Journal of Public Transportation 25 (2023) 100042

Contents lists available at ScienceDirect

Journal of Public Transportation

journal homepage: www.journals.elsevier.com/journal-of-public-transportation

BRT network design for transit cost reduction in Cape Town, South Africa

Obiora A. Nnene[a,*], Mark H.P. Zuidgeest[a], Johan W. Joubert[b]

[a] Department of Civil Engineering, University of Cape Town, Cape Town, South Africa
[b] Centre for Transport Development, Industrial & Systems Engineering, University of Pretoria, South Africa

ARTICLE INFO

Keywords:
Bus rapid transit
Operational cost
Multi-objective optimisation
Public transport network
Service improvements

ABSTRACT

Car dependency in Cape Town, South Africa, like in many modern cities is high but comes at a high cost. One of these is significant network congestion, especially during peak commuting periods. A strategy adopted by the City's transport development authority to alleviate this situation is the phased development of an integrated public transport network. A bus rapid transit service known as the MyCiTi BRT is proposed to be the backbone of the integrated network. However, an investigation of the existing service reveals a pattern of low ridership and high operating costs which is not economically sustainability. The main objective of this work is therefore to propose solutions to improve the existing service particularly in terms of increasing ridership and reducing operational cost. To achieve this, a network design approach known as simulation-based optimisation that combines simulation with optimisation algorithms is proposed. The results are compared with the baseline MyCiTi network. The outcome shows that the proposed methods are capable of improving the MyCiTi network and thereby its service.

1. Introduction

The City of Cape Town (see Fig. 1) is located in the Western Cape province of South Africa and has an estimated land area of 2455 km^{2}TDA (2015). According to RHDHV (2014), the city's population was 4.04 million inhabitants in 2017, and it is projected to grow to about 4.5 million by 2032. Public transport planning in the city is done by the Transport Development Authority (TDA), with travel demand modelling and network design being some of the authority's most important activities. An estimated 68% of the city's population are of working age, hence, economically active. This translates to a very high demand for travel and quality transport services. In terms of the modal network characteristics for the morning and evening peak periods, there is a 53:38 split between private cars and public transport among all travellers RHDHV (2014). Non-motorised transport accounts for the remaining 9% of the modal share. The high dependence on private cars give rise to significant network congestion in the city, especially during the peak commuting periods. This trend will likely worsen as the population of Cape Town grows and travel demand increases. One of the strategies adopted by the TDA in Cape Town to alleviate this situation is the phased development of an integrated public transport network (IPTN), which is a public transport network planned in anticipation of the future effect of urban growth on travel demand in the city with the goal of having the IPTN fully operational by 2032. This plan entails an expansion of the city's existing public transportation network. This is in order as the population of the city is estimated to increase by about 37% in the target year. The current network comprises a bus network known as the Golden Arrow Bus Service (GABS), a bus rapid transit (BRT) network, minibus taxis and a metro rail service. When completed, it is expected that BRT and rail would form the backbone of the IPTN. The MyCiTi service which is the main focus in this paper began its operations in 2010 as part of a nationwide roll out of BRT services across South Africa. It is expected that a bus rapid transit (BRT) service known as the MyCiTi BRT will be a significant component of the IPTN when the latter is completed. The system utilises recent technologies like automated fare collection (AFC), closed transfer facilities and level boarding platforms. It will be rolled out in four stages, with the full system ready for service in about twenty years. The first phase routes of the service were launched officially in 2011. Since then, new routes have been incrementally developed to expand the service's coverage within the city. Presently, two express services are undergoing testing between the Cape Town Central Business District (CBD) and the city's South-Eastern axis for the second phase of operations. The network consists of 472 nodes and about 46 operational routes. The stop locations have different configurations, namely main stations and smaller stop couplets. Stations are closed areas on the system, that allow

* Corresponding author.
E-mail address: obiora.nnene@uct.ac.za (O.A. Nnene).

https://doi.org/10.1016/j.jpubtr.2023.100042
Received 10 April 2022; Received in revised form 22 November 2022; Accepted 9 March 2023
Available online 17 March 2023
1077-291X/

Fig. 1. Map showing the Western Cape Province and the City of Cape Town adapted from Viljoen and Joubert (2019).

passengers to enter the system before they board a vehicle. The platforms allow travel in either direction along a route. On the other hand, the stop couplets, are open areas consisting of two stops on opposite sides of a roadway. Fig. 2 shows the current network of the MyCiTi service and its coverage in the Cape Town. Also visible in the map are the other land-based services mentioned earlier, namely GABS and minibus taxis. The BRT network runs from the CBD to Atlantis in the north and southeast towards Mitchells Plain and Khayelistsha. The service also extends southwest to the Hout Bay.

Observed trends on the MyCiTi BRT reveals the need to improve the service on different fronts such as network optimisation, operational cost reduction and even improvement in terms of its administrative structure. Officials at the TDA indicate that current trip patterns are not conducive to operate economically sustainable transit services. The primary travel demand problems recognised by the MyCiTi service planners is that of low patronage on the network. Hence, without a substantial change in the way commuters travel or user perception of the service, the revealed demand on existing routes would not change. Another impact of the low passenger ridership on the BRT network is the increase in the operational costs and subsidy requirement of the system. Therefore, increasing ridership on the system would be beneficial, since this would reduce the amount of subsidies required to operate the service. In line with this, the objective of this research is to improve the MyCiTi BRT network in the areas of passenger ridership increase and operational cost reduction. To achieve this a novel transport network design model which combines agent-based simulation and multi-objective optimisation is used. This approach is generally known as Simulation-based optimisation (SBO). A key feature of this approach is that it adequately accounts for the randomness in commuter behaviour on the transport network, by simulating their individual choices with agent-based modelling. Commuters normally prefer high frequency transit services that are safe, direct, comfortable and affordable. Operators, on the other hand, prefer to provide their service in a way that minimises operational expenses. Hence, this conflict in stakeholder perspectives is also taken into account in the proposed solution. The remainder of the paper is structured as follows. In section two a literature review for the proposed work is presented. Section three discusses the data requirements for the work and how they are processed. Section four presents the model, its components and verification. In section five, the results of testing and applying the proposed model is

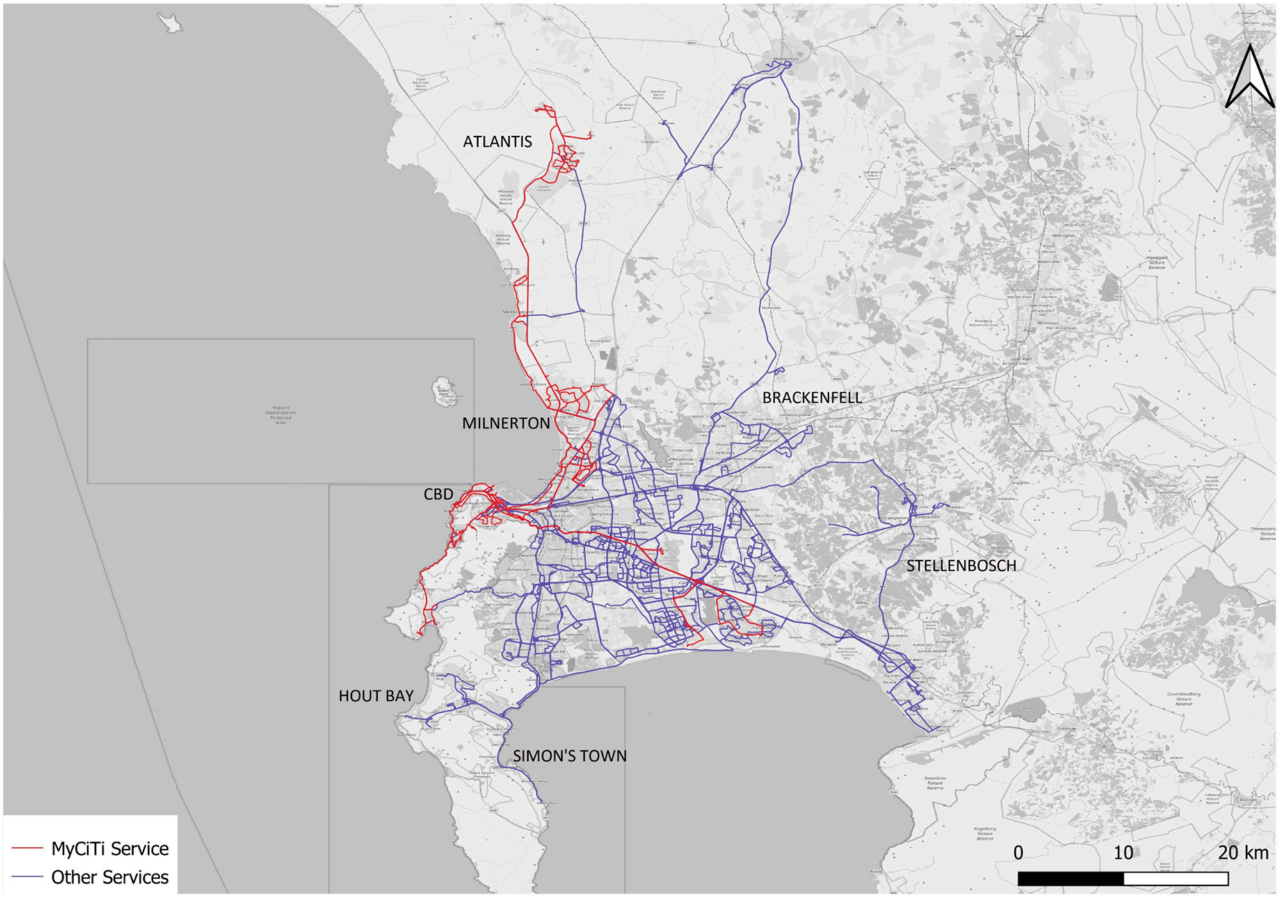

Fig. 2. Map showing the MyCiTi BRT and other land-based services.

presented and discussed. In the final section, possible areas of future research are highlighted.

2. Literature review

2.1. Transit network design

Public transport systems play a critical role in improving mobility and access to opportunities, which is crucial for the socio-economic growth and well-being of individuals and the society in general. A direct benefit of an optimised network can be seen in the reduction of transit costs that accrue network user (travel time) and operators(operational cost). Furthermore, optimising a public transit network will most likely lead to increased ridership from mode shifts and latent demand. This is normally considered a desirable outcome by transit agencies since it inmproves system efficiency and resource utilisation. A public transit network is a system of connected transit lines that support the operation of public transportation services. In the literature, the design of public transit networks is broadly classified as a Transit Network Design Problem (TNDP) as it deals with the optimized design of public transit route networks and the determination of their operational parameters. According to Ceder (2015) and Ibarra-Rojas et al. (2015), the network design and frequency setting activities of the transit network design problem (TNDP) are respectively classified as strategic and tactical while vehicle and crew scheduling are operational activities. Network design is considered strategic because it involves long term planning and requires less detail. Conversely, frequency setting is done in a shorter time horizon but involves a higher level of detail, hence, its description as a tactical activity. Lastly, the vehicle and crew scheduling or operational activities take place in the shortest time and require microscopic route level details (see Fig. 3). The work discussed in this paper gives a strategic outlook to network design and improvement with a focus on total network cost reduction. This entails that a detailed service planning scheme involving crew and vehicle scheduling is outside the scope of the paper.

Transit network design formally deals with finding a set of routes and their operational frequencies that best address the stated goals of multiple network stakeholders. The network stakeholders are normally commuters and operators who aim to minimise the cost they incur on the network. Hence, the objective of a TNDP is to improve a public transport network through traveltime reduction for passengers and operational cost minimisation for service providers. The problem is an optimisation problem, subject to a set of discrete or continuous constraints such as network configuration, route choice and service headway. Other considerations are the feasibility constraints on route length, vehicle capacity and fleet size. The above description highlights the features of the TNDP as a combinatorial multi-objective optimisation problem. It is *combinatorial* in the sense that the goal is to find an optimal set of routes and their operating frequencies among a finite set of alternatives Schrijver (2003). It is also a *multi-objective* optimisation problem because of the many conflicting objectives of different stakeholders such as network users, operators and even the public transit authorities Buba and Lee (2018).

Methods of solving the TNDP are categorised as analytical and heuristic. Analytical methods utilise *exact* algorithms such as the *branch and bound* and *branch and cut* algorithms which attempt to find the closed form of an objective function in the search for a unique solution to the problem. Instances of research done with analytical methods in the literature are (Ouyang et al., 2014; Chien et al., 2001). However, analytical solution models for the TNDP are often criticised as being inadequate to solve the problem due to the non-convex and np-hard nature of the problem which renders them very difficult to solve (Chen

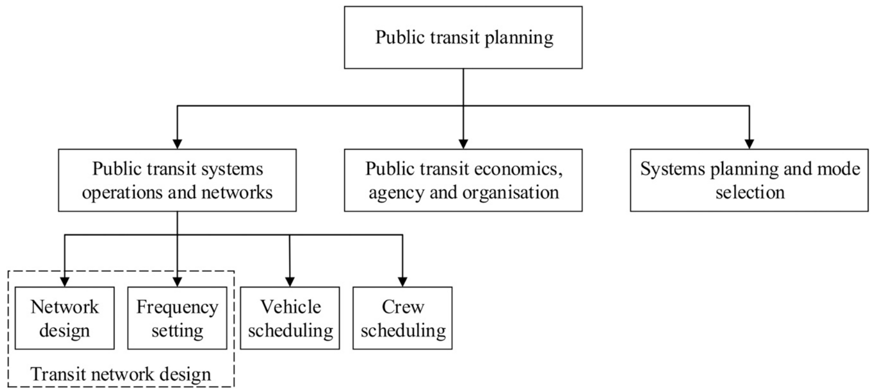

Fig. 3. Components of the public transport planning process adapted from Ceder (2007).

et al., 2017; Nnene et al., 2017). Another criticism of analytical solution methods is their limited scope of application to idealistic or small-sized problems. More recently, these shortcomings have led to the development of a type of heuristic algorithms known as *metaheuristics*. These techniques use *approximate* algorithms which can find *good* solution(s) in a reasonable amount of time. Their major strength is that they are not problem dependent Fan and Machemehl (2004), hence, they can be used to solve a broader array of problems spanning various knowledge domains. To this end, the metaheuristic algorithms can be applied to large scale or realistic TNDP problems, since they can find acceptable network solutions within a limited period. Examples of such works are (Huang et al., 2018; Madadi et al., 2019; Nnene et al., 2019a,b). Other works like Brands and Berkum (2014) and Heyken Soares et al. (2019) use multiobjective metaheuristic algorithms that yield a Pareto set of optimal solutions rather than one single solution, with the set normally represents different tradeoffs in the problem. Similarly, this paper adopts the multiobjective metaheuristic known as the non-dominated sorting genetic algorithm (NSGA-II), which is known to be more amenable to multiobjective problems like the TNDP.

2.2. Simulation-based optimisation of transit networks

SBO deals with integrating optimisation and simulation to solve complex optimisation problems like the TNDP Gosavi (2015). A detailed review of the SBO literature may be found in (Alrabghi and Tiwari, 2015; Arisha and Abo-Hamad, 2010) and Amaran et al. (2016). In these works, some of the classifications used, include input variables (quantitative and qualitative), and the number of objectives (single and multi-objective) Hachicha et al. (2010). Others are by the type of parameter (discrete or continuous), and the optimisation procedure used. In terms of solution approaches, four prominent ones described in Arisha and Abo-Hamad (2010) are gradient-based methods, statistical methods, meta-models, and metaheuristics, though more are available in Amaran et al. (2016). In the context of transportation planning, SBO applications based on meta-models and metaheuristics are among those recorded in the literature. In the meta-model based techniques, an analytical model is used to approximate the objective function. Usually a mathematical function that mimics the input and output behaviour of a simulation's stochastic response is used to replace parts of the simulation. Doing this reduces the computational burden of the problem as the meta-model can generally be resolved with deterministic optimisation techniques. Public transport-related research that has adopted this solution approach is Osorio and Bierlaire (2013). In the work, a meta-model is used to integrate information from a simulator with an analytical queueing network model. The resulting SBO framework is computationally efficient and can be applied to complex problems with tight computational budget constraints. The method was also used to evaluate the performance of a traffic signal control problem for the Swiss city of Lausanne, under different demand scenarios. SBO solutions involving metaheuristics have an advantage over the previously stated methods. This advantage is their ability to find good solutions even when the search space is high-dimensional and not continuous or when qualitative decision variables are involved Arisha and Abo-Hamad (2010). A typical application of this approach to public transport systems is Song et al. (2013), who proposed a SBO method for evaluating and optimising sustainable transportation systems. Four major parts of their model are the strategy, simulation, evaluation, and optimisation. The tools used to implement the model were the traffic simulation software PTV-Visum in combination with a Genetic Algorithm (GA). The model has been used to study a small multimodal network in China to show its feasibility. As previously highlighted in the introductory section, the NSGA-II and MATSim are key components of the model developed in this paper. Both of these are elaborated on next.

2.2.1. NSGA-II

The Non-dominated sorting genetic algorithm-II is a multi-objective genetic algorithm. It was developed by Deb et al. (2000) due to inherent limitations in its predecessor the NSGA Knowles et al. (2008). In the literature the algorithm is classified as *bio-inspired* since its operations mimic the principle of natural genetics. It works by enabling the realisation of newer and presumably better generations of solutions from existing ones. The algorithm's framework consists of a population of solutions or chromosomes. Each chromosome is made up of genes that depend on the particular representation of the chromosome. To adapt the NSGA-II for public transport network design, an initial population of transit networks is selected as a first set of feasible solutions or chromosomes. The networks generally have different configurations and other operational parameters. The main task is, therefore, to find a network and its parameters among the alternatives, which best address the stated optimisation goals.

To search for a network solution, the algorithm combines both traditional single objective genetic operators, like crossover and mutation, with other multi-objective ones such as non-dominated sorting (NS) and crowding distance (CD). Crossover and mutation are known as genetic operators, since their action on the current population of network solutions give rise to offspring, which are generally assumed to be fitter or perform better than their *progenitors*. They achieve this by altering the stops and route configuration of randomly chosen networks within the population. NS and CD on the other hand play the role of advancing the search for a better network solution by providing a mechanism to assign fitness scores to the networks that are being evaluated and ranking them based on their fitness. This ultimately makes it possible for the algorithm to identify better solutions that will make up the next generation. At the end of the optimisation process, the best performing chromosome or network in the population represents a globally optimum solution. However, it should be pointed out that for very difficult problems like the TNDP, it is not feasible to know if a solution is the *global optimum*. This is especially true in a multi-objective context where one seeks a set of optimal solutions also known as a Pareto optimal front rather than a single solution. Therefore, an efficient, locally optimal front that is obtained within a reasonable time frame is generally considered acceptable.

To solve a TNDP with the NSGA-II, the chromosomes or networks needs to be encoded in a way that is amenable to the algorithm's operators. String and binary representation are the most common representations used when solving the TNDP. In Buba and Lee (2018), a string is used to represent the network route, while a tuple is used to represent the route's operational frequency as the number of vehicles operated per hour and the unique identifier for that route. However, in this paper, a JavaScript Object Notation (JSON) Crockford (2011) encoding is used. This representation therefore enables genetic operations to be carried out directly on the candidate networks and their detailed operational schedules. This, in turn, allows for the simultaneous handling of the route network design and frequency setting problems.

2.2.2. MATSim

MATSim is an agent-based modelling tool, designed to model large scale transportation scenarios in very fine details. In this paper, MATSim is used to simulate the microscopic activities of people on a public transportation network over a 24-hour duration. Conceptually, the simulation consists of two layers that are characterised as *physical* and *mental* in Nicolai (2013) and Rieser (2010). The *physical* layer is also called a mobility or traffic flow simulation, and it represents the tangible parts of a transit system, like agents or travellers, their activity plans, activity locations, vehicle fleet, network infrastructure and other concrete elements of the system. The physical layer reflects the agents and how they interact on a transit network, while the *mental* layer represents the abstract part of a transit network system. It describes how the agents receive and process information from the network environment; and how this affects their decisions and choices on the network, with the ultimate goal of improving their overall travel experience. For example, if a traveller who has planned to use a specific transit route, now receives information about a sudden closure on that route. This will likely make the traveller choose a different route to their destination. A typical MATSim cycle with its steps is shown in Fig. 4. It comprises of five steps that are discussed briefly below, further details about these steps may be found in Horni et al. (2016).

1. Initial demand generation: The initial demand is generated by creating daily activity plans, from the socio-economic and demographic data of agents within a given transportation area. Statistical sampling and discrete choice modelling are two techniques which can be used to create the plans. A plan comprises the sequential ordering of all activities an agent will engage in within a 24-hour period and the trips that connect the activities. Typically, each MATSim agent stores a fixed number of daily plans in their memory. Furthermore, the mode of travel and other time-based information like activity departure, arrival and duration are defined in the plan. However, the estimation of mode choice is done using utility based functions that are normally defined within the third step of the simulation known as scoring.
2. Execution: Execution involves simulating the generated initial demand. In this step, the travel plans of agents are executed sequentially by their time of occurrence. It is also done in a way that respects certain boundary conditions like the closing hour of a shop or the maximum link and flow capacity of a road. Another name for this step is mobility simulation, or *mobsim*, for short. When the simulation begins, an agent chooses a single plan from its memory, which is then executed during the simulation. The overall simulated effect of the individual agents decisions and actions on a network at a given time, defines the prevailing network condition at that time. It also determines the subsequent choices people will make on the network.
3. Scoring: The score is obtained, by evaluating a plan using a utility function known as a *scoring function*. MATSim uses the scores to measure and compare the quality of a passenger's plan and determine if it should be dropped or not. MATSim uses the scores to measure and compare the quality of a passenger's plan and determine if it should be dropped or not. It describes a traveller's perception of time spent travelling or engaging in an activity with components like; waiting time, travel time or time spent on the activity. The general form of the scoring function, can be seen in Equation (1), while its constituent variable are presented in (2).

$$U_{\text{plan}} = \sum\nolimits_{i=1}^{N} U_{\text{activity},i} + \sum\nolimits_{i=1}^{N-1} U_{\text{travel,mode}(i)} \tag{1}$$

$$U_{\text{plan}} = \sum\nolimits_{i=1}^{N} \left(U_{\text{perf},i} + U_{\text{wait},i} + U_{\text{late},i} + U_{\text{early},i}\right) + \sum\nolimits_{i=1}^{N-1} U_{\text{travel},i} \tag{2}$$

The function is the total utility derived from performing an activity and travelling to and from the activity. The utility derived from engaging in an activity ($U_{\text{activity},i}$) is a sum of utilities derived from performing the activity ($U_{\text{perf},i}$), arriving late ($U_{\text{late},i}$) or early ($U_{\text{early},i}$) at the activity, and waiting to perform the activity ($U_{\text{wait},i}$). It is reckoned that arriving late is considered a negative utility. However, arriving on time at the activity location does not necessarily increase utility. The second component is the negative utility derived from travelling to an activity location.

2.2.3. Performing activities

The utility for performing an activity, which is usually positive, is depicted as:

$$U_{\text{perf},i} = \beta_{\text{perf}} \cdot t_{\text{typ},i} \cdot \ln\left(\frac{t_{\text{perf}}}{t_{0,i}}\right) \tag{3}$$

where:

1. $U_{\text{perf},i}$ = utility of performing an activity.
2. β_{perf} = the marginal utility for performing an activity.
3. $t_{\text{typ},i}$ = the typical duration for activities, say eight hours for work.
4. t_{perf} = the actual time spent on an activity.
5. $t_{0,i}$ = the duration when utility starts to be positive.

2.2.4. Waiting

The penalty for arriving late at an activity, which is usually negative, is then:

$$U_{\text{wait},i} = \beta_{\text{wait}} \cdot t_{\text{wait},i} \tag{4}$$

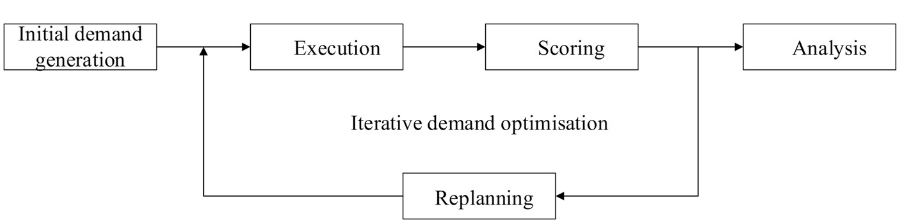

Fig. 4. MATSim simulation process adapted from Horni et al. (2016).

where: .

1. $U_{\text{wait},i}$ = utility associated with waiting to start an activity.
2. β_{wait} = the negative marginal utility of time spent waiting.
3. $t_{\text{wait},i}$ = the time spent waiting such as for the transit vehicle to arrive.

2.2.5. Arriving late at the activity

The penalty for arriving late at an activity, which is usually negative, is:

$$U_{\text{late.arr},i} = \begin{Bmatrix} \beta_{\text{late.arr}} \cdot (t_{\text{start},i} - t_{\text{latest.arr},i}) & \text{if } t_{\text{start},i} > t_{\text{latest.arr},i} \\ 0 & \text{else} \end{Bmatrix} \tag{5}$$

where: .

1. $U_{\text{late.arr},i}$ = penalty for arriving late.
2. $\beta_{\text{late.arr}}$ = marginal utility of arriving late.
3. $t_{\text{start},i}$ = start time for the activity.
4. $t_{\text{latest.arr},i}$ = latest possible penalty-free activity starting time.

2.2.6. Leaving early from the activity

The reward for arriving late at an activity, which is usually negative, is depicted as:

$$U_{\text{earliest.dep},i} = \begin{Bmatrix} \beta_{\text{earliest.dep}} \cdot (t_{\text{end},i} - t_{\text{earliest.dep},i}) & \text{if } t_{\text{end},i} > t_{\text{earliest.dep},i} \\ 0 & \text{else} \end{Bmatrix} \tag{6}$$

where: .

1. $U_{\text{earliest.dep},i}$ = penalty for leaving early.
2. $\beta_{\text{earliest.dep}}$ = marginal utility of earliest departure.
3. $t_{\text{end},i}$ = end time for the activity.
4. $t_{\text{earliest.dep},i}$ = earliest possible activity end time.

2.2.7. Travelling

The disutility or a penalty for travelling is seen in (7).

$$U_{\text{trav}} = \beta_{\text{trav,mode}} \cdot t_{trav} + \beta_m \cdot m_{\text{trav}} + (\beta_{\text{dist,mode}} + \beta_m + \gamma_{\text{dist,mode}}) \cdot d_{\text{trav}} + \beta_{\text{transfer}} \tag{7}$$

where: .

1. U_{trav} is the penalty for travelling
2. $\beta_{\text{trav,mode}}$ is the marginal utility of travelling by a given mode
3. t_{trav} is the time spent travelling
4. β_m is the marginal utility of money
5. m_{trav} is the change in monetary value of a fare caused by the travel
6. $\beta_{\text{dist,mode}}$ the marginal utility of the distance travelled by a given mode
7. $\gamma_{\text{dist,mode}}$ cost per unit distance travelled
8. d_{trav} is the distance of the leg
9. β_{transfer} is a transfer penalty e.g. incurred in using transfer points on transit systems.

Equation (2) reflects the mode choice of travellers and its is assumed in MATSim that the mode corresponding to their best calculated plan reflects their mode of choice Grether et al. (2009).

4. Replanning: When agents adapt their plans, in response to changes in the transit network, it is known as *replanning*. This is the main *innovative* component of MATSim, as it allows the agent to modify their plans as they *learn* about prevailing network conditions. This enables the agent to maximise their experience on the public transport network and it is linked with travel behavioural changes on the network.
5. Termination: MATSim usually specifies a termination criterion such as number of iterations, to enable the simulation stop once the condition is met. Ideally, the simulation should terminate after it reaches equilibrium. Meister et al. (2010) describes the termination point as an agent-based stochastic user equilibrium. At the end of the simulation, post-analysis which involves collecting and aggregating network performance indicators is done to gain insight into the travel demand and simulated behaviour of agents within the study area.

In terms of how it models public transport, MATSim organises public transit system data in terms of stops, routes and lines which is a format commonly used by public transit services worldwide Horni et al. (2016). In the case of the MyCiTi BRT network a line modelled in MATSim will comprise two or more transit routes. The route itself is a sequence of road links that facilitate the MyCiTi buses to run on the route. Each route serves one direction of travel and enables buses to move to and from the depot at the end and beginning of a day, respectively. The routes also have as an attribute the list of `departures`, which gives information about the time a vehicle starts at the first stop on that route. Furthermore, a route includes a sequential list of transit stops that are served, alongside operating timetables, which indicate when vehicles arrive or leave a stop. The times are specified as offsets in time units from the departure at the first stop. At each subsequent stop, the offset is added to the initial departure time at the first stop. Each departure contains a vehicle's start time on the route and a reference to the vehicle. As the timing information is part of the route, it becomes possible to have routes with identical stop sequences but different time offsets. Stop locations are described by their coordinates and an optional `name` or `id`. They must be assigned to unique lines of the network for the simulation.

2.3. Data requirements

The supply side data used in this paper include transit network, schedules and vehicles. These are obtained from general transit feed specification (GTFS) feed for the MyCiTi service. On the other hand, the agent's *plan* files representing travel demand are extracted from the MyCiTi AFC dataset. The AFC system reads the passenger's smart card to determine that they have sufficient credit for their trip and by so doing a record of valid trips made by the passenger may be stored. The smart card unique number enables people's movement to be tracked anonymously within the system, without violating their privacy or obtaining personal information. This facilitates the automated collection of passenger flow and network utilisation information; such as unique passenger transaction number, boarding and alighting locations, bus route interactions with the route name, along with the date and time of transactions. The interactions are typically in the form of bus transactions—*boarding, alighting* and *transfers* at stops. Before using the data, the following steps are taken to pre-processing it are listed below. 1) Import and clean data by deleting incorrectly recorded data points. 2) Split the data into daily trips since the AFC is stored in monthly batches. 3) Make trip chains. 4) Create the plan data.

3. Models

The objective of the mathematical model presented in this section represents a cost minimization for both users and operators of the MyCiTi BRT network. This translates to reducing travel time for passengers and operating cost for operators. This has the effect of increasing the perceived attractiveness for commuters which can lead to increased ridership and access while the operators who are now able to reduce inefficiencies like operating redundant routes, can improve system utilization, service delivery and network coverage. The network is represented as a graph $\mathbf{G} = (\mathbf{N}, \mathbf{E})$, which is a multiple connection of a finite sets of $n \in \mathbf{N}$ nodes and $l \in \mathbf{E}$ links. The objective functions in (8), represent the costs that accrue to two major transit network stakeholders; users as in (9) and operators as seen in (10). Transit users generally view *generalised cost* in terms of their total travel time (access, waiting, and in-vehicle travel times plus transfers where applicable). On the other hand, operators are concerned with the total operational cost comprising of the total distance and time operated. Operational distance is the cost that accrues from the wear and tears on the operators' vehicles as they traverse the designated routes to satisfy passenger demand. It is typically measured in kilometres. However, the operational time consists of personnel cost element like salaries that accrue throughout operations. Therefore, by minimising these objective functions, the total cost incurred on the network will be optimised for the earlier mentioned stakeholders.

$$Min: Z_1, Z_2 \tag{8}$$

$$Z_1 = \beta_{\text{time}}{}^*\Big(\sum_{r\in R}\sum_{r_i\in R_m} t^{r_i}{}_r q^{r_i}{}_r + \sum_{r\in R}\sum_{r_i\in R_m} t^{r_i}{}_a q^{r_i}{}_a + \sum_{r\in R}\sum_{r_i\in R_m} t^{r_i}{}_w q^{r_i}{}_w\Big) + \psi_{\text{time}}\cdot \sum_{n\in N} n_t \tag{9}$$

$$Z_2 = \beta_{\text{dist}}{}^* \sum_{r\in R}\sum_{r_i\in R_m} l^{r_i}{}_{tr} f^{r_i}{}_{tr} + \beta_{\text{op}}{}^* \sum_{r\in R}\sum_{r_i\in R_m} t^{r_i}{}_{tr} f^{r_i}{}_{tr} \tag{10}$$

subject to agent-based stochastic user equilibrium on the network Horni et al. (2016), which describes the individual traveller's behaviour on a public transportation network, and represented by (11). Modelling travel behaviour in this manner extends the conventional stochastic user equilibrium. This is because individual traveller's demand and behaviour is modelled rather than aggregating route-based passenger volumes as aggregated productions and attractions, as is the case in conventional trip-based models. Also, the route and mode choices used in the traditional user equilibrium is broadened to include other dimensions such as destination choice. Lastly, passenger demand is loaded onto the network, with stochastic algorithms that use time-dependent trip departure times.

$$q_r^n = \tau(c(x\{q_{r_i}^n\})) \tag{11}$$

and some feasibility conditions on route length, frequency and vehicle fleet:

$$l_{min} \le l_{r_i} \le l_{max} \tag{12}$$

$$f_{min} \le f_{r_i} \le f_{max} \tag{13}$$

$$r_{tot} \le R_{max} \tag{14}$$

Where:

1. $\mathbf{N}$ = set of nodes on the network (-)
2. $\mathbf{R}$ = set of transit routes (-)
3. R_m = set of segments r_i that serves demand on route r (-)
4. r = route on the network (-);
5. r_i = segment r_i that serves demand on route r (-);
6. r = route on the network (-);
7. Z = objective function (-);
8. z_1 = user cost objective function (-);
9. β_{time} = monetary unit value for user travel time ('000);
10. ${}^{r_i}t_{tr}$ = travel time on route segment r_i (hr);
11. $q_{tr}^{r_i}$ = travel demand on route route segment r_i (pax);
12. $t_a^{r_i}$ = access time on route segment r_i (hr);
13. $q_a^{r_i}$ = passengers boarding on route segment r_1 (pax);
14. $t_w^{r_i}$ = waiting time on route segment r_i (hr);
15. $q_w^{r_i}$ = passengers waiting on route segment r_i (pax);
16. ψ_{time} = time penalty associated with transfers (-);
17. n_t = transfers on r (-);
18. z_2 = operator cost objective function ('000);
19. β_{dist} = monetary unit value for vehicle mileage ('000);
20. $l_{tr}^{r_i}$ = length of route segment r_i (km);
21. $f_{tr}^{r_i}$ = frequency on route segment r_i (veh/hr);
22. β_{op} = monetary unit value for vehicle operating time ('000);
23. $q_{r_i}^n$ = individual agent demand on the route segment r_i (pax);
24. n = index of the agent (-);
25. τ = agent-based probabilistic route choice model (-);
26. $c(x)$ = network costs (-);
27. $\{q_{r_i}^n\}$ = set of all individual agent route demands on the network (-);
28. l_{min} = minimum route length (-);
29. l_{max} = maximum route length (km);
30. f_{min} = minimum frequency value (veh/hr);
31. f_{max} = maximum frequency value (veh/hr);
32. r_{tot} = number of designed routes (-);
33. R_{max} = maximum number of routes that are allowed on the network (-);

The feasibility constraints for the model are those on route length, frequency and the vehicle fleet size seen in (12) to (14). These constraints are used to set the allowed limiting conditions for the allocation of resources on the transit network. Equation (12), which is a route length constraint, is introduced to define the upper and lower bounds outside which it would be illogical to operate a bus service. Usually, public transit operators will not run a service on routes that users may conveniently traverse by walking. They also avoid developing excessively long routes Cipriani et al. (2012). Such routes make schedule adherence difficult or may result in the need to provide too many transfers, which users find unattractive in transit services. Equation (13) is a feasibility constraint on transit service frequency. The constraint represents the maximum and minimum operable frequency on each transit route within the bus network. It depends, typically, on the available fleet size and transit demand for each route. Lastly, (14) puts a constraint on the maximum number of routes or network size. This is generally determined by transit authorities who stipulate the number of routes they can provide. In practice, this constraint depends on the available financial resources, which the authorities can invest in operating the network.

3.1. Model assumptions

1. A complete trip or satisfied demand may be in two forms: *boarding–alighting* (B–A) or *boarding–connection–alighting* (B–C–A). The former is a direct trip without transfer, while the latter is a trip satisfied with one transfer required. This specification aligns with how demand coverage is defined in this article: demand that is satisfied with zero or one transfer. It is assumed that commuters generally find a trip less attractive beyond one transfer and that this would lead them to search for alternative, more direct routes or even in some cases to change their mode of travel Owais (2015).
2. In agent-based travel demand models, demand is generated from people's activities at different locations based on various land uses, however, in this work it was not possible for to obtain information concerning activities or activity locations outside the transit network. Consequently, activities refer strictly to transactions like passenger boarding, alighting transfers and others that occur on the network.
3. In this work automated fare collection data is used to create the daily trip chains of the commuters which is subsequently converted to the initial demand used in the MATSim simulation.
4. Vehicle and crew scheduling is not performed as it is outside the scope of this work.
5. Latent travel demand is assumed given the influence of other modes of public transports on the IPTN and BRT.

4. Solution procedure

The model used in this paper is named a Simulation-based transit network design model (SBTNDM). Two main components of the

SBTNDM are Agent-based travel demand simulation (MATSim) and multi-objective public transit network optimisation(NSGA-II). The most important interaction within the model, is in translating the NSGA-II's network solutions to a format that is readable by MATSim and vice-versa. This will ensure that the network solutions are adequately evaluated and the results can be used in subsequent stages of the optimisation. Three main steps are involved in network design with the SBTNDM are, network generation, network evaluation and lastly, a procedure used to search for an optimised network. They are described next.

1. Network generation: In this stage of the model, computer based heuristic algorithms are used to create an initial population of feasible solutions for the network design problem. Feasibility criteria like network size and route length guide the process. The input for this stage includes: nodes of an existing transit network, minimum and maximum route length; and the number of routes per network. The network generation heuristic reads in the MyCiTi stops data, it then gets the shortest paths between all origin destination pairs in the data set with a *k*-shortest path algorithm Yen (1971). Each path is checked for the route length feasibility condition. After the checks, networks are created by selecting 46 routes at random. This number was used, to match the network size of the MyCiTi network. Through this process a pool of feasible networks were generated. From this pool, the initial population of networks used in the design will be initialised.
2. Network analysis: This step involves evaluating the generated network solutions against the objective functions. This is achieved by initially simulating travel demand on the networks using MATSim to obtain user behaviour parameters. Thereafter, the result of network simulation are used as input in the objective functions to calculate the objective scores and other indicators for each network. The obtained score is assigned to the current network solution, which is then returned to the optimisation module for further processing. One way to account for the randomness associated with transport decision making and service planning is to simulate the process multiple times and utilise the mean realisation of the different simulation runs. This is highlighted in cases where the same service plan incurs different costs due to the stochastic decision making of commuters and their impact on the operating environment or poor decision making in operational planning. Similarly, in this paper, multiple instances of MATSim is run in each evaluation of the candidate network solutions. A MATSim scenario comprising of multiple runs of the simulation is set up and configured to run in parallel. Internally, each *simulation* comprises a user specified number of MATSim iterations running sequentially. For this work, it was determined experimentally that each simulation converges after 80 iterations. However, since multiple simulations are required, they are set up to run in parallel. An image depicting the structure of the MATSim simulation utilised in this work can be seen in Fig. 5.
3. Search for optimised solution: In this final stage, the integration between simulation and optimisation occurs. The result of the MATSim simulation is used by the NSGA-II to rank and compare the solutions through the action of its genetic operators. This process continues iteratively until a predefined termination criterion is satisfied. The key inputs used here are the outputs of the network generation and analysis, respectively, namely the feasible candidate solutions and objective function scores from simulating each solution with MATSim. This implies that at different stages of its operation, the procedure will call generation and analysis sub-routines. The flow chart for the SBTNDM algorithm may be seen in Fig. 6.

5. Results

5.1. Numerical results

The result of the network design exercise is a Pareto optimal set consisting 10 different networks representing trade-off solution for the problem. The obtained network results are analysed and compared with the status-quo MyCiTi network, which is essentially the existing network with its operating parameters. The analysis is performed by evaluating each network solution against the objective function previously used in the mathematical model of the problem. The resulting indicators representing the total cost incurred by commuters in terms of travel time and operators alike in terms of operational costs, then serves as the basis of comparison The results are presented as tables below. Table 1 shows the raw values of the indicators like the user and operator objective costs. In Table 2 differences between indicator values for the existing network and those designed with the SBTNDM are shown as percentages. In the table, all the designed networks show improvement over the existing MyCiTi network. Furthermore, network 1 has the largest enhancement of 26.42% for user cost and the least operator cost reduction of 6.29%. On the other hand, network 10 has the best operator's cost reduction of 13.06% and the least user cost reduction of 10.34%. In essence, the network with the highest travel time also has the least operator cost and vice versa. It may also be noted that network 5 strikes a balance between the previously discussed user and operator inclined solutions, hence it will be considered the trade-off solution in the Pareto set. This is important to note because in transport service provision both extreme solutions only satisfy one key stakeholder, leaving the other with very high costs. Hence, finding a solution that strikes a balance between both commuters and operators is the imperative of the service provider. These results show that the *trade-offs* between both the transit users' and operators' perspectives are captured within the set of network solutions. The three network discussed earlier are further analysed to obtain transport related indicators. These results can be seen in Table 3. It may be observed that the designed networks show a clear improvement on the existing network on all indicators. The results are also reflective of the trade-offs previously alluded to. For indicators relating to commuters like satisfied demand, the Solution 1

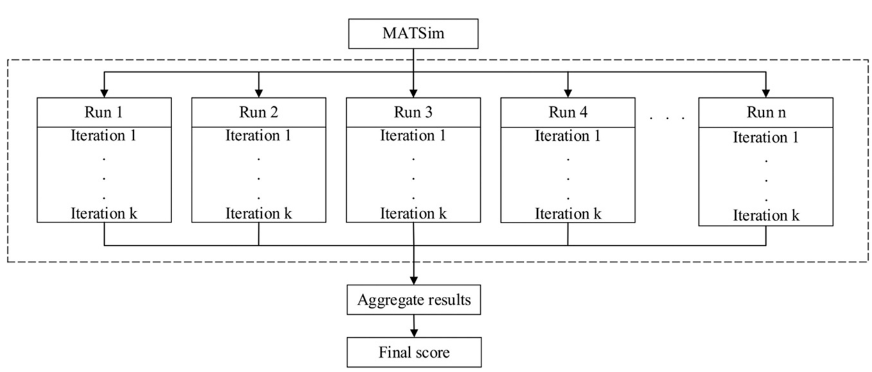

Fig. 5. The parallel implementation of MATSim used in the SBTNDM.

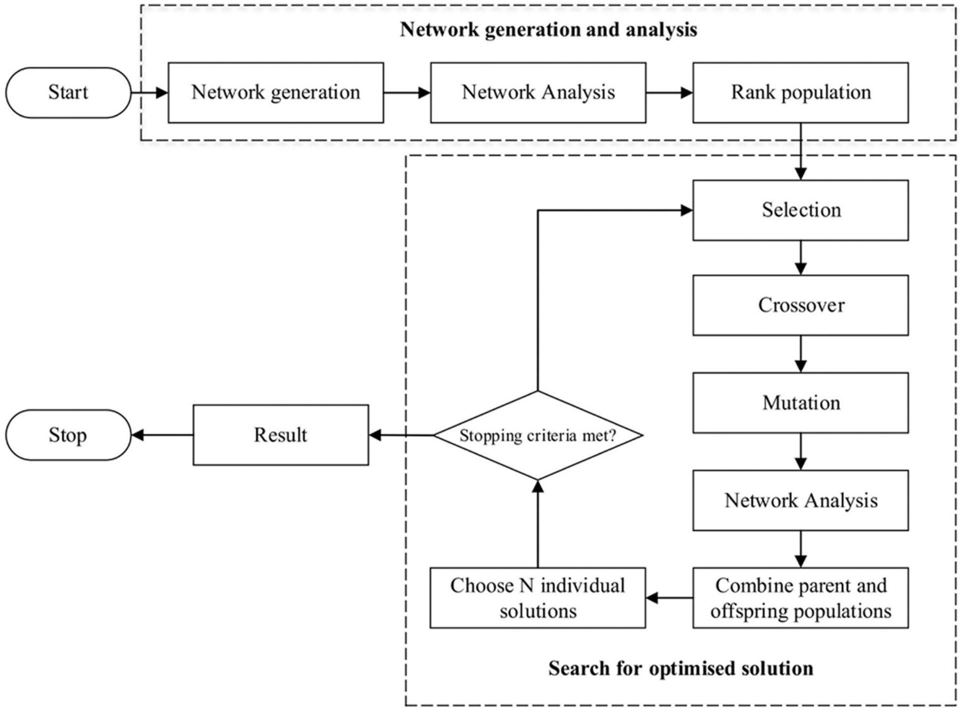

Fig. 6. The simulation based transit network design model.

Table 1
Raw indicator values for existing and designed networks.

Network	User cost (hr)	Operator cost ('000)
Base	777.06	29,073.76
1	571.77	27,244.27
2	578.44	26,735.19
3	589.05	26,374.50
4	598.69	26,119.96
5	616.18	25,759.27
6	629.60	25,653.11
7	646.70	25,504.73
8	662.42	25,398.57
9	670.25	25,292.42
10	696.71	25,276.98

Table 2
Difference between existing and designed networks.

Network	Difference in User cost (%)	Difference in operator cost (%)
Base	–	–
1	26.42%	6.29%
2	25.56%	8.04%
3	24.20%	9.28%
4	22.95%	10.16%
5	20.70%	11.40%
6	18.98%	11.77%
7	16.78%	12.28%
8	14.75%	12.64%
9	13.75%	13.01%
10	10.34%	13.06%

outperforms the others, though this comes at a higher cost of operational parameters like Vehicle time. In terms of the operator perspective, Solution 10 does better with a lower value of operator cost. Lastly, it can be observed that the solution 5 strikes a balance between solutions 1 and 10.

Fig. 7 is a map showing the CBD and the immediate areas north of the CBD such as Milnerton. In the map the base network is coloured red while routes of the designed network are highlighted in green. Since both networks overlap, the visible green routes are those supplied by the designed network solution which are not available in the base network. This would allow for coverage of demand on parts of the network that were not previously covered, hence, passengers whose demand were unmet in the existing network can now be satisfied. The above discussion reflects the effect of latent demand on the network design process. In public transport planning two common sources of latent demand are from people without prior access to transport services and those who are attracted due to service improvements. Both cases are applicable in Cape Town as there remains a large section of the society, especially those living on the city's periphery that have limited to no access to transport services. On the other hand, there are passengers who will change their mode of travel if they found a better alternative.

Overall, these results show that the SBTNDM is capable of designing networks that reduce operator cost on the MyCiTi, while boosting passenger ridership. From a decision support and policy making standpoint, the SBTNDM can serve as an important tool in the public transport planning process in Cape Town. This is because the model can identify many suitable transit network alternatives and reveal the trade-offs associated with each network option. A key feature of this work is its adoption of technology-driven tools like big data in the form of AFC data and agent-based simulations for addressing some of the challenges associated with planning, designing and operating the MyCiTi BRT system. The shift to technology based solutions for Cape Town's transport is crucial because of the potential technology has to improve the operational efficiency. For instance the trip making data obtained on the MyCiTi AFC system offers the service operator better insights into the behaviour of its users, which in turn makes it possible to upgrade their services in ways that increase the attractiveness of the system. Therefore, the fact that this work utilises some of the earlier mentioned tools to develop network solutions that reflect the changing landscape of travel behaviour in the city makes the proposed SBTNDM even more relevant. In light of this, the authors recommend that the SBTNDM be adopted as a public transport network planning and design tool for the MyCiTi BRT in Cape Town.

Table 3
Aggregate transit network performance indicators for the identified scenarios.

Indicators	Base network	Solution 1 (User)	Solution 5 (Balanced)	Solution 10 (Operator)
Satisfied demand (pax)	37,392	51,324	47,541	44,385
Utilisation (%)	62.69	86.05	79.70	74.42
Objective function value (km)	52,619.35	48,567.20	45,215.15	42,452.99
Veh. time (hr)	2468.96	1942.69	1808.60	1618.12
Op. cost ('000)	33,919.40	34,674.53	32,859.63	32,170.70

Fig. 7. An overlay of the base and designed networks.

6. Conclusion

The current operational environment for the MyCiTi BRT network in Cape Town, South Africa, reveal inefficiencies that have resulted in low passenger ridership among other challenges. The status quo has resulted in low passenger ridership in public transport and a high dependence on private cars which has worsened congestion and environmental pollution. Furthermore, due to the difficulties involved in balancing operational service levels with cost, many public transport systems in South Africa suffer from low productivity, high costs, and a need for large government subsidies. To alleviate this situation, a simulation-based network design model was used to improve the network. The results from testing and applying the so-called SBTNDM shows that the model is capable of designing improved network solutions that reflect the objectives of network stakeholders like passengers and operators. Possible research directions that may extend from this work in the future, is to widen the application of the SBTNDM to a multi-modal network context, to improve modal integration in Cape Town.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the Centre for High Performance Computing (CHPC), South Africa, and the University of Cape Town's ICTS High Performance Computing team for providing computational resources for this research project.

References

Alrabghi, A., Tiwari, A., 2015. State of the art in simulation-based optimisation for maintenance systems. Comput. Ind. Eng. 82, 167–182.
Amaran, S., Sahinidis, N.V., Sharda, B., Bury, S.J., 2016. Simulation optimization: a review of algorithms and applications. Ann. Oper. Res. 240 (1), 351–380.
Arisha, A., Abo-Hamad, W., 2010. Optimisation methods in supply chain applications: a review. Ir. J. Manag. 95–124.

Brands, T., Berkum, E.C.V., 2014. Performance of a genetic algorithm for solving the multi-objective. Multimodal Transp. Netw. Des. Probl. 2 (1), 1–20.
Buba, A.T., Lee, L.S., 2018. A differential evolution for simultaneous transit network design and frequency setting problem. Expert Syst. Appl. 106, 277–289.
Ceder, A., 2007. Public Transit Planning and Operation: Theory, Modelling and Practice, first ed. Elsevier, Burlington, MA.
Ceder, A., 2015. Public Transit Planning and Operation: Modeling, Practice and Behavior, second ed. CRC Press.
Chen, J., Wang, S., Liu, Z., Wang, W., 2017. Design of suburban bus route for airport access. Transp. A: Transp. Sci. 13 (6), 568–589.
Chien, S., Yang, Z., Hou, E., 2001. Genetic algorithm approach for transit route planning and design. J. Transp. Eng. 127 (3), 200–207.
Cipriani, E., Gori, S., Petrelli, M., 2012. Transit network design: a procedure and an application to a large urban area. Transp. Res. Part C: Emerg. Technol. 20 (1), 3–14.
Crockford, D. Introducing JSON; 2011. URL https://www.json.org/json-en.html.
Deb, K., Agrawal, S., Pratap, A., Meyarivan, T., 2000. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In: Schoenauer, M., Deb, K., Rudolph, G., Yao, X., Lutton, E., Julian, J. (Eds.), Parallel Problem Solving from Nature PPSN VI. PPSN 2000. Lecture Notes in Computer Science, vol 1917. Springer-Verlag, Berlin, pp. 849–858.
Fan, W., Machemehl, R.B., 2004. Optimal Transit Route Network Design Problem: Algorithms, Implementations, and Numerical Results. Southwest Region University Transportation Center Texas Transportation Institute Texas A\&M University System College, Texas.
Gosavi, A., 2015. Simulation-based Optimization: Parametric Optimization Techniques and Reinforcement Learning, second ed. Springer, New York.
Grether, D.Y., Chen, M., Rieser, M., Nagel, K. Effects of a Simple Mode Choice Model in a Integrating a simple model for mode choice into a multi-agent simulation of travel behavior and traffic flow. Complexity and Spatial Networks In Search of Simplicity, Advances in Spatial Science.2009; (March 2014): 167–186.
Hachicha, W., Ammeri, A., Masmoudi, F., Chachoub, H. A comprehensive literature classification of simulation optimisation methods.In: International Conference on Multiple Objective Programming and Goal Programming - MOPGP10. 27652; 2010.13.
HeykenSoares, P., Mumford, C.L., Amponsah, K., Mao, Y., 2019. An adaptive scaled network for public transport route optimisation. Public Transp. 11 (2), 379–412.
Horni, A., Nagel, K., Axhausen, K.W. The Multi-Agent Transport Simulation MATSim. The Multi-Agent Transport Simulation MATSim.2016: 662.
Huang, D., Liu, Z., Fu, X., Blythe, P.T., 2018. Multimodal transit network design in a hub-and-spoke network framework. Transp. A: Transp. Sci. 14 (8), 706–735.
Ibarra-Rojas, O.J., Delgado, F., Giesen, R., Muñoz, J.C., 2015. Planning, operation, and control of bus transport systems: A literature review. Transp. Res. Part B: Methodol. 77 (3), 38–75.
Knowles, J., David, C., Kalyanmoy, D., 2008. Multiobjective Problem Solving from Nature, first ed. Springer-Verlag Berlin Heidelberg, New York.
Madadi, B., Nes, R., Snelder, M., Arem, B., 2019. A bi-level model to optimize road networks for a mixture of manual and automated driving: an evolutionary local search algorithm. Comput. -Aided Civ. Infrastruct. Eng. 1–17.
Meister, K., Balmer, M., Ciari, F., Horni, A., Rieser, M., Waraich, Ra, et al., Large-Scale Agent-Based Travel Demand Optimization Applied to Switzerland, Including Mode Choice. 12th World Conference on Transportation Research, 11 - 15 July 2010, Lisbon, Portugal.2010; 1: 1–30.
Nicolai, T.W. MATSim for UrbanSim: Integrating an urban simulation model with a travel model. PhD; 2013.
Nnene, O.A., Zuidgeest, M.H.P., Beukes, E.A., 2017. Application of metaheuristic algorithms to the improvement of the MyCiTi BRT network in Cape Town. J. South Afr. Inst. Civ. Eng. 59 (4), 56–63.
Nnene, O.A., Joubert, J.W., Zuidgeest, M.H.P., 2019a. An agent-based evaluation of transit network design. Procedia Comput. Sci. 151, 757–762 (jan).
Nnene, O.A., Joubert, J.W., Zuidgeest, M.H.P., 2019b. Transit network design with metaheuristic algorithms and agent based simulation. IFAC-Pap. 52 (3), 13–18.
Osorio, C., Bierlaire, M., 2013. A simulation-based optimization framework for urban transportation problems. Oper. Res. 61 (6), 1333–1345.
Ouyang, Y., Nourbakhsh, S.M., Cassidy, M.J., 2014. Continuum approximation approach to bus network design under spatially heterogeneous demand. Transp. Res. Part B: Methodol. 68 (18), 333–344.
Owais, M., 2015. Issues related to transit network design problem. Int. J. Comput. Appl. 120 (8), 40–45.
RHDHV.Integrated Public Transport Network Plan 2032. City of Cape Town; 2014.
Rieser, M. Adding transit to an agent-based transportation simulation. PhD; 2010.
Schrijver, A., 2003. Combinatorial Optimization: Polyhedra and Efficiency, first ed. Springer, Berlin, Heidelberg.
Song, M., Yin, M., Chen, X.M., Zhang, L., Li, M., 2013. A simulation-based approach for sustainable transportation systems evaluation and optimization: theory, systematic framework and applications. Procedia - Soc. Behav. Sci. 96 (Cictp), 2274–2286.
TDA.MyCiTi Business Plan 2015 Update; 2015.April.
Viljoen, N.M., Joubert, J.W., 2019. Supply chain micro-communities in urban areas. J. Transp. Geogr. 74 (August), 211–222.
Yen, J.Y., 1971. Finding the K shortest loopless paths in a network. Manag. Sci. 17 (11), 712–716.