland

A South African scenario for emissions modelling

Johan W. Joubert*, Ruan J. Gräbe

Centre for Transport Development, Industrial & Systems Engineering, University of Pretoria, 0028, South Africa

Abstract

Scenarios with high levels of diversity lend themselves to be represented or imitated by disaggregate, agent-based models as the individual agents can capture the unique attributes of the constituents of the overall system. This paper describes the methodology to establish a baseline transport simulation model, using MATSim, to represent and imitate vehicle-specific emissions. The study area is the multi-metropolitan megacity (and province) of Gauteng, South Africa.

Keywords: Agent-based simulation; MATSim; Emissions; South Africa

1. Introduction

In transport scenarios where there is high diversity, an agent-based approach proves to be beneficial to evaluate the impact of interventions as it is sufficiently expressive to capture the unique attributes of individuals (agents) and the environment [19]. In a country like South Africa with its high economic inequality (a Gini coefficient of 0.65 points in 2015) and extreme socioeconomic and population diversity (eleven official languages), capturing and accounting for the diversity is especially important when you want to ask questions like *"who gets the benefit from transport infrastructure and interventions?"* and, equally important, *"who pays for those benefits?"*

Developing countries are often eager to embrace state-of-the-art or emergent international policies, but often lacks the ability to capitalise on the benefits of those policies as they lack the supporting infrastructure to enable and enforce the policies. One such a case is the trend towards greening the transport industry and adopting environmental standards. Point in case is the South African government who subscribes to making transport more environmentally sustainable, through its *Green Transport Strategy* [3], but fails to adopt or commit to specific quantitative targets. Not adopting quantitative targets is not necessarily bad, in this case, as such targets will likely have been based on (inappropriate) international targets that do not take local conditions into account.

* Corresponding author. Tel.: +27-12-420-2843; fax: +27-12-362-5103.
E-mail addresses: johan.joubert@up.ac.za (Johan W. Joubert)., ruan.grabe@up.ac.za (Ruan J. Gräbe).

10.1016/j.procs.2021.03.092

This paper describes the process to establish a baseline scenario in the Multi-Agent Transport Simulation (MATSim) that can be used to evaluate emissions. The scenario covers the multi-metropolitan economic centre of the country, the province of Gauteng, which is made up of the City of Johannesburg, Tshwane (Pretoria) and Ekurhuleni, along with two district municipalities: Sedibeng and West Rand. The province accounts for less than 2% of the country's land surface but 25% of the population and more than a third of the country's gross domestic product (GDP). Why MATSim?

Macal [14] notes that not all *agent-based* models are equal and one should distinguish between four properties. At the most basic level, the first property is that of *individuality*; accounting for each agent and its unique attributes. The second property, *autonomy*, implies that each agent is able to make its own decisions based on its individual characteristics. Thirdly, *interactivity* as a property implies that those behaviours are independently or interdependently with other agents and its environment. Finally, an *adaptive* agent-based model implies that agents can adapt and learn throughout the course of the simulation. MATSim is attributed with all four the properties [8]. It models a population of agents, each representing a person with unique attributes, as they execute their daily plans on a transport network. A *plan* is a sequence of activities and trips that is richly described in terms of timing (start time and duration), location (detailed facilities or coordinates instead of zones) and mode. The agent's experience is scored using a generalised cost function that accounts for the (positive) utility of participating in value-adding activities and the (negative) utility of travelling and incurring cost to overcome the distance between activities. The experience is based on and influenced by multiple agents trying to execute their individual plans on the limited infrastructure, causing congestion. Autonomous decision-making is embedded in the MATSim machinery as exchangeable and complimentary modules like changing the timing of one's activity, altering your route, or changing mode, depending on the supporting data available and the intent of the model investigation. As agents adapt their daily plans and execute the revised plans iteratively, they build up a memory of plans, favouring those that promise a higher expected utility. This co-evolutionary machinery of MATSim allows for a relaxed state to be achieved over a sufficient number of iterations when no agent can consistently improve their own state of maximising its utility.

The individual attribute specific to this paper is that of the emissions concept (Euro 2, Euro 3, etc) of the vehicle type allocated to each individual. Each vehicle's unique emissions concept is then taken into account during the mobility simulation of MATSim [12] as it looks up the detailed cold and hot emissions (g/km) for different traffic conditions from the Handbook Emission Factors for Road Transport (HBEFA), version 4.1 [15]. While similar studies have used MATSim to evaluate emissions interventions [1, 9, 11, 13], little attention was given to how the initial scenarios is set up.

This paper is structured as follows. In Section 2 we explain how the travel demand was converted from Gauteng's travel matrices. Section 3 describes the process to find the emission characteristics of the population of vehicles in South Africa. The execution of the model and its results are described in Section 4. We also address the issue of model validation in Section 4.1. finally, we conclude with a brief research agenda in Section 5.

2. Demand generation

Traditional transport models heavily rely on origin-destination (OD) matrices to describe the trip distribution patterns. Often these OD matrices only cover a short time window; typically the morning peak as it tends to have a higher traffic density than the afternoon peak. Consequently, the majority of trips accounted for are the routine home-to-work commute. The value of activity-based travel demand is that it typically covers and entire day and a richer activity set. But such activity-based models also require richer data as as input, which are often detailed trip diaries completed through surveys [10, 18].

Because such data detailed data is frequently not that accessible, a number of research contributions already showed how a trip based model's OD matrices can be converted into an activity-based set of travel chains for discrete individuals. Both Fourie [6] and Gao et al. [7] showed how *EMME/2* data is converted into a MATSim population of agents, and both confirmed that the agent-based counterpart yields more accurate travel time predictions and richer overall result sets.

For Gauteng, South Africa, this is the approach that we will follow here too. The Saturn-based transport model of the South African National Road Agency Limited (SANRAL) for Gauteng for 2016 is validated in [16] and its OD matrices are used as a point of departure. Each trip is accounted for with a single agent with a `home` location in the trip's origin zone, and the work location in the destination zone of the trip. Along with the OD matrices there is a detailed Geographic Information System (GIS) shapefile depicting the set of transport analysis zones, $\boldsymbol{Z}$, on which the matrices are based. For the Gauteng model there are five square matrices, $\boldsymbol{T}^m$, where $m \in \boldsymbol{M} = \{1, \dots, 5\}$ such that

$$\boldsymbol{M} \triangleq \begin{cases} 1 & \text{private car commuter trips,} \\ 2 & \text{light vehicle business trips,} \\ 3 & \text{other light vehicle trips,} \\ 4 & \text{light heavy goods vehicle trips, and} \\ 5 & \text{heavy goods vehicle trips.} \end{cases}$$

For each zone $i \in \boldsymbol{Z}$, the trip production, denoted by $T_i^{\text{prod},m}$, for trips of type $m \in \boldsymbol{M}$ is calculated as the ceiling of the column sum, expressed in (1),

$$T_i^{\text{prod},m} = \left\lceil \sum_{j \in \boldsymbol{Z}} \left(t_{ij}^m \right) \right\rceil \qquad \forall i \in \boldsymbol{Z}, m \in \boldsymbol{M} \tag{1}$$

where t_{ij}^m is the matrix value for the number of trips of type $m \in \boldsymbol{M}$ from zone $i \in \boldsymbol{Z}$ to zone $j \in \boldsymbol{Z}$. Since we work with discrete agents, we need an integer number of trips. To generate the synthetic population, you work your way through each origin zone $i \in \boldsymbol{Z}$ of each type of trip. The home location of an agent travelling from zone $\in \boldsymbol{Z}$ is a random point (coordinate) sampled within the spatial polygon of zone i. The work location is, similarly, a random point in the spatial polygon of destination zone $j \in \boldsymbol{Z}$. The destination zone is sampled using a cumulative probability distribution where the probability of travelling to each zone j, denoted by p_{ij}^m, is calculated using (2).

$$p_{ij}^m = \frac{t_{ij}^m}{\sum\limits_{j' \in \boldsymbol{Z}} t_{ij'}^m} \tag{2}$$

For the Gauteng model, there were a total of 903 origin and destination zones, 560 415 car commuter trips ($m = 1$), 16 128 car business trips ($m = 2$), 33 471 other light vehicle trips ($m = 3$), 13 026 light heavy goods vehicle trip ($m = 4$) and 8 101 heavy goods vehicle trips ($m = 5$).

The departure time from the home activity is uniformly distributed between 6 a.m. and 9 a.m. Since the matrices are focussing on the morning peak trips only, there is no explicit return journey home. In our scenario, we add a final home activity to the agents' activity chain and sample the departure time from the preceding activity, work, to be randomly distributed between 3:30 p.m. and 6 p.m.

For the road network, we extracted the coarse road network from *OpenStreetMap*. The coarse network include all roads and links associated with (high)ways that are tagged with motorway, trunk, primary, secondary and tertiary. This road network is already much denser than the original Saturn road network linking the TAZs.

3. Vehicle population

Once the population of individuals is created, the next step is to assign a vehicle to each. A vehicle should have a specific type that accounts for its emissions concept. That is, it should distinguish on the vehicle class (passenger car or heavy goods vehicle), fuel type (petrol or diesel, in South Africa) and on the Euro class, for example Euro 4.

But such a detailed breakdown of the vehicle type is not available in public data. For example, the electronic National administration Traffic Information System (eNaTIS) provides a live vehicle population [5]. eNaTIS only indicates the number of registered vehicles in each province, and only distinguishes between the classes shown in Table 1. No fuel-related or emissions data is available for any of the vehicle types. Consequently, this paper employs a

Table 1. Live eNaTIS vehicle population for the province of Gauteng (July 2020).

Vehicle class	Province Gauteng	Percentage of total
Motor cars and station wagons	3 128 479	**70.34**
Minibuses	130 207	2.93
Buses, bus trains, midibuses	20 460	0.46
Motorcycles, quadrucycles, tricycles	139 445	3.13
LDV's, panel vans, other light load veh's GVM ≤ 3500kg	853 293	19.18
Trucks (Heavy load vehicles GVM > 3500kg)	140 033	3.15
Other	36 029	0.81

strategy to infer the emissions concepts using second-hand vehicle sales data as a proxy for the distribution of vehicle types in the country.

Even though we ignore light delivery vehicles (LDVs), comprising about 20% of the vehicle population, this approach still caters for at least 70% with the passenger vehicle class (shown in bold in Table 1). We sample[1] from over 65 000 listed passenger vehicles from [2] to obtain the proportion of fuel types[2]. For each sampled vehicle we infer its emissions concept from secondary sources like [17] or the Australian [4] (see Table 2).

Table 2. Vehicle sampling results for 20 years. Each row contains the Euro emissions concept % of the vehicles sampled in that year.

Year	Sample population	Petrol Euro concept					Diesel euro concept				
		2	3	4	5	6	2	3	4	5	6
2000	0.09	5.3	77.0	—	—	—	2.7	15.0	—	—	—
2001	0.14	5.3	77.0	—	—	—	2.7	15.0	—	—	—
2002	0.13	5.3	77.0	—	—	—	2.7	15.0	—	—	—
2003	0.21	5.3	77.0	—	—	—	2.7	15.0	—	—	—
2004	0.32	5.3	77.0	—	—	—	2.7	15.0	—	—	—
2005	0.50	—	37.1	46.8	—	—	—	8.1	8.1	—	—
2006	0.67	na	na	na	na	na	na	na	na	na	na
2007	1.03	na	na	na	na	na	na	na	na	na	na
2008	1.14	—	8.5	63.1	—	—	—	13.1	15.3	—	—
2009	1.27	—	6.5	62.0	5.0	—	—	4.5	20.5	1.5	—
2010	2.20	—	3.6	57.1	11.4	—	—	4.3	18.6	5.0	—
2011	3.11	0.7	—	23.0	46.0	—	—	0.7	2.2	27.3	—
2012	4.29	3.0	3.0	18.0	56.0	—	—	4.0	4.0	12.0	—
2013	5.39	—	—	9.7	53.6	1.9	—	—	7.1	27.7	—
2014	7.00	—	—	19.5	36.3	10.6	—	—	5.3	25.7	2.7
2015	8.50	—	0.7	9.2	29.0	30.9	—	2.0	2.0	18.4	7.2
2016	9.70	—	1.8	10.0	23.8	19.6	—	1.1	4.4	26.5	11.1
2017	9.30	—	—	8.7	31.4	30.8	—	—	5.4	15.1	8.7
2018	10.10	5.0	4.0	15.0	13.0	44.0	—	1.0	3.0	6.0	9.0
2019	na	na	na	na	na	na	na	na	na	na	na
2020	18.50	—	—	3.0	24.1	37.6		—	6.8	15.2	12.0
Median		0.86	2.32	13.45	28.15	23.97	0.03	1.28	5.56	16.82	7.11

[1] "na" values indicate years from which sampling is yet to be done.
[2] We disregard all but the main fuel types (petrol and diesel) in our South African context.

With the detailed emissions concepts known for each year, we follow a Monte Carlo sampling approach to estimate the overall likelihood of a specific emissions concept. A single run will first sample the production year of the vehicle, using the probabilities from the *Sample population* column in Table 2. Next we sample the Euro emissions concept conditional on the production year. This is repeated for just over 10 000 times. One can now, for this run, tally the total number of each of the ten emission concepts.

We repeat this process for en ensemble of 1 000 runs and report the median values of the (symmetric) distributions for each emissions concept at the bottom of Table 2.

4. Results

With the probabilities for each emissions concept estimated, the next step is to assign a vehicle type to each agent. This is done by sampling a concept using a cumulative probability distribution of the median probabilities. At this point in time, no additional person attributes are taken into account to discern, for example, that a more affluent individual is more likely to have a newer and environmentally friendlier (higher Euro concept) vehicle. This is left for future work as it will require additional proxy variables like the estimated value of each vehicle. Also, the current scenario does not simulate the two heavy vehicle classes as there is no reliable way (yet) to estimate the vehicle sizes and emission concepts for these vehicles. Consequently, these agents are removed from the population of agents.

The remaining initial travel demand of light vehicles, along with the vehicles and the road network, is loaded into the MATSim scenario and executed. Without optimising the computational setup, the simulation runs for approximately 8 200 seconds per iteration for the 611 452 agents on a multi-core DELL server with 80GB of RAM allocated to the process. Within the MATSim configuration, 8 threads are made available for both the mobility simulation (taking up the majority of computational time) and the events processing.

For the size of the network, and the number of agents, a total of 251-million events are generated during the mobility simulation. For this paper, only the cold and warm emission events are of interest. After the simulation, these two event types are extracted, and aggregated per vehicle. One valuable characteristic of an high-resolution, agent-based setup is that one can aggregate to any required level. While we only report the per-vehicle emissions here, one can easily introduce both spatial and temporal dimensions to the results because for each emissions event, we also know the detailed link on the network (spatial) and the time when the vehicle left the link (temporal). The distribution of the per-vehicle CO_2, computed as the total carbon dioxide from fuel consumption, is shown in figure 1. Overall, a

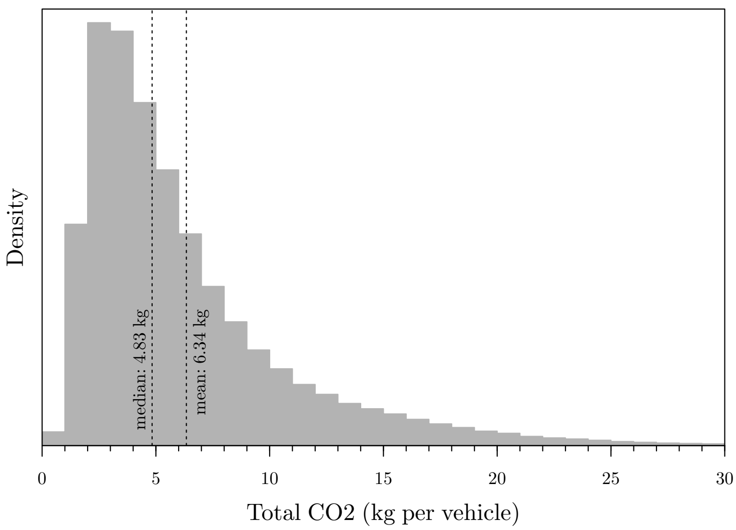

Fig. 1. Distribution of the total CO_2 produced from home-work-home trips.

total of ±3 900 tonnes of CO_2 was emitted over a 24-hour period; 28.8 tonnes of CO (carbon monoxide); 8.0 tonnes of NO_x.

4.1. Validation

The HBEFA database is specifically geared towards European countries: Germany, Austria, Switzerland, Sweden, Norway and France. To calibrate the emissions for (South) African conditions, while outside the scope of the current paper, is worth commenting on here. The University of Pretoria's Centre for Transport Development has acquired a Portable Emissions Measurement System (PEMS) that allows the research group to measure the actual emissions of a vehicle while driving under normal conditions. With *normal conditions* we imply the actual elevation of Gauteng (in excess of 1 300 m above sea level), the road grade, weather conditions and driver behaviour. The PEMS unit is, at the time of writing this article, used to build up a database of emissions for a variety of South African vehicles (both light and heavy duty) on standardised routes and using multiple drivers. For each combination of vehicle, route and driver, multiple trips are conducted to establish the variation observed during real driving. Estimating the variation will allow us to correct the simulated emissions (over an ensemble of simulated runs) to that of the real drive emissions (RDE) observed.

5. Conclusion

In this paper, we present the methodology to establish a baseline agent-based (MATSim) scenario for Gauteng, South Africa, to model the emissions emitted from (mainly) commuting vehicles over a 24-hour period. The authors would like to acknowledge Gabriela Murta, Lynnet Mhlongo, Paul Jacobs, Dawid Steenkamp and Sheunesu Toro, members of the Vertically Integrated Projects (VIP) Programme's MOBility team for their contributions.

References

[1] Agarwal, A., Kickhöfer, B., 2015. Agent-based Simultaneous Optimization of Congestion and Air Pollution: A Real-World Case Study. Procedia Computer Science 52, 914–919. doi:10.1016/j.procs.2015.05.165.
[2] AutoTrader, 2020. Find new & used cars for sale. Available at: https://www.autotrader.co.za. [Accessed: 2020-09-18].
[3] Department of Transport, 2019. Green Transport Strategy for South Africa: (2018–2050). Available online from https://www.transport.gov.za/.
[4] Drive, 2020. What's my car worth? Available at: https://www.drive.com.au/whats-my-car-worth. [Accessed: 2020-09-18].
[5] eNaTIS, 2020. Live Vehicle Statistics from July to August 2020. Available at: http://www.enatis.com/index.php/downloads/doc_download/1857-livevehpopperclassprov20200831. [Accessed: 2020-09-18].
[6] Fourie, P.J., 2009. Agent-based transport simulation versus equilibrium assignment for private vehicle traffic in Gauteng, in: 29th Annual Southern African Transport Conference.
[7] Gao, W., Balmer, M., Miller, E.J., 2010. Comparison of MATSim and EMME/2 on Greater Toronto and Hamilton Area Network, Canada. Transportation Research Record 2197, 118–128. doi:10.3141/2197-14.
[8] Horni, A., Nagel, K., Axhausen, K.W. (Eds.), 2016. The Multi-Agent Transport Simulation MATSim. Ubiquity Press.
[9] Hülsmann, F., Gerike, R., Kickhöfer, B., Nagel, K., Luz, R., 2011. Towards a multi-agent based modeling approach for air pollutants in urban regions, FGSV-Verl.. pp. 144–166. Available at https://depositonce.tu-berlin.de/handle/11303/10353.
[10] Ilhai, A., Balac, M., Li, A., Axhausen, K.W., 2019. The first agent-based model of greater Jakarta integrated with a mode-choice model. Procedia Computer Science 151, 272–278.
[11] Kickhöfer, B., 2014. Economic policy appraisal and heterogeneous users. Ph.D. thesis. Technische Universität Berlin.
[12] Kickhöfer, B., 2016. Emission modeling, in: The Multi-Agent Transport Simulation MATSim. Ubiquity Press, pp. 247–252. doi:10.5334/baw.36.
[13] Kickhöfer, B., Agarwal, A., Nagel, K., 2018. Mind the price gap: How optimal emission pricing relates to the EU CO2 reduction targets. International Journal of Sustainable Transportation 13, 378–391. doi:10.1080/15568318.2018.1472321.
[14] Macal, C.M., 2016. Everything you need to know about agent-based modelling and simulation. Journal of Simulation 10, 144–156. URL: https://doi.org/10.1057/jos.2016.7, doi:10.1057/jos.2016.7, arXiv:https://doi.org/10.1057/jos.2016.7.
[15] Notter, B., Keller, M., Cox, B., 2019. The Handbook Emission Factors for Road Transport. Available at: https://www.hbefa.net/e/index.html. [Accessed: 2020-08-17].
[16] Robinson, A., Venter, C., 2019. Validating traffic models using large-scale Automatic Number Plate Recognition (ANPR) data. Journal of the South African Institution of Civil Engineering 61, 45–57.
[17] True Rating, 2020. The true rating - the real urban emissions. Available at: https://www.trueinitiative.org/true-rating. [Accessed: 2020-09-18].
[18] Zhuge, C., Shao, C., Gao, J., Meng, M., Xu, W., 2014. An Initial Implementation of Multiagent Simulation of Travel Behavior for a Medium-Sized City in China. Mathematical Problems in Engineering 2014, 1–11. doi:10.1155/2014/980623.
[19] Ziemke, D., Kaddoura, I., Nagel, K., 2019. The MATSim Open Berlin Scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. Procedia Computer Science 151, 870–877.