PUBLIC TRANSPORT CAPACITY PROVISION AND ITS SENSITIVITY TO DEMAND ESTIMATION

J Reddy and R Behrens*

Hatch Goba, P O Box 25401, Gateway, 4321; Tel: 0315369400;
Email: jreddy@hatch.co.za
* Centre for Transport Studies, University of Cape Town, Private Bag X3,
Rondebosch, 7701; Email: roger.behrens@uct.ac.za

ABSTRACT

The selection of the appropriate transportation technology to satisfy travel demand has far reaching implications in terms of its ability to either effectively support economic growth and prosperity, or be a drain on public resources. In this regard, ensuring a high level of infrastructure productivity is of great importance. While the applicability of various transit technologies ranging from standard bus services to bus rapid transit and rapid rail transit is widely understood, the scalability of individual modes is an aspect that requires greater definition to ensure that a particular mode can have its capacity enhanced over a period of time in suitable increments that can conveniently and efficiently accommodate growth and fluctuations in demand. Rapid implementation of scalable infrastructure solutions is of particular importance in developing countries. This paper investigates: how scalable road-based transit systems are; how demand estimation accuracy effects the provision of transit capacity; and how much spare capacity should be provided. It is concluded that road-based modes are scalable within families of system configurations, and that in low growth scenarios the infrastructure productivity of high capacity modes is poor. It is argued that, while an accurate demand estimate is important from a revenue estimation perspective, demand accuracy from an infrastructure sizing perspective is less important, as the capacity of one component can be influenced by the capacity of others. It is contended that the amount of spare capacity provided at the outset, or at key upgrade points, depends upon the mode capacity configuration, its capacity increase increment size and the forecast rate of growth, not necessarily the ultimate demand.

1. INTRODUCTION

Transport infrastructure is widely considered as central to supporting economic growth and enhancing quality of life (PICC, 2013). While transport infrastructure is an enabler of city development, it is also one of the biggest consumers of public resources and, potentially, a drain if the wrong mode and infrastructure investment choices are made. From this perspective, within the context of rapid urbanisation, the future prosperity of our cities will rely in part on utilising more efficient and sustainable transport technologies such as public transport and non-motorised travel modes. With transport being a significant consumer of public fiscal resources, and a contributor of some 17% of all greenhouse gas emissions (Cervero, 2013), just implementing public transport modes will not be enough. We have to ensure that the public transport technologies employed are productive (i.e. utilise capacity efficiently) and resilient (i.e. capable of surviving changing demand conditions).

Developing countries in particular often face numerous challenges with respect to the provision of transport infrastructure. Rapid delivery of infrastructure is a key requirement to respond to immediate needs. By implication, responding quickly to needs means that planning and implementation timeframes are reduced. In combination with the limited available data to derive accurate travel demand estimates, reduced delivery timeframes potentially means compromising on the detail and accuracy of market research. This is often perceived to be detrimental to the selection of the appropriate mode and to the initial scale and extent of infrastructure implemented. Conversely, spending a significant amount of time on extensive market research can be detrimental to the point that the *status quo* would have changed by the time a project of scale is implemented (Vasconcellos, 2001). Within this context, rapid implementation of scalable infrastructure solutions is of great importance in ensuring productive and resilient systems. Success in achieving productivity and resilience is largely influenced by the sensitivity of capacity provision to demand accuracy.

Traditionally, matching capacity with demand has been viewed as a primary driver for mode selection. However, as technologies have evolved with greater options, and as resources become more constrained, the investment choices have become more complex, especially when coupled with the need for more productive and resilient transport infrastructure assets. In this regard, the aim of the paper is to share guiding principles derived from research that can aid planners, engineers and other key decision-makers with the selection of modes and the provision of capacity at the appropriate level within the context of incremental demand changes and uncertainty with regard to its forecasting.

These principles were formulated by developing better insight into the scalability of various road-based mode capacity configurations, while considering the interdependency of the individual transit system components. For example, the frequency of a bus service influences the sizing of a station platform and not necessarily the pure unconstrained demand typically estimated during planning. To provide this direction, answers were sought to the following questions: (1) How scalable are road-based transit systems?; and (2) How does demand estimation, and its accuracy in estimation, affect the provision of transit capacity? While the above two questions are the focus, a related question is: (3) How much spare capacity should be provided from the start or be provided in increments? The latter question is more difficult to answer because the provision of infrastructure capacity is related to expected demand growth rates and affordability instead of only ultimate demand. In practice this question will likely be addressed by a policy decision, but the outcomes of the research could help inform such decisions.

The paper is divided into four sections. The following section briefly describes the study method (for a detailed explanation of the study method, see Reddy, 2014). Section 3 discusses the results of the study in relation to a number of transit system components. Section 4 concludes by summarising the main conclusions, and reflecting upon some implications the study findings have for public transport system planning and design practice.

2. STUDY METHOD

The study involved the development of a bespoke capacity estimation model named tCAT (*Transit Capacity Analysis Tool*), using Microsoft Excel. A literature review was undertaken to develop an understanding of the factors that affect the provision of transit facilities and to inform the development of the tCAT model. The literature review covered subject matter related to: capacity analysis of transit infrastructure; demand estimation; infrastructure productivity; mode selection; and transportation economics.

The tCAT model estimates the spare capacities for various system components such as station platforms, Right of Ways (ROW), etc. for a range of different road-based transit configurations in relation to incremental changes in demand, while considering the effects of interrelationships between individual system components, such as, the effect of berth provision on station platform sizes. The model was used to compare different transit mode capacity configurations on a synthetic transit corridor (see Figure 2-1) defined as follows:

- The transit system was to be configured to address travel needs within a single linear urban development corridor.
- Being a typical urban development corridor, the equidistant spacing between stops was 800m. This arrangement creates a maximum 400m walk to a stop from any point along the corridor, which represents a five minute walk.
- The spacing of full intersections was set at 800m apart, being characteristic of an Urban Class 2 road according to TRH26 *Road Classification and Access Management Manual* (COTO, 2011).
- The total length of the development corridor was set at 20km, which is close to the average of 23km for 39 Bus Rapid Transit (BRT) systems evaluated as part of the Transit Corporative Research Programme’s BRT case study exercise (TCRP, 2003).
- The assumed density of development limits the need for feeder services with all trips being walk-in trips. This was an appropriate simplification since the focus of analysis is on the main transit mode capacity configuration within the corridor.
- The start and end of the transit route is defined by two terminal facilities which are not dissimilar to on-line halts or stops with the exception of allowing turnaround and longer dwell times for schedule timing purposes.

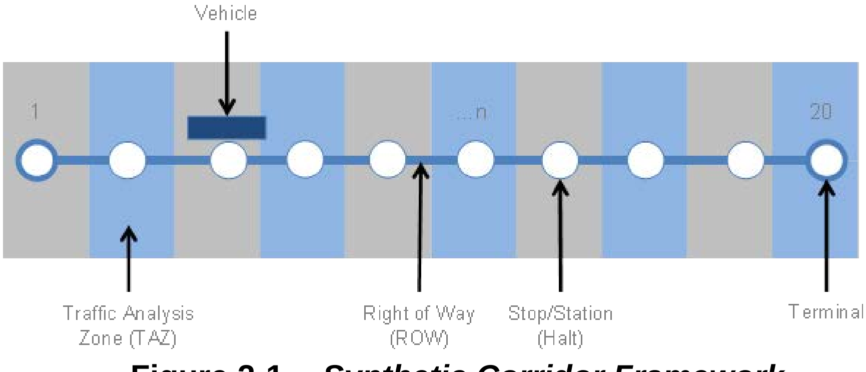


Figure 2-1 ***Synthetic Corridor Framework***

While the use of a synthetic corridor and demand data may bring into question the realism of the results, this approach was considered acceptable for the following reasons:

- The focus of the research was on establishing capacity utilisation across a range of demand conditions.
- The step-wise provision of capacity in response to the step-wise increase in demand defined the incremental provision of capacity with utilisation being measured at each step change.
- Total corridor demand was incrementally increased until failure was reached. The same failure points would be reached regardless of the initial base demand used.

The tCAT model was used to:

- estimate the capacities of a range of road-based transit configurations and solutions;
- evaluate the sensitivity of incremental capacity provision to the step-wise increase in demand; and
- evaluate how the capacity of different system components is utilised.

The overall model structure can be defined according to three main analytical process groups. These process groups include the following:

- demand (i.e. passenger ridership);
- supply (i.e. capacity provision); and
- objective interpretation (i.e. capacity utilisation).

Each process group has its own set of inputs and outputs, but are interrelated as the outputs of one process group are in many instances inputs to another.

The model was configured to allow for the supply process group (which defines station platform sizes, ticket gate numbers, berth provision) to dynamically respond to changes made in the demand process group. Changes in demand result in supply dynamically responding to the need within the confines of the global parameters that define the rules for capacity provision. For example, while demand may drive the need for a large number of berths at a specific location, practical considerations such as distance to the next intersection, accounted for through special parameters, would limit the number of berths that can be provided at that location. The objective interpretation process group is then a reflection of how the capacity provided is utilised.

The calculation engine of the model is based on the equations and parameters derived from vehicle manufacture specifications and the *Transit Capacity and Quality of Services Manual* (Reddy, 2014). The data structures for each process group are illustrated in Figure 2-2. The relationships between the process groups are also shown.

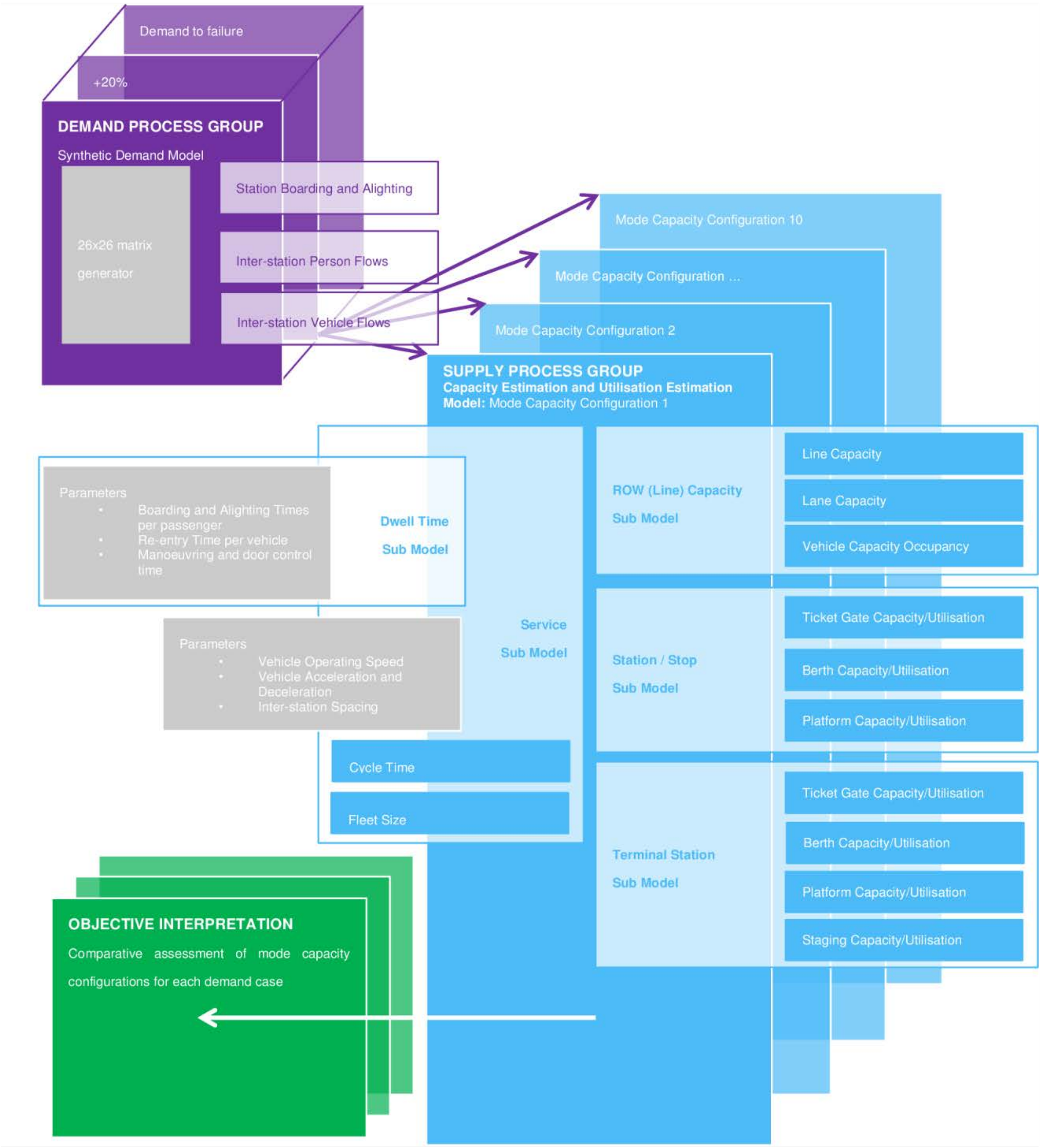

Demand to failure
+20%
DEMAND PROCESS GROUP
Synthetic Demand Model
26x26 matrix
generator
Station Boarding and Alighting
Inter-station Person Flows
Inter-station Vehicle Flows
Mode Capacity Configuration 10
Mode Capacity Configuration …
Mode Capacity Configuration 2
SUPPLY PROCESS GROUP
Capacity Estimation and Utilisation Estimation
Model: Mode Capacity Configuration 1
Parameters
Boarding and Alighting Times per passenger
Re-entry Time per vehicle
Manoeuvring and door control time
Dwell Time
Sub Model
Parameters
Vehicle Operating Speed
Vehicle Acceleration and Deceleration
Inter-station Spacing
Service
Sub Model
Cycle Time
Fleet Size
ROW (Line) Capacity
Sub Model
Line Capacity
Lane Capacity
Vehicle Capacity Occupancy
Station / Stop
Sub Model
Ticket Gate Capacity/Utilisation
Berth Capacity/Utilisation
Platform Capacity/Utilisation
Terminal Station
Sub Model
Ticket Gate Capacity/Utilisation
Berth Capacity/Utilisation
Platform Capacity/Utilisation
Staging Capacity/Utilisation
OBJECTIVE INTERPRETATION
Comparative assessment of mode capacity configurations for each demand case

Figure 2-2 ***tCAT Calculation Engine Structure***

A total of ten mode capacity configuration scenarios were considered based on the combinations of system components that would typically constitute a transit system. These configurations are summarised in Table 2-1. For each configuration, an analysis of the utilisation of individual components was undertaken for a given scenario.

Table 2-1 *Mode Capacity Configurations*

Configuration No.	Description	Fare Collection Method		Right of Way)			Vehicle Type					
		On-board Fare Collection	Pre-boarded Fare Collection (Closed System)	ROW C	ROW B	ROW A	Minibus (16 persons)	Midibus (39 persons)	Bus (106 persons)	Articulated Bus (158 persons)	Articulated Bus (183 persons)	Bi-Articulated Bus (220 persons)
CNG 1	Minibus (Conventional)	x		x			x					
CNG 2	Minibus (Priority)		x		x		x					
CNG 3	Midibus (Conventional)	x		x				x				
CNG 4	Midibus (Priority)		x		x			x				
CNG 5	Bus (Conventional)	x		x					x			
CNG 6	BRT Rigid		x		x				x			
CNG 7	BRT Articulated		x		x					x		
CNG 8	BRT Articulated Capacity		x		x						x	
CNG 9	BRT Bi-Articulated		x		x							x
CNG 10	BRT Bi-Articulated Priority		x			x						x

Note: CNG=configuration

Two types of graphical output were produced:

- utilisation of capacity and productivity potential (which depicts the relationship between utilisation of capacity at different total corridor demands for each mode capacity configuration); and
- comparative assessment of system saturation and 'bottlenecking' (which provides a comparative illustration of total system saturation and when bottlenecking first occurs as total corridor demand increases).

The combined interpretation of these graphics formed the primary basis upon which conclusions were drawn.

2.1. Interpretation of Graphical Outputs: Utilisation of Capacity and Productivity Potential

The output of the utilisation of capacity and productivity potential analysis takes the form of two graphs (see, for example, Figure 3-1). One of these graphs represents average utilisation of a particular system component across the entire transit system, while the other represents maximum utilisation for individual components of the transit system. The latter representation is useful in identifying 'bottlenecking' effects, which tend to occur well in advance of full average system utilisation being reached. 'Bottlenecking' is defined as the first point at which failure occurs or when demand exceeds the capacity provided at a particular location, potentially affecting the utilisation of capacity in other parts of the system. Each line represents the utilisation of the capacity of a system component, of a particular mode capacity configuration. The positioning of the lines relative to each other also indicates the relative performance of each mode capacity configuration.

The oscillations represent the introduction of additional capacity that is greater than the additional demand. Since the provision of capacity is largely modular, there is in most instances the inevitable release of spare, unintentional capacity with the attempt to satisfy incremental demand growth. This results in decreased utilisation of the system component. The greater the *amplitude* of the oscillation, the greater the step change in capacity provided due to the fact that capacity can only be provided in larger capacity modules. On the other hand, the *wave-length* (distance between two consecutive crests) of the oscillations indicates the resilience of the capacity provided to increasing demands. The greater the wave-length, the more resilient the system component is to requiring capacity enhancements as demand increases.

As the provision of additional capacity becomes a practical challenge based on various, physical constraints, the oscillations become less pronounced and eventually disappear once all the maximum capacity that can physically be provided is provided. For example, the availability of kerb-side length would be a constraint to the number of berths that can be provided. This ultimately impacts on the berth capacity of any single stop or station and eventually the system capacity. Once the ‘ultimate’ capacity is provided, the utilisation increases as demand increases. Once in this system state, full capacity utilisation and eventually ‘component failure’ occurs. At this stage, the graphical output illustrates utilisation ‘flat lining’ at 100 percent utilisation at a particular corridor travel demand.

2.2. Interpretation of Graphical Outputs: Comparative Assessment of System Saturation and ‘Bottlenecking’

The graphical output that represents the comparative assessment of system saturation and ‘bottlenecking’ of the various mode capacity configurations takes the form of a single graph illustrating the disjuncture between when bottlenecking first occurs and when the remaining capacity in other parts of the system is eventually utilised (see, for example, Figure 3-2).

The bottleneck point (i.e. the first link in the chain to ’break’) would typically occur before system-wide saturation is reached. For a particular system component such as the station platforms, the graph illustrates at which total corridor demand that particular component would first experience demand in excess of its capacity (i.e. ‘bottlenecking’) at a specific spatial location in the transit system. Total corridor demand could theoretically continue to increase with system saturation eventually occurring when the capacity of the specific transit component(s) is exceeded at all spatial locations. Since in practice, a bottleneck is likely to inhibit the utilisation of adjacent capacity or overall system performance, there is a linkage between the productivity potential of a particular infrastructure asset and the difference between when ‘bottlenecking’ first occurs and the system saturation point. This difference has been termed the Productivity Potential Gap (PPG). The larger the PPG, the greater the spare capacity is in locations apart from where the bottlenecking first occurs.

This spare capacity illustrates potential underutilisation and associated low productivity potential. The converse would also hold true.

3. RESULTS

The results of the comparative analysis described in Section 2 are contained in the Sections 3-1 to 3-5. These sections should be read with reference to Table 2-1 which provides a more detailed definition of the individual mode capacity configurations.

3.1. Vehicle Occupancy

From a vehicle occupancy perspective, modes with lower capacity tend to be scalable over smaller demand ranges, with the ability to provide capacity in smaller increments. Minibus or mode capacity configuration 1 (CGN1) is a good example of this. High capacity systems such as CGN9 are scalable over a broader demand range: however, the capacity increments provided are larger. A representation of capacity utilisation of the mode capacity configurations is given in Figure 3-1 with an associated illustration of system saturation and 'bottlenecking' points given in Figure 3-2.

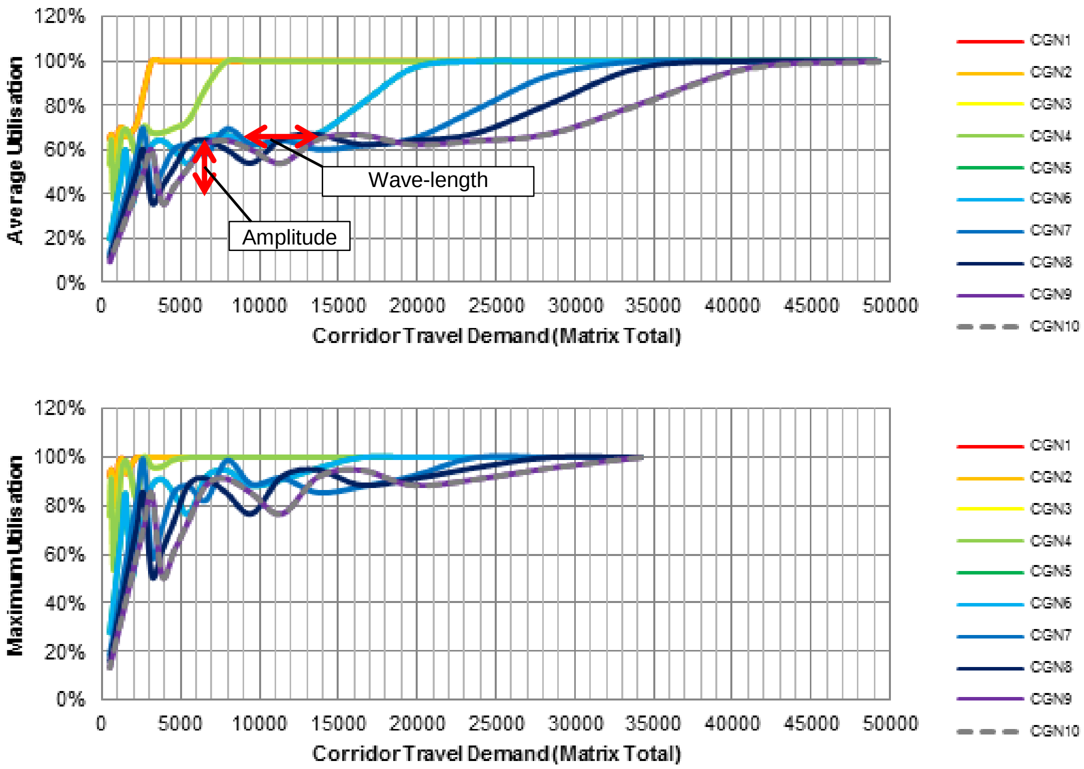


Figure 3-1 ***Utilisation of Capacity and Productivity Potential: Vehicle Occupancy***

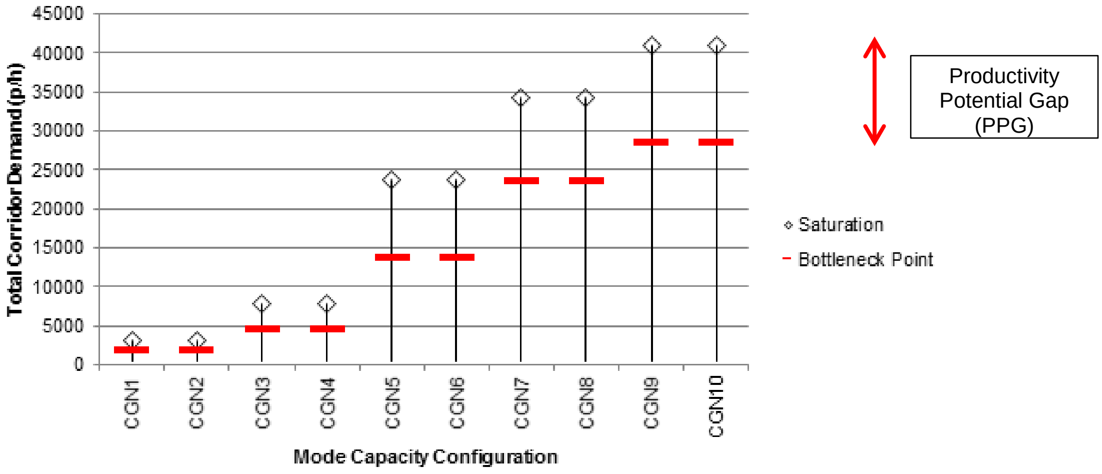


Figure 3-2 *Comparative, Assessment of System Saturation and 'Bottlenecking': Vehicle Capacity*

3.2. ROW Line Capacity

Right-of-Way (ROW) Line Capacity is directly influenced by the frequency of service and the capacity of the vehicles providing the service. As a result, the utilisation of ROW line capacity is also a reflection of vehicle capacity utilisation as shown in Section 3-1. However, for purposes of completeness, a representation of ROW capacity utilisation of the mode capacity configurations is given in Figure 3-3, with an associated illustration of system saturation and 'bottlenecking' points given in Figure 3-4.

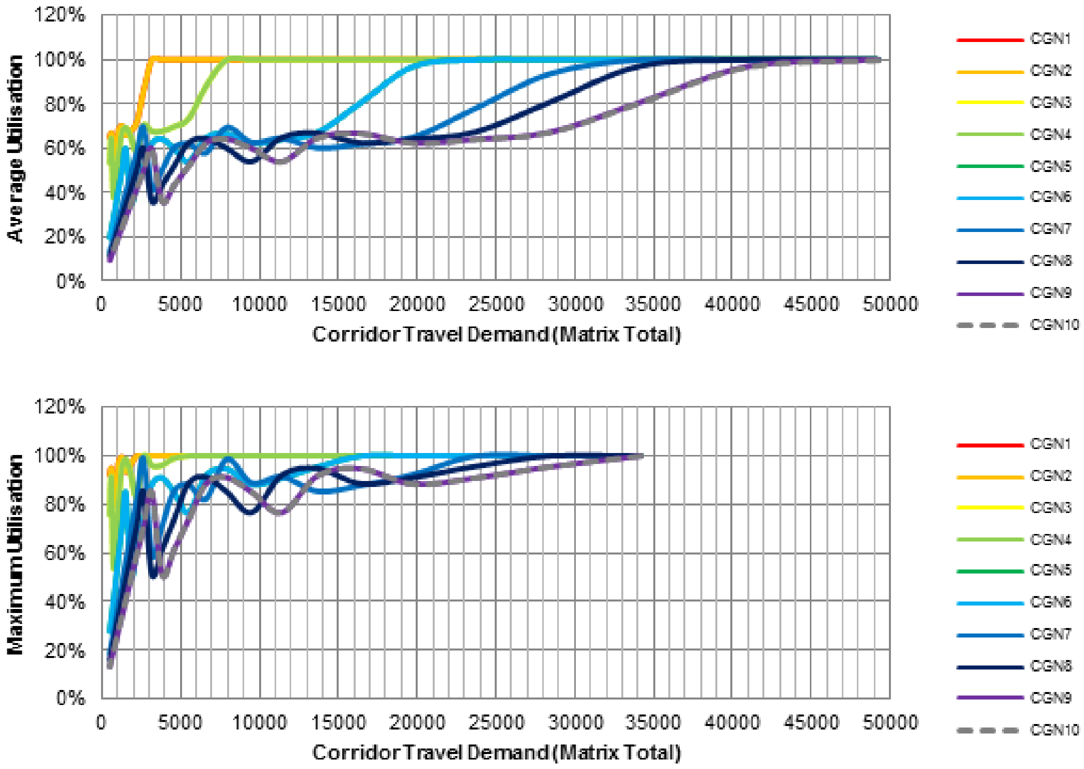


Figure 3-3 *Utilisation of Capacity and Productivity Potential: ROW*

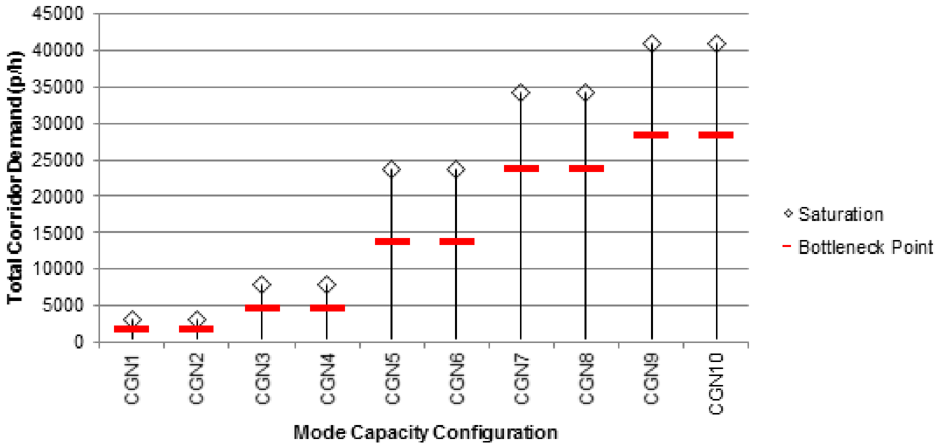
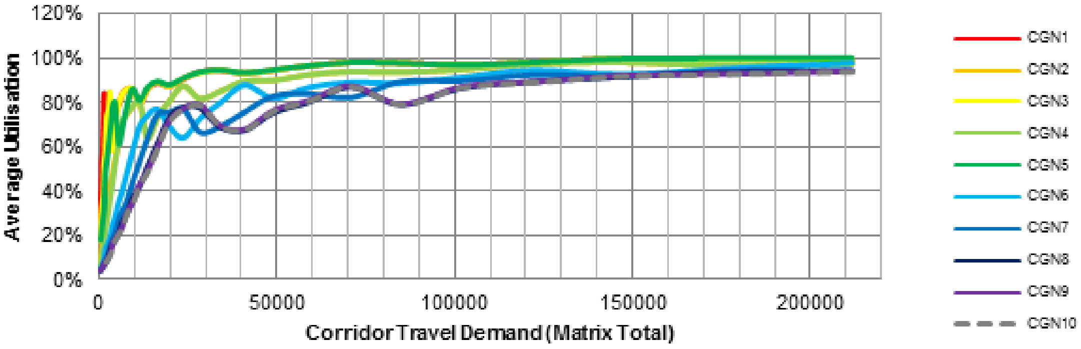


Figure 3-4 ***Comparative, Assessment of System Saturation and 'Bottlenecking': ROW***

3.3. Stations/Stops and Berths

The provision of berth capacity is related to the availability of kerb side space and the dwell time of vehicles stopping at the berth. The corridor was configured to have significant kerb side space given the generous intersection spacing assumed. System wide saturation is not easily reached for any of the mode capacity configurations from a berth perspective. Other system components such as line and vehicle capacity reach saturation well in advance. This is attributed to the amount of space each station has to expand. This space was determined by the length available to stack berths in series and is an adjustable parameter in the model. Based on this parameter value, the available space allowed additional berths to be provided at all stops across the corridor as total corridor demand increased. At the stations with the lowest demand, the combination of space allowed for berth capacity provision and the effects of trip distribution, required a substantial increase in total corridor demand before saturation was reached at these stations and stops, driving up the system saturation values. On the other hand, 'bottlenecking' effects do occur because the station or stop that experiences peak loading eventually runs out of space for providing additional berths. From the results there is no clear best mode capacity configuration, although some are more resilient than others against bottlenecks occurring. In this regard, the high order mode capacity configurations tend to be more resilient. A representation of capacity utilisation of the mode capacity configurations is given in Figure 3-5 with an associated illustration of system saturation and 'bottlenecking' points given in Figure 3-6.



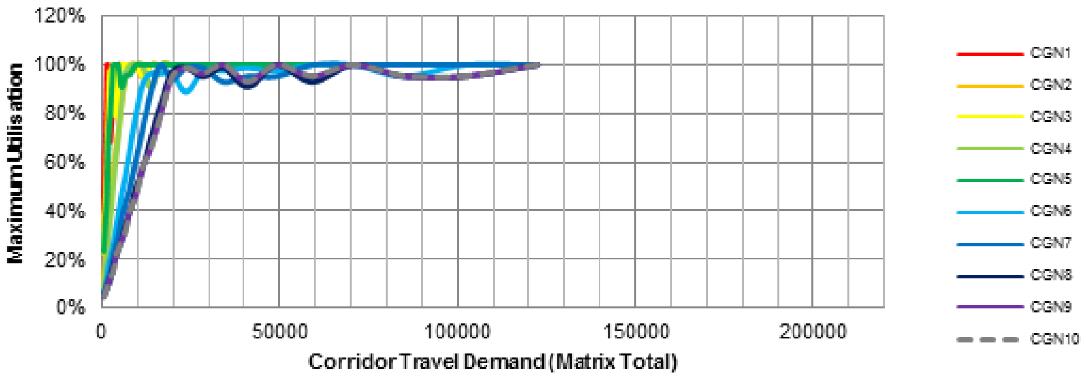


Figure 3-5 Utilisation of Capacity and Productivity Potential: Station/Stops: Berths

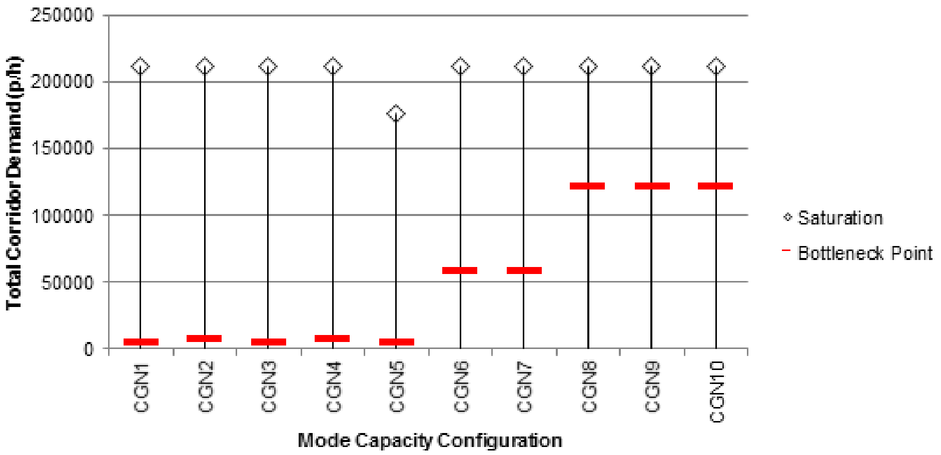


Figure 3-6 Comparative, Assessment of System Saturation and 'Bottlenecking': Station/Stops: Berths

3.4. Stations / Stops: Platforms / Waiting Areas

The provision of platforms and waiting areas is driven by demand, frequency of service and berth requirements. While keeping a fixed platform or waiting area width, the platform face has to extend to cover all berths to support simultaneous loading. In some instances, while demands are not excessive, high berth requirements can drive the need for a larger platform or waiting area. From this perspective, platform and waiting area utilisation is in many instances linked to berth capacity utilisation. The lower the berth capacity, the more berths are required, which in turn drives longer platforms and ultimately creates more waiting area capacity. From the results, there is clearly no configuration that performs best, with the exception of CGN5 which is significantly underutilised due to its low, unit berth capacity. A representation of capacity utilisation of the mode capacity configurations is given in Figure 3-7 with an associated illustration of system saturation and 'bottlenecking' points given in Figure 3-8.

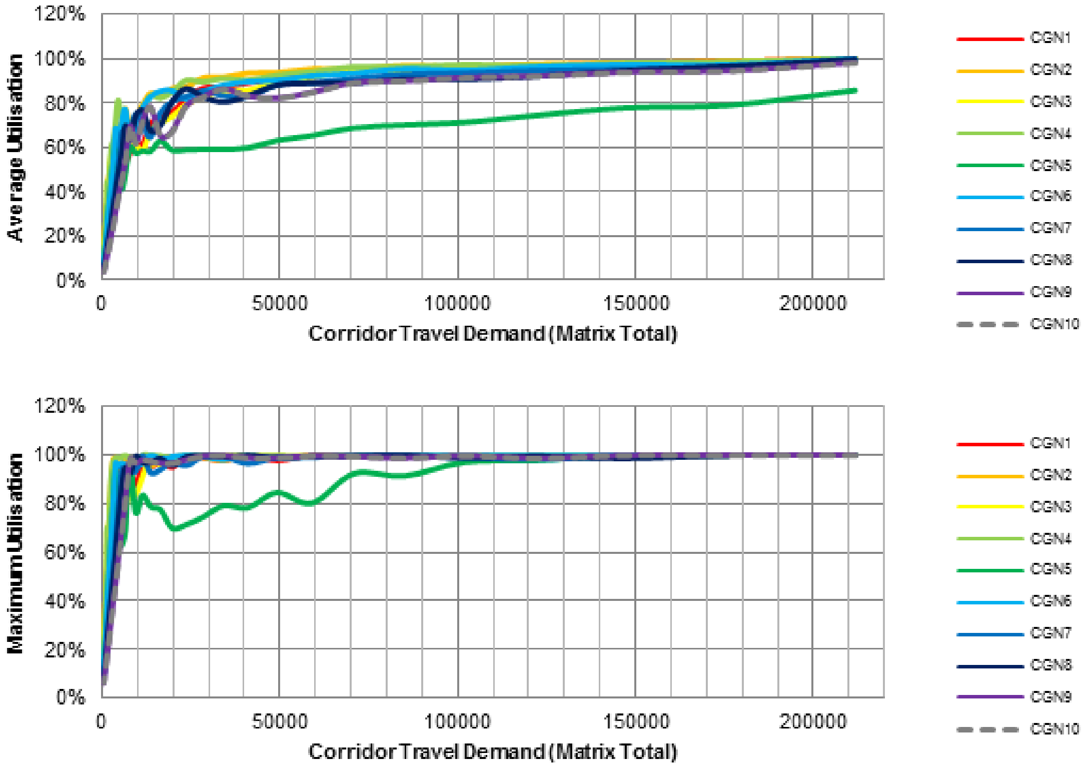


Figure 3-7 Utilisation of Capacity and Productivity Potential: Station/Stops: Platforms and Waiting Areas

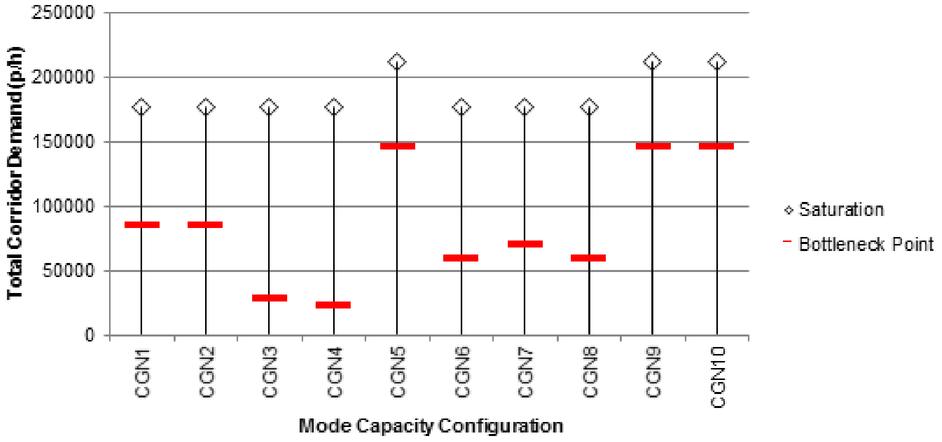
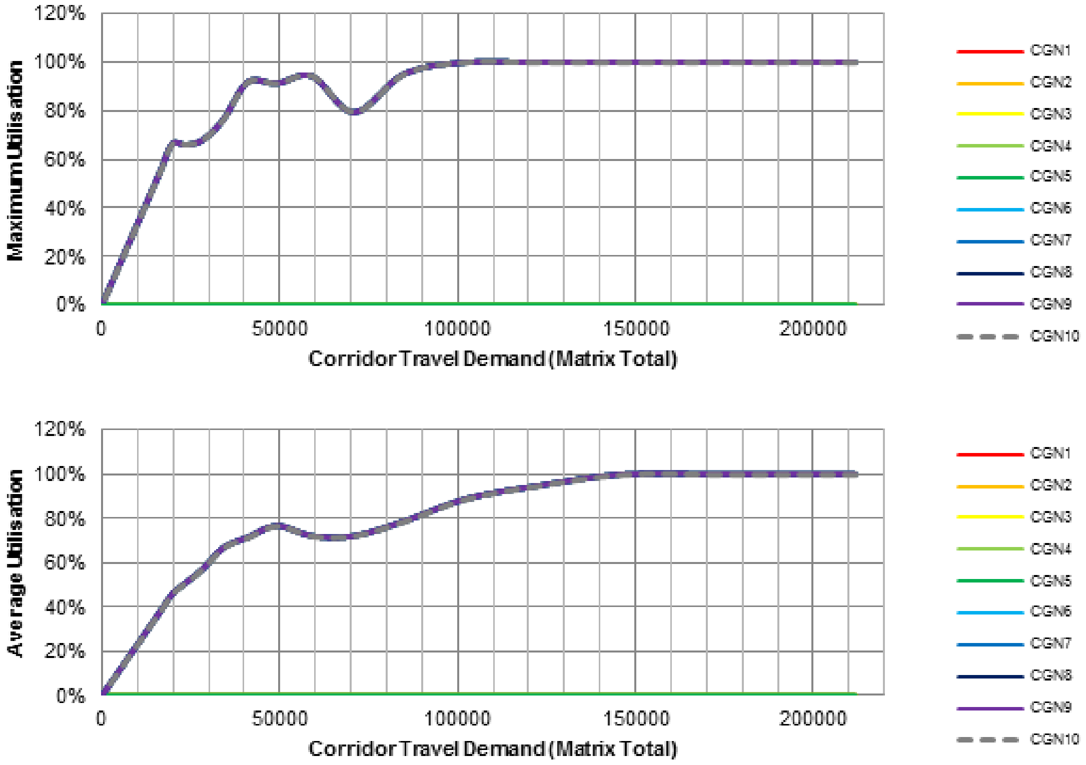


Figure 3-8 Comparative, Assessment of System Saturation and 'Bottlenecking': Station/Stops: Platforms and Waiting Areas

3.5. Stations / Stops: Ticket Gates

Ticket gate utilisation is driven by per-minute demand and the number of ticket gates provided. While the demand influences the number of ticket gates required, the maximum number of ticket gates that can be provided is governed by the platform and waiting area widths. For the purposes of this research, a fixed width of 5m was considered, which allows a maximum of six ticket gates. At a minimum, two ticket gates were provided to allow for directional flows. On this basis, the utilisation of ticket gate capacity for each mode capacity configuration that has a closed preboarding fare collection system is the same since demand is common across all configurations.

The ticket gate capacity utilisation results for CNG2, CNG4, CNG6, CNG7, CNG8, CNG9 and CNG10, revealed that system saturation from the perspective of ticket gate capacity utilisation occurs at a total corridor demand of 147 020p/h. At approximately, 65 000p/h all six ticket gates are required from which point average capacity utilisation across the system climbs until saturation is reached. 'Bottlenecking', however, occurs sooner than total system saturation and represents the point in the system at which saturation first occurs. This is typically the point most loaded and where additional capacity cannot be practically provided. In this instance, bottlenecking occurs at 102 090p/h. A representation of capacity utilisation of the mode capacity configurations is given in Figure 3-9 with an associated illustration of system saturation and 'bottlenecking' points presented in Figure 3-10.



Note: CNG 1,3 and 5 have on-board fare collection. Consequently, ticket gates are not required.

Figure 3-9 Utilisation of Capacity and Productivity Potential: Stations/Stops: Ticket Gates

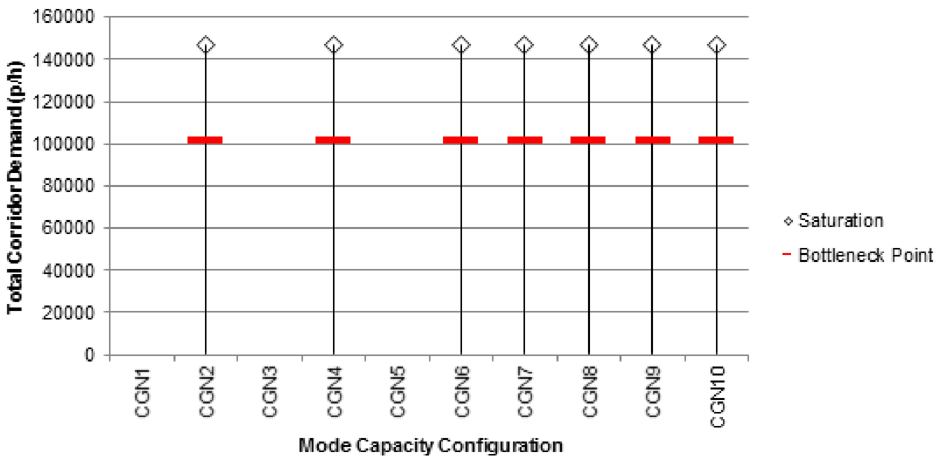


Figure 3-10 Comparative, Assessment of System Saturation and 'Bottlenecking': Stations/Stops: Ticket Gates

4. CONCLUSION

This paper set out to investigate: how scalable road-based transit systems are; how demand estimation accuracy effects the provision of transit capacity; and how much spare capacity should be provided from the start or in increments.

With regard to the scalability of road-based transit systems, it is concluded that road-based modes are scalable within families of configurations, with scalability being understood as the ease with which additional capacity can be provided. Lower capacity modes are more scalable up to a point in low to moderate growth scenarios, whereas higher capacity modes are more scalable in high growth scenarios. This illustrates that anticipated growth rate and not just ultimate demand should play a role in mode selection and capacity configuration. In low growth scenarios, the infrastructure productivity of high capacity modes is poor. The scalability of transit modes, and the incremental provision of capacity, are themes not well considered in many system planning guidelines. The tCAT model outputs provided estimates of the basic increments in which capacity could be provided for each mode. This information provides greater insight into which modes are capable of more precisely matching capacity to incremental increase in demand over time, thereby preserving productivity of the transit asset throughout its life cycle.

With regard to the importance of demand estimation accuracy for capacity provision, it is concluded that, while an accurate demand estimate is important from a revenue estimation perspective, demand accuracy from an infrastructure sizing perspective is less important, especially if a robustness analysis is undertaken as defined in the tCAT. The resultant capacity of certain components is influenced by the capacity of other components, as opposed to being directly influenced by forecast demand. Many existing transit system planning guidelines focus on selecting modes based on line capacity, or the design of infrastructure elements as discrete components of a system. It is important to consider overall system capacity, as well as the interrelation between the various system components. For example the capacity provision of one component could drive a higher unintended capacity of another, reducing the importance of the accuracy of demand estimation for the sizing of certain system components.

With regard to optimal spare capacity, it is concluded that the amount of spare capacity provided at the outset or at key upgrade points depends upon the mode capacity configuration, its capacity increase increment size and the forecast rate of growth, not necessarily the ultimate demand.

Ultimately, scalability, spare capacity provision and sensitivity to demand estimation is related to driving more productive, resilient infrastructure, and ultimately, doing more with less. This philosophy in turn ensures better economic returns and a more sustainable future.

REFERENCES

Cervero, R. 2013. “Transport Infrastructure and the Environment: Sustainable Mobility and Urbanism”, Paper prepared for the 2nd Planocosmo International Conference, Bandung Institute of Technology, 2, 4, 9, pp. 10-11.

Committee of Transport Officials (COTO). 2011. “South African Road Classification and Access Manual”, South African National Roads Agency Limited, Pretoria, 46.

Presidential Infrastructure Coordination Comission. 2013. “A Summary of the South African National Infrastructure Plan”, Republic of South Africa, Pretoria

Reddy, J. 2014. “The Sensitivity of Public Transport Infrastructure Design to Demand Estimation”, Master of Engineering (Transport Studies) 60-credit minor dissertation, University of Cape Town, Cape Town.

Transit Cooperative Research Program. 2003. “Transit Capacity and Quality of Service Manual, 2nd Edition”, Transportation Research Board, Washington DC, 1-13 – 1-19, 1-21, 2-14, Part 4.

Vasconcellos, E. 2001. “Urban transport, environment and equity: The case for developing countries”, London: Earthscan.