Developing an alternative approach to mode choice modelling with the application of modelling Gautrain patronage

by
André Louis Marais

Thesis presented in partial fulfilment of the requirements for the degree Master of Science in Engineering at Stellenbosch University

Supervisor: Dr Simen Johann Andersen
Faculty of Engineering

April 2014

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof, that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2014

Abstract

Mode choice modelling is an important and versatile tool that can aid decision makers with transit related strategies and scenario planning. The traditional approach to modelling public transport is labour intensive and requires many resources. The expensive nature of developing mode choice models can also act as a deterrent for developing a model. Not having access to a functional mode choice model can force decision makers to make important decisions without having access to proper information. There is therefore a need to provide a simplified solution for developing a functional mode choice model that can be developed and maintained with fewer resources.

This research project explores the possibility of developing a simplified alternative approach to public transport modelling that can model mode choice behaviour with the same degree of accuracy as traditional models. The modelling steps employed in this research project were the typical four step demand modelling approach, but the principles employed differ slightly. The focus area of this research project is the development of simplified utility functions and the calibration thereof. Typical mode choice models coincide with many assumptions, variations and uncertainties. In this research project the proposed utility functions are simplified by incorporating most of the assumptions and intangible components of the utility function into a single station to station specific calibration factor. The hypothesis is that a simplified alternative approach to the utility functions can still provide a model that is purpose built and functional.

The application of the proposed mode choice model is to model the mode choice between the Gautrain and private vehicles as the major mode of transport.

The following dates are used for the purpose of this research project:

- February 2013 is the base year scenario
- August 2012 is used to do backward predictions
- August 2013 is used as a short term future scenario testing.

A constant utility approach combined with a Logit probability function was utilised in determining the mode split probabilities for each station to station OD pair.

With the implementation of Intelligent Transport Systems high detailed traveller information is utilised. The model can therefore be accurately calibrated to precisely replicate the base year scenario.

The following tangible/measurable attributes are incorporated into the utility functions:

- Journey lengths in both distance and time
- Value of time
- Vehicle operation cost
- Gautrain ticket pricing
- Gautrain bus fares
- Parking cost at stations
- Train frequency.

The following three variables are the only additional factors that are included:

- Calibration Factors
- Sensitivity Factor
- Seasonal Factor.

The model's prediction capabilities are deemed adequate with an acceptable goodness of fit between the observed and modelled patronage. The table indicates the capability of the model to accurately model the total patronage for the various scenarios. The R-squared values and overall correlation between the observed and modelled patronage for the individual station to station OD pairs are also listed.

Scenario	Observed Patronage	Modelled Patronage	Overall Correlation	R-Squared
August 2012	5 691	5 691	0.961	0.924
August 2013	5 380	5 375	0.975	0.950

With the development of the simplified mode choice model, it is concluded that the Gautrain patronage can be modelled with acceptable accuracy. The overall correlation achieved in this research project proved to be higher than what is typically accepted with a traditional modelling approach. The proposed simplified model is therefore considered as a feasible solution to modelling public transport ridership with fewer resources and in a shorter time frame.

Table of Contents

List of Figures

List of tables

List of Acronyms

AA	Automobile Association of South Africa
BRT	Bus Rapid Transit
CPIX	Consumer Price Index
EFC	Electronic Fare Collection
GITMP25	Gauteng Integrated Transport Master Plan for the next 25 years
ITS	Intelligent Transport Systems
OD	Origin Destination
PRASA	Passenger Rail Association South Africa

1. INTRODUCTION

All major metropolitan areas within South Africa are faced with extreme challenges in meeting the ever-increasing travel demand. Gauteng, being the heart of the South African economy, is confronted with major transport related challenges, currently and in the future. With the ever increasing private vehicle ownership, urbanisation and influx of people seeking economic prosperity, government needs to provide the necessary infrastructure and solutions in order to accommodate this ever growing transport demand.

According to the Gauteng Integrated Transport Master Plan for the next 25 years (GITMP25) the core principles for future planning are to limit urban sprawl by land-use densification and improve the mobility, safety and capacity of the transport network through the enhancement of the public transport network. The following systems listed below, some of which are still in design phase while others are already operational, prove the government's dedication towards the vision of moving people from private vehicles to public transport.

- Gautrain – Rail service between Pretoria and Johannesburg
- Passenger Rail Association South Africa (PRASA) improvement plan
- Rea Vaya – City of Johannesburg Bus Rapid Transit (BRT)
- City of Pretoria BRT
- City of Rustenburg BRT
- City of Ekurhuleni BRT.

All of the above systems can benefit vastly from having access to a model that is designed specifically around each individual public transport system.

Transport modelling is a powerful tool that can aid decision makers with forward planning, evaluating various scenarios and the development of transport related strategies. The traditional approach to modelling public transport is labour intensive and requires many resources. The high cost involved in developing a mode choice model generally requires that the model be developed over a long period of time and therefore typically incorporates outdated or inadequate surveyed data. The high cost of developing a model can also act as a deterrent for developing a model. Occasionally it is considered to be too expensive and then no model is developed. It is also expected that in many instances, given budget constraints, once the model has been developed, it is incorrectly used for extended periods of time without being rebuilt, recalibrated or re-evaluated.

Human behaviour will change over time and so will the decision making process around mode choice. It is important that a mode choice model is updated and recalibrated regularly in order to accommodate for the change in human behaviour.

A need therefore exists to provide a simplified solution for developing a functional mode choice model that can be developed and maintained with fewer resources. The aim of this research project is to explore such a simplified alternative approach to public transport modelling that can model mode choice behaviour with the same degree of accuracy as traditional models. The core design principle is to develop a model with fewer resources, but without sacrificing the functionality and accuracy of the model. The development of a simplified model that is less dependent on large amounts of resources can be developed and implemented in less time. A shorter development time will decrease the cost of developing a model and increase the possibility that a model can be developed for individual public transport systems. As the overall development cost of the model decreases, it is expected that more iterations of the model will become feasible. Validating and recalibrating the model on a regular basis will ensure the improvement of the model resulting in a model that stays applicable and useful.

Most public transport analyses are done with the aid of highly detailed transport planning models. These models are based on the traditional four step modelling approach i.e. trip generation, trip distribution, mode split and assignment. The assigned network values are compared with surveyed data and the model is calibrated in order to resemble the surveyed data as closely as possible. Regression analysis is used to determine various coefficients and factors that are included into the mode split calculations in order to calibrate the model. With these models the desired outcomes are not always guaranteed. These models typically require many resources and surveyed data for the calibration process of the various road link segments and route choice models. With the above mentioned approach the mode split, and therefore the ridership of the public transport network, is usually embedded in many other network calibration factors.

Traditionally, the collection of traveller related information is extremely expensive and labour intensive. With the implementation of various technological devices in modern society, the amount of usable data created every day increased dramatically. Compared to traditional surveyed methods, the sample size of data collected from technological devices are typically much larger and results in a higher degree of certainty. In some instances, the sample size might even be equal to the population size. With technological devices, the information is created automatically and is stored on a continuous basis. With data created and stored continuously, no sampling is required and information for any period can be extracted from the data banks.

In order to calibrate and validate a model, a couple of iterations of the model are required. With the availability of data on a continuous basis, it is possible to do multiple iterations for various time periods without repeating surveys or making unnecessary assumptions. As the available data increases and the analysis processes are optimised, so does the speed at which iterations are done. Future iterations of the model can also be completed with fewer resources because no additional surveys are required.

As the availability of data changes over the years, so should the way it is utilised and implemented.

The application of the proposed mode choice model is to model the mode choice between the Gautrain and private vehicles as the major mode of transport for a journey. With the implementation of Intelligent Transport Systems (ITS), such as Electronic Fare Collection (EFC) on the Gautrain network, very detailed traveller information is available. Because of the use of smartcards and a tap-in; tap-out fare pricing strategy, accurate Station to station origin destination (OD) patterns can be obtained. This information is generated automatically and on a continuous basis and no sampling is required. With accurate Station to station OD patterns and Station to station OD specific calibration factors, one can calibrate the model to exactly replicate the base year scenario. Thus the relevance of this research project is not how accurate one can model the base year, but rather how accurate the model can predict patronage, given the change in certain input parameters.

The following dates are used for the purpose of this research project:

- February 2013 is the base year
- August 2012 is used to do backward predictions
- August 2013 is used as a short term future scenario testing.

This research project utilises the basic principles of the four step model, but also explores an alternative approach to the development and calibration of the mode split calculations. Typical network calibration involves the calibration of the model against surveyed data that are usually aimed predominantly at the private vehicle mode. Typical surveyed data include surveyed roadside interviews, link counts and stated preference surveys. This approach to data collection is labour intensive, typically coincides with a small sample size and cannot deliver a single unique answer. Many assumptions must be made and the calibration process is heavily reliant on the interpretation and experience of the modeller. The proposed approach is to focus predominantly on the applicable public transport mode, i.e. the Gautrain, and to consolidate the rest of the network into a single mode. The applicable public transport mode becomes the focal point and the model is calibrated against the observed patronage of the applicable public transport mode. Little to no attention is given to the rest of the network and the other modes. No additional network assignment and calibration is required. Because the model developed for this research project only focuses on the Gautrain patronage, many of the deeply embedded calibration factors are removed and the main emphasis can be on the calibration of the Gautrain network. The above mentioned approach thereby reduces the amount of surveyed data requirements and assumptions. The hypothesis is that the reduction in model complexity will still provide adequately accurate results.

The mode choice between utilising the Gautrain versus private vehicle is done with the aid of detailed utility analysis. The perceived cost to complete a desired journey is calculated for the various modes and a Logit probability distribution function is used to determine the mode split for each OD pair. The use of high detailed, yet simplified, utility functions in conjunction with accurate patronage information gears the model to better predict the influence of various input parameters on the public transport ridership.

It is generally accepted that, because a model is a simplified representation of reality and because of the challenges in modelling subjective decision making processes, the outputs will not be 100% accurate. The usefulness of a model is consequently measured on whether the outputs are sufficiently accurate in comparison to traditional models. The typical process in determining if the model is fit for purpose is by comparing the model outputs with observed data. A correlation between the two data sets is determined and a goodness of fit is established.

2. LITERATURE REVIEW

"Where policies and strategies are developed without recourse to modelling, these are likely to be ineffective, short-lived, have unintended consequences and may even be counter-productive." (Furnish & Wignall, 2009)

Developing traditional mode choice models are considered as resource intensive. A traditional model was therefore not developed in order to directly compare its performance with that of the proposed model. This chapter provides the reader with more detail on general modelling principles and the extensive resource requirements associated with developing a traditional mode choice model. The first part of this chapter provides more detail on the traditional four step modelling approach, after which a couple of case studies are presented. The purpose of the case studies is to illustrate the extent of the typical requirements and the accepted goodness of fit associated with traditional mode choice models. The chapter is concluded with a discussion on simplified and alternative approaches to traditional modelling methods.

Transport modelling is the mathematical representation of the supply and demand of numerous elements in a transport network. The demand is based on the desired journeys within the network and the supply is based on the available infrastructure that enables one to reach a desired destination. The demand is typically divided into the highway network and the transit network. Transport models are widely used all over the world with the following applications: (Metropolitan Washington Council of Goverments):

- Demand forecasting
- Estimating demand in the absence of observed data
- Scenario testing
- Project planning.

Conventional transport models are based on the traditional four step modelling approach and are typically utilised to replicate the following (Furnish & Wignall, 2009):

- Current levels of demand
- Movement patterns
- System capacities.

The four step modelling approach is widely used for transport modelling. The basic four steps are trip generation, trip distribution, mode choice and assignment. The proposed mode choice model utilises the basic principles of the four step model, but also explores an alternative approach to the development and calibration of the mode choice calculations. The proposed mode choice model is developed to explore the possibility of developing an acceptably accurate mode choice model that requires far less resources than what is typically associated with traditional models.

Some studies also refer to the fourth step as “route choice” rather than assignment (McNally, 2007). After the completion of the four steps, the calculated flows are compared to surveyed data and the model is calibrated accordingly. The calibration is done by making alterations to the assumptions, input parameters and calculations utilised in the last three steps of the four step approach. Figure 1 illustrates the four step model approach (McNally, 2007).

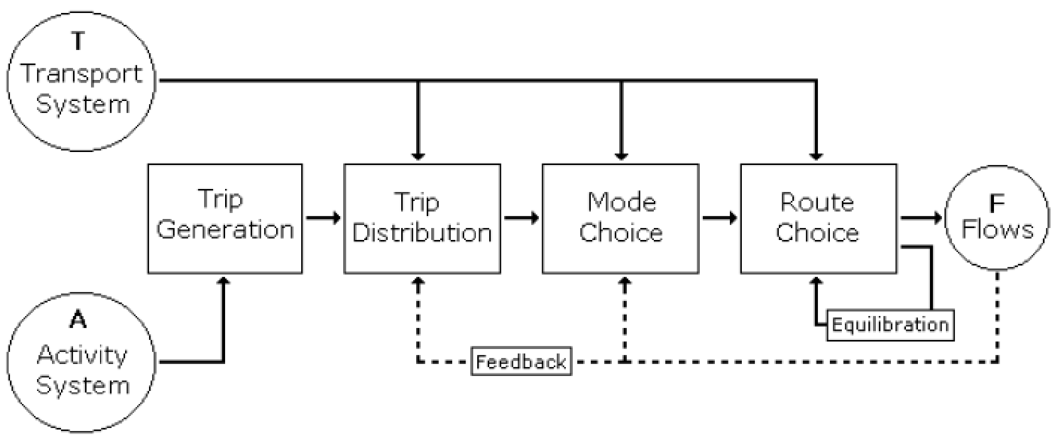


Figure 1: Four Step Model source: (McNally, 2007)

The transport system incorporates the infrastructure and transport related services and defines the available supply of the transport network. The transport system has an impact on the trip distribution, mode choice and route choice. The activity system defines the demand of the transport system and incorporates the spatial developments, land use, economic activities and demographics of the population. The activity system will have an impact on the trip generation rates utilised in the model.

According to McNally (2007) the application of typical demand models is a continuous process that may extend over a couple of years. The duration of data collection alongside the development of the model can sometimes extend over such a long period that the transport environment can actually change considerably during the analysis period. If the development of the model takes too long, the surveyed data become obsolete and the modelling exercise becomes less effective.

Traditionally the cost of developing and maintaining a model is expensive. “The cost of applying the models constitutes almost half of the region’s transportation planning budget, including data support” (Metropolitan Washington Council of Goverments). It is therefore important to strive to develop models that are less dependent on time consuming surveys and laborious data processing approaches. The availability of comprehensive and accurate data derived from technologies inside and outside the transportation industry is ever increasing. It is therefore of outmost importance to keep changing the way transport modelling is done in order to optimise the utility of the ever increasing available data. Section 2.6 provides more detail on the alternative approach to traditional modelling.

The following sections provide more background on the trip generation, trip distribution, mode split and assignment steps of the four step demand modelling approach.

2.1 Trip Generation

Trip generation is a process that determines the number of trips and frequency thereof in the network. According to Ortuzar & Willumsen (1990) the two different approaches are either to make use of discrete choice models, or to use data obtained from the household socio-economic attributes of each zone. Discrete choice modelling determines the probability of a trip from disaggregate data, calculated from observed trip patterns of individuals. This approach requires high detailed information and is feasible for small area studies, but becomes unfeasible for larger networks as the required sample size becomes too large. The second approach i.e. the use of household socio-economic attributes, determines the trip ends in the network. The trip ends are the number of trips being generated by and attracted to each of the zones in the network. Socio-economic attributes are determined from land use data and include attributes such as residential, commercial and industrial areas in conjunction with the population’s demographics.

The socio-economic factors affecting the trip productions from zones in the network are typically:

1. Household income
2. Number of occupants per household
3. Daily activities of members i.e. working, unemployed, school or other educational activities
4. Vehicle ownership of households
5. Residential density
6. Value of land within the zone
7. Ease of access to the zone
8. Residential areas in the zone.

The factors affecting the trip attraction towards zones in the network are:

1. Employment opportunities
2. Floor space for commercial and industrial areas
3. Type and density of commercial and industrial areas
4. Accessibility.

The total number of trips generated within the network needs to be equal to the number of trips attracted within the network. If they are not equal, the number of trips generated or the number of trips attracted should be altered to correspond with each other. According to Wegman & Everett (2012), one has more confidence in the trip productions than in the trip attractions. This is largely due to the fact that one is generally more confident in the surveyed data relating to the population and the housing thereof, rather than the data relating to job opportunities and employment. Thus, if the two sums are not equal, it is general practice to alter the trip attractions in order to correlate with the number of trip productions.

2.2 Trip Distribution

Trip distribution is the step during which one determines the transport demand between the various zones in the network. Each zone pair (OD pair) needs to be assigned a certain number of trips, where the total number of trips originating and terminating at the various zones should be in line with the trip generation rates determined during the trip generation step.

"Choosing an adequate representation of transportation demand comprises of a trade-off between model complexity and data accuracy" (Gupta & Shah, 2012).

According to Gupta & Shah (2012) the following trip distribution methods can be used to calculate the network trip distribution.

1. Simple methods
    a. The growth-factor method
    b. Tri-proportional method
2. Theoretical Models
    a. Gravity Models
    b. Entropy models
3. Counting based methods.

Growth-factor and Tri-proportional methods are done by altering existing or surveyed trip distribution patterns. OD data can be derived from household survey data and roadside interview data. This approach is feasible for small study areas with only a few zones within the study area. A trip distribution pattern is calculated from the data and is adjusted to correlate to the trip-ends of the trip generation. As the study area increases in size, so does the required sample size in order to obtain sufficient usable information. As the sample size increases, the more expensive and time consuming the study becomes.

In many instances, if surveys are considered too costly and time consuming for the project, previously calculated trip distribution patterns can be utilised to create an OD matrix that fits with the calculated trip generation.

In any event, the trip generation matrices need to be altered for the total trips generated and trips attracted to be equal to each other and to be equal to the calculated trip generation rates. The most common process to do this is the Furness, or bi-proportional method. This method converts the "old" OD matrix into a "new" OD matrix with the use of balancing factors.

The new OD matrix "T_{ij}" can be calculated as follow (Gupta & Shah, 2012);

$$T_{ij} = t_{ij} A_i \frac{O_i}{o_i} B_j \frac{D_j}{d_j}$$

Where:

t_{ij} = Original number of trips from origin zone *i* to destination zone *j*

A_i = Balancing factor for origin trips from zone *i*

B_j = Balancing factor for destination trips to zone *j*

O_i = Sum of all trips originating from zone i in the new matrix

D_j = Sum of all trips with destination to zone j in the new matrix

o_i = Sum of all trips originating from zone i in the old matrix

d_j = Sum of all trips with destination to zone j in the old matrix.

Some of the major benefits of using previous OD patterns are that the process is easy to understand, it is comparable with the observed trip matrices and the original trip distribution patterns are preserved during the calculations. The latter is a benefit for short term planning, but might be problematic for long term planning. Trip distribution patterns will change because of future developments and this approach does not allow for a change in the trip distribution patterns. Another shortcoming of this approach is that data needs to be available for each and every OD pair, because if the sampled OD pair value is zero, it will remain zero regardless of the balancing factor. The effect of this can however be mitigated by introducing a minimum number of trips for each OD pair using seed numbers or making certain assumptions where no data is available.

Given all the factors mentioned above, it is clear that this approach requires large amounts of data and is resource intensive.

Trip distribution can also be calculated from theoretical models, such as gravity and entropy based models. Both of these models are derived from laws of physics, where the gravity model is derived from Newton's gravitational law and the entropy model is derived from the second thermodynamics principle.

The gravity based model is adapted from Newton's law of attraction, where the force of attraction is proportional to the mass of the two objects and inversely proportional to the distance between the two objects (Jewett, 2004).

In short, the number of trips between an origin zone and destination zone is a function of the number of trips generated at the origin zone, the number of trips attracted to each destination zone and the deterrence function between different zones. The deterrence is usually a function of the travel cost between the zones.

Some popular deterrence functions are (Ortuzar & Willumsen, 1990):

1. Exponential function $f(c_{ij}) = e^{-\beta c_{ij}}$
2. Power function $f(c_{ij}) = c_{ij}^{-n}$
3. Combined function $f(c_{ij}) = c_{ij}^{-n} e^{-\beta c_{ij}}$

Where:

c_{ij} is the associated travel cost between zone i and j

β and n are calibration constants.

The entropy model is based on the second principle of thermodynamics. According to thermodynamics, an isolated system tends toward disorder and the amount of disorder is measured as entropy (Jewett, 2004).

According to Jewett (2004) a Microstate is a particular configuration of all the individual elements of the system and a Macrostate is the overall condition of the system. In the transport environment the Microstates can be interpreted as all the different individual combinations of inter-zone trip distributions that will produce the specific Macrostate i.e. the desired trip generation rates as determined in step one. One of the assumptions is that all the Microstates are equally probable with the resulting OD pattern corresponding to the highest number of Entropy (Gupta & Shah, 2012).

Counting based methods estimate the trip distribution patterns from link counts. Each OD pair has a path of least resistance. These paths are a combination of various routes and links in the network. Given the various link counts in the network and knowing which OD pairs contribute to each one of the counted values, a trip distribution pattern can be estimated. This approach to trip distribution does not produce a single unique answer and the number of variables and combinations usually far exceed the number of counted data points. The generalised least-square method can be used to determine the best fit, given the various input parameters.

The above mentioned approaches to trip distribution calculations coincide with expensive surveys, small sample sizes and many assumptions. With the implementation of mobile devices such as GPS tracking units and mobile phones, more detailed trip generation and trip distribution data is available to be utilised by transport modellers. Tracking of mobile devices can provide a very high detailed view of the movement patterns of people in the study area and it has the potential of providing a very large sample size.

2.3 Mode Split

Mode split is the process in which the number of trips between zones is divided between various modes of transport. The typical modes include private vehicle, bus, rail, walking, cycling or a combination of them. The choice of mode plays a big role in policy making. Public transport is an essential part of any transport network, and it is important to determine what decision makers can do in order to increase public transport ridership and where it is economically feasible to implement, expand or upgrade the public transport network.

According to Ortuzar & Willumsen (1990) the factors influencing mode choice may be classified into three groups, based on the characteristics of the trip maker, the journey and the transport facility.

The characteristics of the trip maker include factors such as:

1. Vehicle ownership
2. Possession of drivers licence
3. Household structure
4. Income
5. Residential density.

The characteristics of the trip maker are probably the biggest contributing factor for the mode split between public or private transport. If the person does not have access to a private vehicle, does not have a driving licence, or cannot afford a private vehicle, he/she is forced to make use of public transport. This can be considered as a situation of forced mode split between private and public transport rather than mode choice.

The characteristics of the journey may also play a role in a person’s mode choice. These factors include the purpose of the trip and the time of day. People are more likely to make use of public transport for structured predetermined journeys like commuting and rather use their private vehicles for random unplanned journeys. Knowing the schedules, routes and capacity of the public transport system increases the perceived comfort level of the traveller, whereas a certain discomfort can arise when confronted with the unknown. With the use of ITS such as traveller information systems, more information becomes available to the traveller and unfamiliar intermodal journeys can be planned before the time of departure, or while on route.

The characteristics of the transport facility include the tangible/measurable factors such as out of pocket costs for the fare and parking, and the travel time associated with the journey. Other intangible/objective costs include factors such as comfort, safety, mode preference and reliability of the mode. The intangible contributing factors cannot really be quantified and are usually derived from surveyed data.

The mode split can either be done before or after the trip distribution step (Ortuzar & Willumsen, 1990).

In some instances, where the predominant decision making factors are the characteristics of the trip maker, the mode-split calculations can be done prior to the trip distribution step. This approach assumes that the mode choice is heavily reliant on the income level of the individual. This assumption makes the mode choice insensitive to factors such as fare price and travel time. This type of modelling particularly holds true in an environment where people in a lower income group do not really have a choice in the mode they use and the people in the higher income group do not consider public transport as a mode of choice.

Until recently this was particularly true in South Africa, where the public transport services were neglected, unreliable and considered dangerous. Low income workers had to make use of a limited public transport network and a very small percentage of the middle and higher income people used public transport. This however is not the case anymore. Government initiatives, such as the implementation of BRT systems, the proposed upgrading of the passenger rail system by PRASA and the recent construction of the Gautrain, are encouraging more and more middle and high income people to consider public transport as a mode of choice. This is in line with the vision of the GITM25 to move people from their private vehicles to public transport.

If the mode split is done prior to the trip distribution, the personal characteristics play the deciding role in the mode choice and the journey characteristics do not influence the mode choice. If the mode split is done post trip distribution, the journey characteristics are considered for the mode choice, but the individual characteristics are aggregated during the trip distribution process. It is however possible to determine the trip generation and trip distribution separately for each individual user group, but this approach drastically increases the complexity of the model.

According to Koppelman & Bhat (2006) the disaggregate approach (mode split done prior to trip distribution) explains the individual's mode choice based on the circumstances of individual travellers and can therefore incorporate the change in individual characteristics and attributes of alternatives. The aggregated approach (mode split done after trip distribution) however relies on the statistical association among relevant variables on a non-individualised level and can therefore incorporate alterations in the network, services and the population.

Ben-Akiva (Ben-Akiva M. E., 1985) proposed the following decision making process: The decision maker first determines the available alternatives and then considers the various attributes of each of the alternatives. With the above mentioned information, he/she then determines the desired choice by making use of a decision rule. This research project followed the same approach for decision making. The alternatives are whether to use a private vehicle or the Gautrain to reach a destination. Various attributes are considered for each alternative and the mode choice is based on the decision rules that are quantified with the aid of utility functions. The decision rules and utility functions are described in more detail in sections 2.3.1 and 2.3.2.

2.3.1 Decision Rule

According to Koppelman & Bhat (2006), "Discrete choice models can be used to analyse and predict a decision maker's choice of one alternative from a finite set of mutually exclusive and collectively exhaustive alternatives."

Discrete choice models and analysis are based on the theories of individual choice behaviour. The individual's choice can be interpreted as the outcome of a sequential decision-making process (Ben-Akiva M. E., 1985). According to Ben-Akiva (1985) the sequential decision-making process follows the following steps:

1. Definition of the choice model
2. Generation of alternatives
3. Evaluation of attributes of the alternatives

4. Choice
5. Implementation.

Within a transport environment, the decision making process to determine the chosen mode of transport for a particular journey can be interpreted in the following way: During the definition of the choice model, the individual determines the desired outcome of his/her journey. This includes factors such as the origin and destination of the journey as well as the time of day. During the next step the individual will determine possible alternatives that can be utilised in order to meet the desired outcomes of his/her journey. The possible alternatives are the various modes of transport available to the individual. After all the available alternatives have been determined, the various attributes for each alternative are determined and weighed up against each other in order to determine the preferred mode choice. The various attributes of the different modes are consolidated in a mode and journey specific utility function that incorporates attributes such as travel time, travel cost and value of time. By evaluating the various utility functions, the individual will choose the mode with the highest utility. The utility analysis is described in more detail in section 2.3.2. Note that the various mode attributes can also be expressed as a disutility, in which case the preferred mode will be the mode with the lowest utility. The individual then implements his/her choice and completes the journey with the decided mode.

Depending on the number of alternatives to be considered the choice model is either referred to as a binary choice model for two alternatives, or a multinomial choice model where more than two alternatives are considered. For the purpose of this research project, only binary choice models will be considered, but note that the techniques and principles are transferable to multinomial choice models.

The discrete choice theory implies that, given the above mentioned decision making process, all the various individuals confronted with the same journey purpose and utility function should all choose the same mode. This in fact is not true, as human behaviour is inherently probabilistic and contains a certain degree of randomness. Another factor contributing to the observed variation in mode choice is due to the aggregation of the disaggregated individual’s choice behaviour. Given the above mentioned randomness and variation in mode choice, a probabilistic choice model is derived from the discrete choice theory.

According to Ben-Akiva (1985) there are two different approaches to incorporate the variability within a probabilistic choice model - one can either implement a constant utilities method, or a random utilities method.

2.3.1.1 Random Utility

The random utility approach assumes that the variation is due to the observational deficiencies and that the utility function varies between users. It is assumed that the individual will always choose the alternative with the highest utility and the variation in choice between individuals is incorporated by utilising random variables within the utility functions. The random utility thus comprises of a deterministic/systematic component and a random variable component. See section 2.3.2 for more detail.

The probability "$P(i)$" of choosing mode "i" given the alternative choice "j" are as follows (Celikoglu, 2007):

$$P(i) = P\left[U_i > U_j, for\ all\ j \neq i\right]$$

Where U_i and U_j are the utility functions for options *i* and *j* respectively.

2.3.1.2 Constant Utility

For the constant utility approach it is assumed that the utility functions are constant and do not change between various users. In order to incorporate the necessary variation the decision maker's decision is not based on choosing the utility with the highest value, but is instead based on a probability distribution function. The probability distribution functions are functions of the various utilities for the different alternatives. See section 2.3.2 for detail on the utility functions.

The following equations illustrate the various probability functions typically used in order to determine the choice probability "$P(i)$" for choice "i" given the alternative choice "j". (Ben-Akiva M. E., 1985):

Linear probability function:

$$P(i) = \frac{V_i - V_j + L}{2L}$$

for

$$–L > V_i - V_j < L$$

Probit probability function:

$$P(i) = \Phi \frac{V_i - V_j}{\sigma}$$

Logit probability function:

$$P(i) = \frac{e^{V_i}}{e^{V_j} + e^{V_i}}$$

Where:

V_i and V_j are the deterministic/systematic components of the utility functions for alternatives *i* and *j*

L is a predefined constant defining the upper and lower limit of the linear function

Φ denotes the standardised cumulative normal distribution.

From the above mentioned probability functions the Logit function is the function most commonly used to determine travel behaviour (Khan, 2007).

2.3.2 Utility analysis

Utility functions are the mathematical representations of the perceived cost for completing a journey. As mentioned in section 2.3.1, the utility approach can either utilise a constant utility function, or a random utility function. The following attributes are among the many attributes that can be incorporated into the utility functions (Khan, 2007):

- In vehicle travel time
- Out of vehicle travel time
- Access time to transit point
- Waiting time
- Interchange time
- Traveling fares
- Parking

- Other level-of-service attributes.

According to Khan (2007) a utility function is typically expressed as a linear function of the various attributes. These attributes are weighted by the multiplication of a coefficient in order to incorporate the relative importance of the various attributes.
The following linear equations illustrate the various methods of calculating the utility "U_m" given a mode "m" (Khan, 2007):

For a constant utility functions the utility can be expressed as:

$$U_m = \theta_1 X_{m1} + \theta_2 X_{m2} + \ldots + \theta_k X_{mk}$$

Where:
U_m is the net utility function for mode m
$X_{m1}, \ldots, X_{mk}$ are k numbers of attributes of mode m
$\theta_1, \ldots, \theta_k$ are k numbers of coefficient (or weights attached to each attribute) which need to be inferred from survey data.

Note that for the constant utility approach, the utility "U_m" is deterministic, with no randomness factor.

For a random utility function the utility can be expressed as:

$$U_m = V_m + E_m$$

Where:
V_m is the deterministic component of the utility of the mode m
E_m is the error component of utility m.

The systematic component of the random utility function closely correlates to the overall utility of the constant utility function. The random utility function can therefore be expressed in more detail by incorporating the formula used to calculate the fixed utility function (Bierlaire, 1995):

$$U_m = V_m + E_m$$

$$U_m = c_m + \sum_i \theta_i X_i\,(m) + E_m$$

Where c_m is the alternative specific constant.

The number of attributes and coefficients included in the utility functions are important and it has a direct impact on the model complexity (Bierlaire, 1995). The more attributes incorporated into the utility function, the more surveyed data, weighted factors and assumptions are required. The number of attributes and weighted factors also has a big impact on the complexity of the calibration process. Providing additional variables, assumptions and uncertainties into the choice behaviour modelling process will not necessarily produce more accurate results.

2.3.3 Variable External Factors that Influence Mode Choice

The final choice of mode does not only depend on the above mentioned decision rules and utility analysis. If an individual decided on the preferred mode to utilise, random external factors can cause the user to alter his/her original choice. The external random factor affecting the final mode choice includes among other the following:

- Occurrence of public holidays
- Occurrence of special events, such as big concerts and sporting events
- Network delays caused by incidents
- Adverse weather conditions.

Public holidays normally coincide with a large reduction in peak hour traffic. Some people might utilise public transport in order to avoid long delays in congested areas. If there is no congestion, people might choose to utilise their private vehicle instead of public transport. In some instances the transit ridership can increase on public holidays. An increase in public transport leisure related journeys and the influx of people to holiday destinations will increase the demand on the transit services and the transit ridership will increase during these periods.

Sporting events and big concerts may also stimulate an increase in transit ridership if the venue is close to a public transport system.

Incidents on the network can also affect the choice of mode. If there is an incident on the motorways causing long delays, people might rather utilise public transport, such as rail, that might not be affected by the incident. The contrary is also true that, if an incident occurs on the public transport network, people might also change the original choice of mode.

Weather conditions can have a substantial effect on transit ridership. According to Guo (2007) the weather conditions have an impact on the traveller's activities and travel experience. Inclement weather conditions generally coincide with a reduction in personal activities and a mode preference towards private vehicles.

Inclement weather conditions will have an effect on the trip demand, as people are likely to reduce the number of required trips because of the adverse weather conditions. The number of transit journeys will consequently reduce, with a reduction in overall trip demand.

Bad weather conditions will also have a negative effect on the experience of a traveller using public transport. Waiting for and transferring between transit services in bad weather conditions may be uncomfortable. Utilising a private vehicle in the same weather conditions is typically more comfortable and, if it is available, will be the preferred choice. The perceived comfort levels will have an impact on the mode choice and the public transport ridership will decline.

Guo (2007) suggests that weather conditions such as temperature, wind, snow and rain do impact the transit ridership, but the impact thereof is dependent on the specific mode, transfer and waiting areas, time of day, days of the week and season.

2.4 Assignment and Calibration

During the assignment step the calculated demand is loaded onto the network. The modelled values are compared with the observed values and a correlation test is done. The correlation between the modelled values and observed values is improved through a process known as calibration. During the calibration process, certain input parameters, coefficients and calibration factors are altered in order to produce a higher correlation. Regression analysis is also utilised to determine the relationship between the respective variables and the statistical significance thereof.

After the input parameters, coefficients and calibration factors have been altered, the assignment step is repeated with the new values. The assignment process is an iterative process, where the calculated demand is loaded onto the network, a correlation test is done and the network is calibrated accordingly.

The calibration process usually comprises the utilisation of many of the available resources. Given the interdependency between all the various input parameters, the many assumptions associated with building a model and the absence of single unique answers, the outcome of the model is based on the judgement and experience of the individual modeller. A more detailed description of the mode choice calibration is described in section 2.4.1.

2.4.1 Mode choice calibration

The calculated mode choice is directly dependant on the various utility functions. The mode choice (or mode split) calibration is done by comparing the observed mode split against the modelled mode split. According to Celikoglu (2007) the calibration process involves the following steps:

- Estimate the parameter values
- Evaluate the statistical significance of the parameters
- Validate the model by comparing modelled prediction with the observed behaviour.

Each of the parameters contained in the utility functions are estimated. These parameters include the measurable and non-measurable attributes of the journey. The measurable (tangible) attributes are the attributes that the modeller can quantify in time or monetary values. These include the travel times and travel costs associated with a particular journey. The non-measurable (intangible) attributes are those attributes of which the perceived value are derived from surveyed data or assumptions. These factors typically include the following (Zhao, Li, Chow, Gan, & Shen, 2002):

- Value of time
- Safety

- Comfort
- Cleanliness
- Appearance
- Pedestrian environment
- Luxury
- Scenery
- Biasedness towards a mode
- Public perception.

The statistical significance of the parameters is represented by the weighted coefficients for each attribute. Attributes with a higher significance, such as direct out of pocket costs, are assigned larger weighted factors than attributes with a lower significance. The weighted factors are derived from previous studies, stated preference surveys and assumptions. The development of these coefficients typically coincides with many assumptions and uncertainties.

The final step of the calibration process involves comparing the modelled results to the observed results and altering the calibration coefficients and mode specific constants.

In some instances certain weighted factors and parameters may be altered if there are grounds supporting the alteration thereof. In general, one will change the intangible input parameters and weighted factors that are based on assumptions rather than the tangible parameters and weighted factors supported by available data.

In order to calibrate the mode split probability, calibration coefficients and a mode specific constant are combined with the utility functions as mentioned in section 2.3.2 (Celikoglu, 2007).

$$U_m = \lambda\left(\sum_i \theta_i X_i\,(m) + \delta\right)$$

Where:

δ is the mode specific constant

λ is the calibration coefficient.

Depending on the level of detail and model complexity, the number of calibration coefficients can vary. For a low complexity model, a single universal calibration coefficient may be used, whereas a more detailed model may incorporate more journey specific coefficients.

There are many assumptions and parameters incorporated within the utility functions that are used to model mode choice behaviour. The challenge of the calibration process is that the various input parameters are typically interdependent and the effect they have on the overall mode split is nonlinear. Neural network analysis can be used to identify the effect that various input parameters have on the modelled choice behaviour and therefore assists in the calibration process (Celikoglu, 2007).

The mode choice model is calibrated against the calculated public transport ridership. Traditional public transport ridership information is obtained from stated preference surveys, ticket sales and passenger counts at stations. These techniques require additional assumptions and calculations to determine the estimated patronage. The ticket sales are typically a combination of seasonal tickets and single journey tickets, passenger counts only provide the number of passengers entering or exiting the station and stated preference surveys only indicate the opinion of some individuals. Deriving the patronage from the above mentioned approaches coincide with a certain amount of error and uncertainty. The typical approach is to estimate the various link flows on the public network and calibrate the network accordingly.

With more accurate OD patronage data the estimated link volumes can be replaced with accurate observed station specific OD patronage. The model can then be calibrated against true patronage volumes for each individual public transport OD pair.

2.5 Case studies

The following case studies are presented to illustrate the required resources and level of uncertainty associated with traditional mode choice models.

The first case study involves the modelling of intercity travel mode choice behaviour for non-business trips within Libya (Bin Miskeen, Alhodairi, & Bin O.K. Rahmat, 2013).

Extensive surveys were conducted at airport terminals and along the major routes between the various cities included in the study. Stated and revealed preference surveyed data was obtained from the private vehicle and aeroplane passengers. The following information was obtained from the surveys:

- Socio-economic aspects of individuals
- Trip information
- Attitudes and perceptions on travel and policy measures.

The following parameters were among the input variables that were considered to be included in the proposed utility functions (Bin Miskeen, Alhodairi, & Bin O.K. Rahmat, 2013):

- Gender
- Nationality
- Educational level
- Household income
- Household vehicle ownership
- Family trip
- Distance of trip
- Access and egress distances to airports
- Total travel cost
- In vehicle travel time
- Out of vehicle travel time
- Duration of stay
- Privacy factor
- Convenience factor
- Comfort factor
- Reliability
- Safety
- Weather conditions.

Considering the above 18 variables, the original utility functions consisted of the following input parameters:

- 2 mode specific constant
- 18 variables as input parameters
- 18 coefficients (weighted factors)
- 2 mode specific error component

The coefficients were approximated with regression analysis and by fitting the data to the model.

A couple of models had to be developed in order to determine the statistical significance of the various variables, coefficients and constants. "A few of the models tested have revealed inadequate statistical goodness-of-fit and/or weird signs, consequently they all were invalidated" (Bin Miskeen, Alhodairi, & Bin O.K. Rahmat, 2013).

The model was further developed through the process of determining and removing the variables with trivial coefficients and variables with incorrect signs associated with them. Table 1 indicates the variables along with their respective coefficients that were included in the utility function of the final model.

Table 1: Variables and corresponding Coefficients

Variable	**Coefficient value**
Nationality	4.179
Age	2.121
Educational level	4.643
Household income	-3.023
Household vehicle ownership	1.290
Duration of stay	-0.189
Access and egress distances to airports	-0.401
Out of vehicle travel time	-0.145
Family trip	-2.907
Total travel cost	0.006
Comfort facto	2.343
Weather conditions	-3.581

The final model produced an R-squared value of 0.664 and was considered as adequately accurate.

The second case study involves the determination of the effect of model specification on valuation of travel attributes with the application of feeder systems in rural India (Maitra, Ghosh, Das, & Boltze, 2013).

The following models were developed for comparison:

- Multinomial Logit

- Heteroskedastic extreme value
- Nested Logit
- Covariance Heterogeneity nested Logit
- Random parameter Logit

The following input variables were included in the utility functions (Maitra, Ghosh, Das, & Boltze, 2013):

- Access mode type
- Seating discomfort
- Access walking distance
- Anxious waiting time at stop
- Relaxed waiting time at stop
- Relaxed waiting time at home
- Cost
- Various access modes schedules and availability variables.

Table 2 lists the R-squared values obtained for the various models. An overall R-squared value of more than 0.2 was considered as a good fit.

Table 2: Summary of R-squared values for various Model approaches

Model approach	**R-Squared**
Multinomial Logit	0.203
Heteroskedastic extreme value	0.191-0.209
Nested Logit	0.205
Covariance Hetrogeneity nested Logit	0.213
Random parameter Logit	0.207-0.236

The third case study involved determining the factors effecting urban transit ridership in Canada (Kohn, 2000). “Factors that affect supply and demand are complex, constantly changing and difficult to identify and discern.” (Kohn, 2000).

The following input parameters were initially considered:

- Data elements including demographics
- Hours of service
- Fare structure
- Vehicle statistics
- Energy consumption

- Employment
- Passenger statistics
- Revenues and expenditures.

The above mentioned data was collected over a period of seven years. Regression analysis was utilised to determine if the change in the various input parameters is the cause of the change in ridership. Initially the model did not provide adequate results and dummy variables had to be added to increase the correlation between the modelled and observed transit ridership values.

The following dummy variables were added:

- Annual dummy variables to compensate for differences on an annual basis
- City specific dummy variable to compensate for the difference in population size between the various cities
- City specific dummy variable to compensate for the difference in the availability of transit systems in various cities.

With the incorporation of the above mentioned dummy variables the R-squared value improved from 0.5 to 0.7. The overall correlation was considered as sufficient, but in most instances the individual data points displayed large residual errors (Kohn, 2000). The model was improved with the incorporation of an additional dummy factor that incorporated the ridership rate of the various cities. The R-squared value increased to 0.88.

"Sensitivity analysis was restricted because of the large number of dummy variables compared to variables with observable and useful data. As a result, it was difficult to conduct meaningful sensitivity analysis since estimations would be similar if the dummy variables were consistent." (Kohn, 2000).

Additional variables were also included and tested to explore the statistical significance thereof. These variables included the following:

- Revenue vehicle hours
- Revenue vehicle kilometres
- A series of population variables.

The population variables had to be removed because all the coefficients had a negative sign. This was deemed as counterintuitive and was removed from the model.

A couple of various combinations of the different variables and dummy factors were tested. The final model with the best fit only incorporated the 2 independent variables, average fare and revenue vehicle hours. This simplified model produced an R-squared value of 0.97.

The following are the benefits of the simplified model (Kohn, 2000):

- The model is easy to use with only two variables
- The level of service is included in the revenue vehicle hours
- The above mentioned incorporates the complex variables associated with urban transit systems, population, ridership level, etc.
- The sign on the average fare is negative and indicates that the increase in fares will result in the decrease in ridership
- The service hours are positive and indicate that an increase in the service hours will result in an increase in overall patronage.
- The model is statistically strong
- The residual error of the individual data points is less than with the other approaches
- Sensitivity analysis is possible with the simplified model
- No dummy variables are needed.

2.6 Alternative and Simplified Approach to Traditional Transport Models

According to McNally (2007) the traditional four step model has significant data demands. Household surveys with travel-activity diaries along with observed traffic studies are needed in order to calibrate and validate the model. In some instances the sample size of the surveyed data may be too small to ensure a high degree of confidence and may not be a true representation of the population.

Traditional models are bulky and complex and are not geared to model individual choice behaviour. The following techniques can be implemented to enhance the model and make it more suitable for choice modelling (Furnish & Wignall, 2009):

- Increase the degree of disaggregation
- Incorporate variable demand techniques
- Include dynamic functions
- Include interactive land use capabilities
- Incorporate multi-modal representation
- Establish pricing responsiveness.

Building such models is particularly challenging and requires large amounts of surveyed data and resources. A fully comprehensive conventional model is likely to be very expensive, potentially perplexing for the user and might never achieve the purpose it was designed for (Furnish & Wignall, 2009).

National scale regional models are important and conventional models are still widely used for large transport planning projects. According to Furnish et al (2009) simplified models can however be used to support and supplement the outputs of conventional models. "Simplified models do provide a rapid, flexible and cost effective testing capability" (Furnish & Wignall, 2009).

The aim of providing simplified models is to make transport models more accessible and usable. Furnish et al (2009) further argues that simplified models are also better at dealing with behavioural and pricing issues than conventional models.

Generally, the more complex the model, the more data is required. The traditional modelling approach to complex models involves many assumptions and the final model is heavy reliant on the choices of the individual modeller. In many instances dummy variables are included to compensate for the inaccurate modelled outputs and to accommodate the error and variability associated with the lack of high detailed and accurate data. These models are typically less sensitive to variation in input parameters that will affect individual choice behaviour. This is also apparent through Kohn's findings that with the inclusion of too many dummy variables, a sensitivity analysis on the actual input parameters becomes obsolete. It is also apparent from the above mentioned case studies that with the traditional approach to modelling, sometimes certain input parameters and coefficients have to be removed from the calculations because the effect thereof is counterintuitive or proved to be statistically insignificant. The effect of the variables and the statistical significance is typically determined through a complex regression analysis process which is inherently inaccurate and probabilistic.

The aim of this research project is to explore the possibility to accurately model mode choice behaviour in a South African environment with the aid of a simplified model. The proposed simplified model will reduce the active involvement of the individual modeller and little to no assumptions will be required. The approach is to include only the tangible and measurable input parameters, for which accurate data is available, into the mode choice functions. All the input parameters that are not supported by available data are considered as complex variables or intangible parameters and are included into a single station to station calibration factor. No regression analysis is required in order to determine the coefficients of the various input variables because no coefficients are implemented in the simplified model. The single station to station calibration factors are utilised to calibrate the model on a detailed level and to incorporate all the uncertainties, variations and errors into a single coefficient.

3. METHODOLOGY

This chapter provides a detailed description of the development and implementation of a simplified and robust mode choice model with the purpose of predicting the passenger demand on the Gautrain network. The proposed mode choice model was developed with the incorporation of certain data obtained from the GITMP25 regional model.

Even though the modelling steps employed in this research project were typical of the four step demand modelling approach, the principles employed differ slightly from traditional demand models. In the previous section it was mentioned that simplified models can be developed in conjunction with large traditional models in order to provide a rapid, flexible and cost effective testing capability that can deal with behavioural and pricing issues.

The Gautrain network is an 80km mass rapid transit railway system consisting of 10 stations with a North-South service from Hatfield station to Park station and an East-West service from Sandton station to Oliver Tambo International Airport (ORTIA). Figure 2 illustrates the extent of the Gautrain network.

Figure 2: Gautrain Network source: (Gautrain Management Agency, 2009)

The Gautrain commenced service in June 2010 when only the link between Sandton station and ORTIA became operational. The rest of the network, with the exception of Park station, became operational in August 2011. On 7 June 2012 the final section from Rosebank to Park station was opened.

The Gautrain service to ORTIA is excluded from this research project. The passenger demand for this service is directly related to the number of passengers arriving and departing from the airport only. People traveling to and from the airport consider different mode choice factors that are not included in the scope of this research project. These factors may include the cost of parking at the airport, the importance of being on time for a flight and the number of days between the date of departure and date of return.

The following dates are used for the purpose of this research project:

- August 2012 is used for the backward prediction scenario
- August 2013 is used as a short term future scenario testing
- February 2013 is the base year.

Each of the above mentioned scenarios are modelled for the morning peak hour (06:30 – 07:30).

The backward and forward predictions are deliberately chosen to be during the same month in order to model the patronage having similar variable external factors. The backward prediction scenario has to be after June 2012 in order for Park station to be included in the analysis. Therefore August 2012 and August 2013 are chosen for the backward and forward prediction scenarios. The base year scenario of February 2013 is in the middle of the other two scenarios with a six month backward and forward prediction period. For the purpose of this research project, all the various external factors that may influence the patronage during a specific month will be consolidated into a single factor i.e. the seasonal factor. The seasonal factor includes factors such as school holidays, special events and weather conditions.

The focus area of this research project is the development of simplified utility functions and the calibration thereof. The hypothesis is that a simplified alternative approach to the utility functions can still provide a model that is purpose built and functional. As mentioned, there is a trade-off between model complexity and model accuracy. The aim of this research project is to determine, given the reduction in model complexity, whether the model can still produce acceptably accurate results.

With all the assumptions, variations and uncertainty coinciding with the development of a mode choice model, the proposed utility functions are simplified by incorporating most of the assumptions and intangible components into station to station specific calibration factors. These factors are calibrated in order to accurately model the observed Gautrain patronage.

In addition to the calibration factors, a universal sensitivity factor is incorporated into the utility functions to calibrate the sensitivity of the model towards change in input parameters. Because the base year scenario is in February and the other two scenarios are in August, a seasonal factor is also included to compensate for the variable external factors on the patronage.

The following factors that may affect the mode choice were included in the mode choice model:

- Location of, and changes to, the Gautrain railway lines, stations and bus routes
- Changes in land use and economic growth
- Changes to road networks and Gautrain railway lines
- Fuel costs and other vehicle operating costs
- Changes to Gautrain rail, bus, and parking fares
- Gautrain service frequency
- Changes to road network delays as a result of congestion
- Transfer times between Gautrain trains and other modes.

Figure 3 shows a process flow diagram with each of the processes allocated to a coloured block, which depicts each of the four traditional steps. The derivation of the mode utilities are shown additionally.

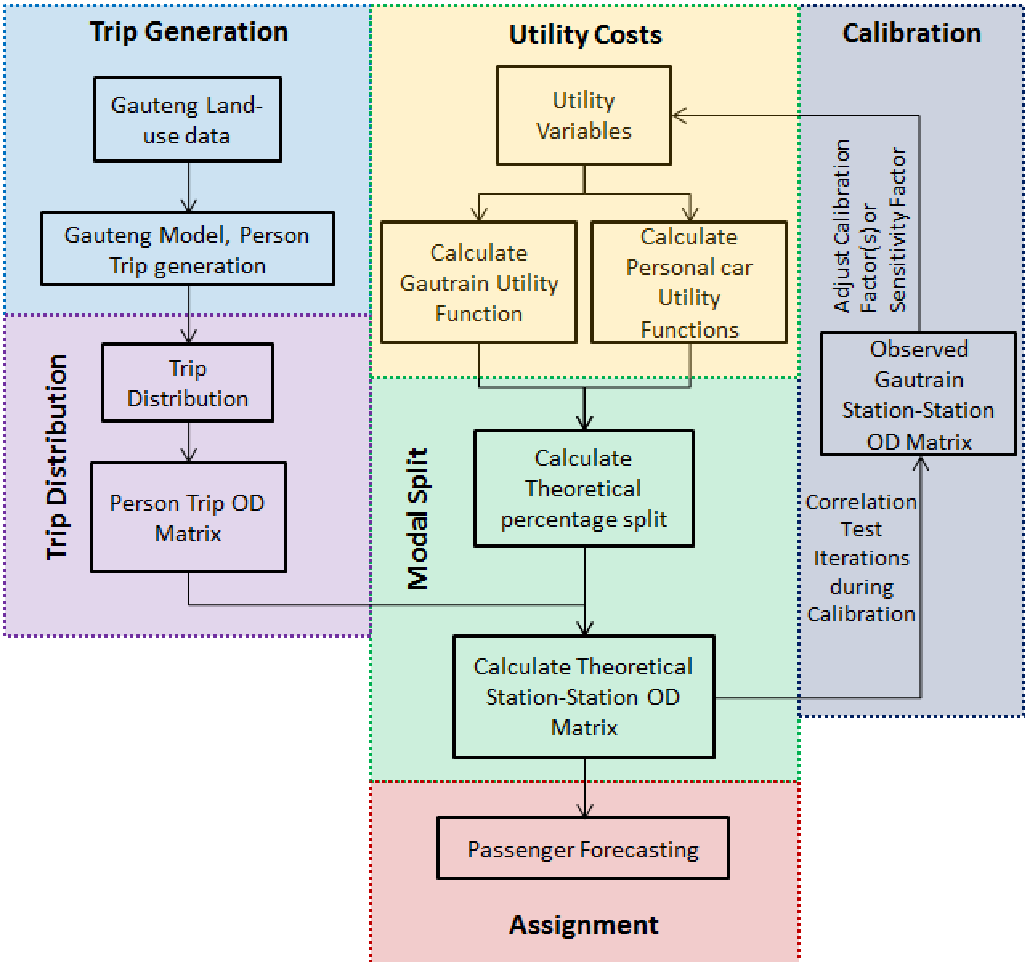


Figure 3: Methodology

In summary, the trip generation is determined from land use data (to determine the trip ends of each zone). These trips are subsequently distributed throughout the network between different OD pairs. The utility functions, in conjunction with the trip distribution rates, are used to calculate the number of users for each mode. Finally, the demand for each mode is loaded (assigned) onto the network to determine the Gautrain station to station OD patterns. Each of the above mentioned steps are discussed in more detail in the rest of this chapter.

The calibration of the model entails an iterative calibration process of the calibration factors, the sensitivity factor and the seasonal factor. The process flow diagram in Figure 4 depicts the iterative calibration process.

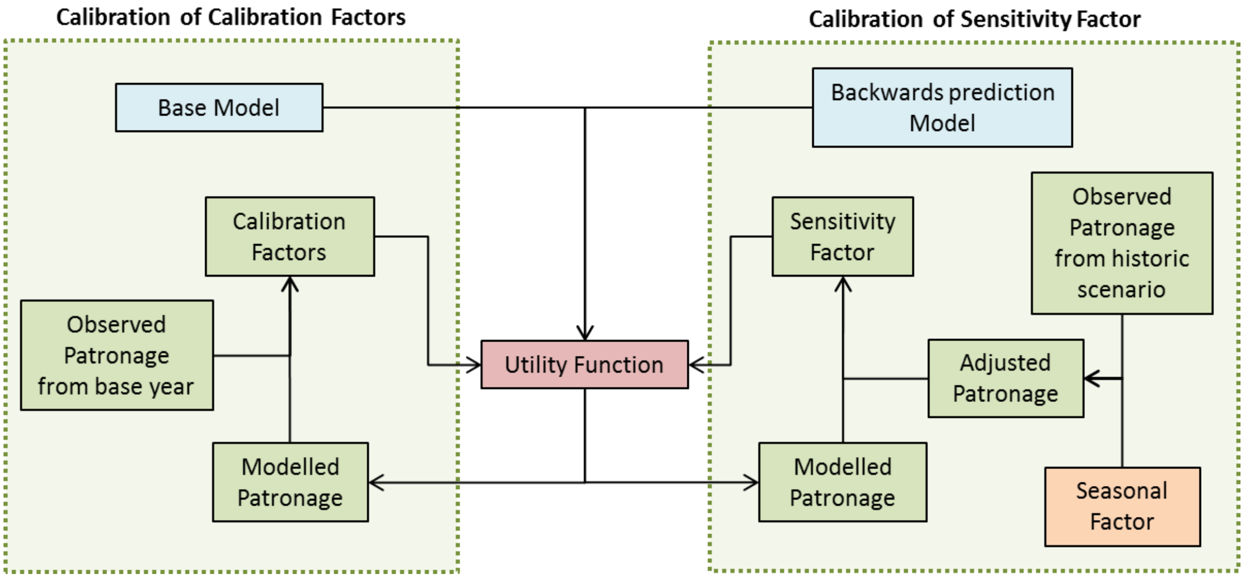


Figure 4: Iterative Calibration Process

The main focus of the calibration process is to develop a utility cost function that can accurately model the base year demand as well as the historic demand. The utility cost function is a combination of various station to station calibration factors, a network sensitivity factor and the deterministic component of the zone-to-zone utility costs. The deterministic zone-to-zone utility costs are the journey attributes expressed in Rand value. Arbitrary seed values are assigned to the calibration and sensitivity factors prior to the calibration.

In summary, the calibration factors are calibrated for the base year scenario (given a certain sensitivity factor) and the historic scenario is used to calibrate the sensitivity factor (given certain calibration factors and a seasonal adjustment factor). The historic model's observed patronage is adjusted in order to compensate for the seasonal effect, if applicable. The calibration process is an iterative process between the base model (calibration factors) and the historic model (sensitivity factor). The iterations are continued until the calibration factors and sensitivity factor converges and the change in iterations becomes negligible.

The development of the mode choice model incorporates the following key tasks which are executed in the following software packages

- EMME Software
    - Extracted a sub network from the GITMP25 network
    - Extracted trip generation rates used in the GITMP25

- PTV Visum Software
  - Imported EMME sub network
  - Modified the network to suit the project needs
  - Calculate various trip distributions
  - Calculate journey specific cost matrices
- Microsoft Excel Software
  - Develop the mode choice algorithms
  - Provide a platform to import the various journey specific utility cost matrices
  - Validate model logic
  - Calculate and calibrate the utility cost functions
  - Calculate the mode split
  - Calculate the Gautrain patronage.

The remainder of this chapter provides more detail on the development of the mode choice model with the application of modelling the Gautrain patronage.

3.1 Gautrain Model Trip Generation

The trip generation for the model is calculated from the land use data that is used in the Gauteng Integrated Transport Master Plan (GITMP25). The land use data utilised for the GITMP25 was updated in 2012, and served as input to the Gautrain Model.

The 2010 and 2012 land use data and the 2018 land use predictions for the Gauteng region was obtained and processed using the trip generation module of the GITMP25 model. The trip generation for future years incorporates the densification and the planned public transport policies and visions.

The three hour morning peak trip generation rates, as used in the GITMP25 for Gauteng, are as follows:

Table 3: 3 Hour Person Trip Generation for Gauteng Region

Year	All Income Groups	Only High Income Group
2010	4 594 492	1 608 693
2012	5 139 684	1 842 258
2018	5 855 493	2 228 357

For all income groups, the percentage growth between 2010 and 2012 is 5.77% per year and between 2012 and 2018 it is 2.2% per year.

For the high income user group the growth between 2010 and 2012 is 7% per year and between 2012 and 2018 it is 3.2% per year.

The unusually high growth rate between 2010 and 2012 should not be perceived as the actual population growth between the periods. The fact is that the 2010 data are determined from the old census data dating back to 2001, whereas the 2012 data are derived from the new census data from 2011. The two data sets (2010 and 2012) are thus not comparable and the perceived high growth rate between 2010 and 2012 is not due to an unusual growth in population between 2010 and 2012.

Given the fact that the sensitivity calibration of the Gautrain model entails doing a backward prediction, it is important to take note of the discontinuity in the data utilised in the GITMP25. If the change in ridership between the above mentioned years is based on a false over-projection of the population growth, the model will be calibrated incorrectly. The 2010 data was discarded and the trip generation for the various scenarios was calculated by interpolating between the 2012 and 2018 data.

The classification used in the GITMP25 to distinguish between the different income groups are as follows:

Table 4: Income Groups Specified by Household Income

User group based on Household income	Household income (2010 nominal value)
Low	R0 – R3 700
Medium	R3 701 – R10 700
High	R10 700 +

One of the assumptions made for this research project is that, given the relatively low bracket for the high income user group, only the high income user group (as determined by the GITMP25) would be likely to consider the Gautrain as a mode of choice. Although someone from a family with a combined family income of between R3 701 and R10 700 (2010 nominal value) will consider using the Gautrain on the odd occasion, they would most probably not commute on the Gautrain regularly.

Therefore, only the high income user group is used as the basis to determine the various trip generation rates for the various time periods in this research project.

The GITMP25 only specifies the trip generation rates for the specified years and not for the various months in each year. Because February 2013 is used as the base scenario, it is assumed that the annual trip generation rates obtained from the GITMP25 are for the month of February, and the monthly trip generation rates will be calculated accordingly.

The trip generation for 2012 and 2018 were extracted from the GITMP25 model. The trip distribution was done (as described in section 3.2) in order to determine the OD matrices for February 2012 and February 2018. Each OD pair within the various scenarios' OD matrices were obtained by interpolating each OD pair individually. The trip generation rates for the various scenarios are in turn calculated from the interpolated OD matrices.

The monthly growth "(r_{ij})" for each OD pair with origin "i" and destination "j", is calculated as follows:

$$r_{ij} = \left(\frac{OD2018_{ij}}{OD2012_{ij}}\right)^{\frac{1}{n}} - 1$$

Where n is the number of months between February 2012 and February 2018.

Each OD pair's growth rate is calculated separately and is independent from each other. The annual average calculated growth comes to 3.13% per annum.

Table 5 contains the calculated total number of trips for the various scenarios.

Table 5: Morning Peak Hour Trip Generation Rates for the Gautrain Model

	Hourly trip generation rates for High income users
GITMP25 Rates from Land Use Data	
∑ February 2012	483 273
∑ February 2018	581 327
Calculated Monthly Rates for the Following Scenarios	
∑ August 2012	486 527
∑ February 2013	490 346
∑ August 2013	494 776

3.2 Gautrain Model Trip Distribution

Developing a trip distribution pattern is renowned to be challenging and many assumptions need to be made. It is particularly difficult to develop an accurate trip distribution function in a South African environment. This is largely due to the laws of segregation during the apartheid years when people were forced to live in certain areas, regardless of where the job opportunities were. This forced population distribution ensures that the trip distribution for the lower income groups do not necessarily adhere to the principle of trip distribution based on the gravity model. The segregation of people in conjunction with urban sprawl and big class differences ensures that there is no general trip distribution pattern in the Gauteng region. At first the trip distribution from the GITMP25 model was used, but it proved not suitable for this research project and caused some unexplained anomalies with the modelling of Gautrain patronage.

It was thus decided to develop a new generic trip distribution pattern for the purpose of this research project. Household survey data from the technical report on trip distribution for the development of the Gauteng model (GITMP25) was used to establish the desired commute travel distance for travellers in Gauteng. Figure 5 illustrates the processed results from the household surveys.

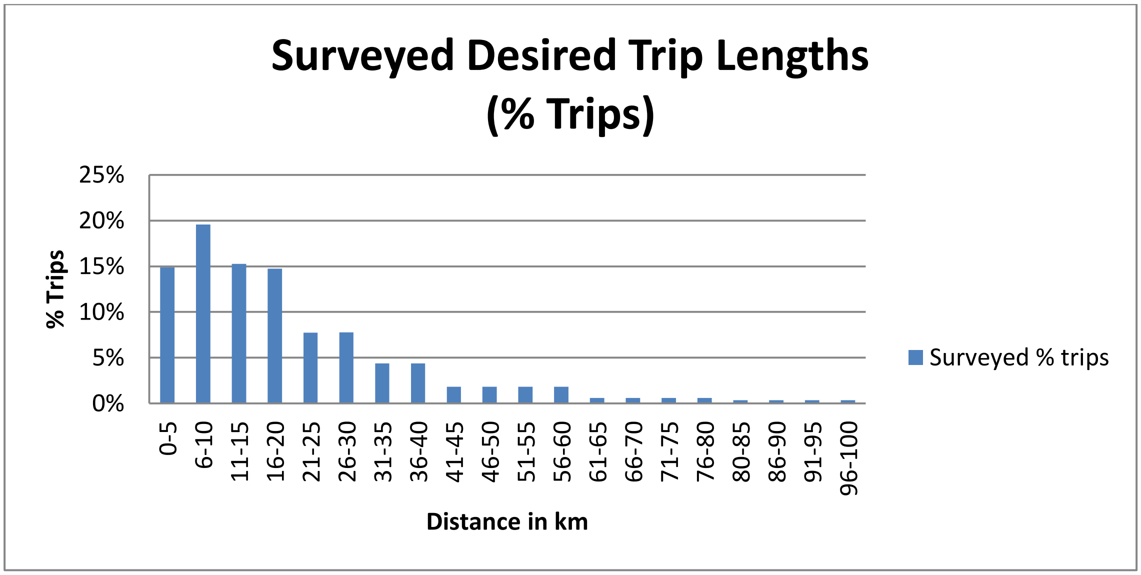


Figure 5: Surveyed Distance to Work

A combination deterrence function was used to fit a synthetic gravity-type model to the above mentioned surveyed data. The strength of the attraction between two zones is calculated using the following formula:

$$f(s) = a \times s^b \times e^{(c \times s)}$$

Where *s* is the journey distance and a, b and c are constants.

The trips from each origin, and trips to each destination, are subsequently distributed according to the relative attraction between the origin and destination zones.

The values for the constants a, b, and c in the above probability function were chosen so that the trip length distribution (from the model output) matches the trip length data from the travel surveys conducted for the Gauteng model (from the technical report on trip distribution for the development of the Gauteng model). The constant values that produced the best fit to the graph, determined by the least absolute deviation method, are as follows:

a = 0.1405

b = 0.228

c = 0.0603.

Figure 6 and Figure 7 show the calculated probability function compared to the surveyed data.

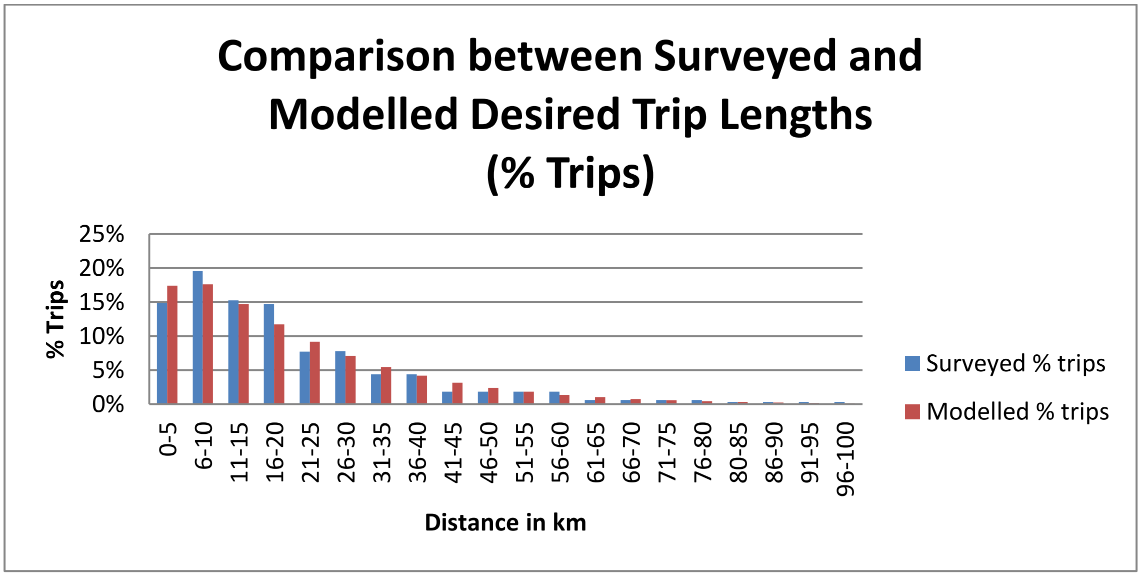


Figure 6: Trip length distribution

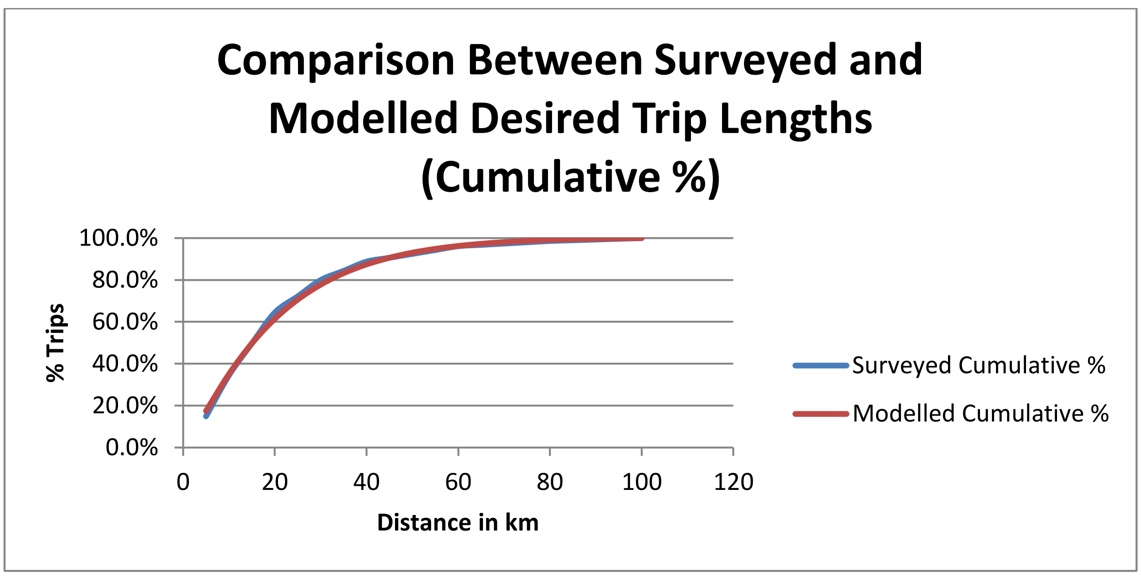


Figure 7: Trip length cumulative distribution

The above mentioned combination probability function, in conjunction with the trip generation rates, was used to calculate the various OD matrices. Given the coarse zone structure and high level road layout in the Gauteng model, the direct distance between the zone's centroids was defined as the journey distance (s) between the zones.

These OD matrices define the trip distribution patterns for the various years and serve as the basis on which the mode split will determine the Gautrain patronage.

3.3 Gautrain Model Mode Split

The mode split for the Gautrain model was done after the trip distribution step and hence the journey characteristics are taken into consideration. The negative effect of having aggregated individual characteristics by performing the model split after the trip distribution is mitigated because only the high income user group is used during this research project. Using only a single user group ensures that a more homogeneous population is used to determine the modal split. Most of the people in the high income user class will have access to a private vehicle and therefore will have a choice between utilising public or private transport. The assumption on the value of time is also more applicable to a specific user group if the group is more homogeneous.

The mode split for each OD pair is calculated by determining the probability mode share for each mode, based on the utility function between the two zones. The utility calculations are used in order to create objective measured values for the subjective choice making processes.

3.3.1 Discrete choice analysis

A constant utility approach combined with a Logit probability function are utilised in determining the mode split probabilities for each OD pair. For the purpose of this research project, the only alternatives are to use either a private vehicle to complete the entire journey, or use the Gautrain in conjunction with the various access and egress modes.

The general formula for a Binomial Logit discrete probability function is as follows:

$$P(i) = \frac{e^{V_i}}{e^{V_j} + e^{V_i}}$$

Where:

$\mathrm{P}(i)$ is the probability of choosing alternative i

V_i and V_j are the deterministic/systematic components of the utility functions for alternatives i and j.

By using the above formula, the following formulas are derived for the purpose of this research project.

The probability "P_{public}" of using the Gautrain and probability "P_{car}" of using one's private vehicle is:

$$[P_{public}] = \frac{e^{Up}}{e^{Up} + e^{Uc}}$$

$$[P_{car}] \quad = \frac{e^{Uc}}{e^{Up} + e^{Uc}}$$

Where Up and Uc are the utility functions for the Gautrain and private vehicle journeys, respectively.

3.3.2 Development of utility functions

A pre-requisite for the modal split calculations is the determination of the utility of the trip for each mode and each OD pair in the model. The utility cost for any particular journey is an expression of the perceived costs to a traveller.

As mentioned, the approach of this research project is to simplify the utility functions by reducing the number of assumptions and coefficients typically implemented. All the journey attributes are expressed as a monetary value with their relative significance equal to each other.

The following attributes are incorporated into the utility functions:

- Journey lengths in distance and time
- Value of time
- Vehicle operation cost
- Gautrain ticket pricing
- Gautrain bus fares

- Parking cost at stations
- Train frequency
- Additional costs incurred, such as prolonged travel times due to congestion.

The factors affecting the utility costs are different for private vehicle trips compared to public transport trips and therefore each has its own formula for calculating the utilities:

- The utility cost "(U_p)" for a public transport journey is calculated as follows:

$$[U_p] = CF(VOT_p[P_{tt}] + [P_{JC}])SF$$

Where:

CF is a Calibration factor

VOT_p is the Value of time for public transport users

P_{tt} is the Travel time between origin and destination zone for public transport

P_{jc} is the Journey cost between origin and destination zone for public transport

SF is a Sensitivity factor.

- The utility cost "(U_c)" for a private vehicle journey is calculated as follows:

$$[U_c] = (VOT_c[C_{tt}] + VOC_c[C_{td}])SF$$

Where:

VOT_c is the Value of time for private vehicle users

C_{tt} is the Travel time between origin and destination zone for a private vehicle

VOC_c is the Vehicle operating cost for a private vehicle

C_{td} is the Travel distance between origin and destination zone for a private vehicle

SF is a Sensitivity factor.

The travel times and travel distances for a private vehicle journey are only dependent on the attributes of a single mode i.e. the private vehicle. The travel times and distances for a public transport journey are however a combination of the various modes utilised in order to complete the journey. The calculations of the various cost factors for the public transport journeys are as follows:

- The Travel Time (TT) for public transport journey $[P_{tt}]$:

$$
\begin{aligned}
[P_{tt}] \; = \; & \mathrm{TT}_{\text{(from Origin Zone to Origin Station)}} \\
& + \mathrm{TT}_{\text{(between Origin and Destination Gautrain Stations)}} \\
& + \mathrm{TT}_{\text{(from Destination station to destination zone)}} \\
& + \text{Average wait time} \\
& + \text{Transfer time}
\end{aligned}
$$

- Travel time between OD zones and Gautrain stations is a combination of public and private transport travel times:

$$\mathrm{TT}_{(Between\ zone\ and\ station)} = \left[\left(P_{Parking}\right)C_{tt} + 2\left(P_{Drop\&go}\right)C_{tt} + P_{Bus}(P_{tt})\right]$$

$P_{Parking}$ is the proportion of trips that use parking at a Gautrain station

C_{tt} is the travel time for a private vehicle between OD zones and Gautrain stations

$P_{Drop\&go}$ is the proportion of trips that are expected to make use of the drop and go facility at the Gautrain stations

P_{Bus} is the proportion of trips that use the Gautrain bus service

P_{tt} is the travel time with public transport.

- The journey cost (JC) between OD zones and Gautrain stations is a combination of public and private transport costs (TC):

$$
\begin{aligned}
[P_{JC}] \; = \; & \mathrm{TC}_{\text{(from Origin Zone to Origin Station)}} \\
& + \mathrm{TC}_{\text{(between OD Gautrain Stations)}} \\
& + \mathrm{TC}_{\text{(from Destination station to destination zone)}} \\
& + \text{Parking}
\end{aligned}
$$

- Travel cost between OD zones and Gautrain stations is a combination of public and private transport travel costs:

$$\mathrm{TC} = \left[\left(P_{Parking}\right)C_{td}VOC_{car} + 2\left(P_{Drop\&Go}\right)C_{td}VOC_{car} + P_{Bus}BusFare\right]$$

$$Parking = (P_{Parking})ParkingCost$$

$P_{Parking}$ is the proportion of trips that use parking at a Gautrain station

C_{td} is the travel distance between OD zones and Gautrain stations

VOC_{car} is the vehicle operating cost for a private vehicle

$P_{Drop\&go}$ is the proportion of trips making use of drop and go facility at the Gautrain stations

P_{Bus} is the proportion of trips making use of the Gautrain bus service

$BusFare$ is the bus fare to and from Gautrain stations

$ParkingCost$ is the cost of parking at the Gautrain stations.

3.4 Gautrain model Assignment

Because the main focus of this research project is to determine the patronage of the Gautrain, the model is only calibrated against the Gautrain patronage. This has a major advantage in keeping the model simple. Given the smartcard tap-in/tap-out electronic fare collection system implemented for the Gautrain, detailed station to station OD patronage is available. The main focus is to split the current network demand OD matrix in such a way as to produce the desired (observed) patronage. What, where and how the other private trips are completed are not incorporated. The following (not limited to) typical network calibration processes are therefore excluded:

- Road network calibration
- Iterative dynamic route assignments
- Road link impedance calibration
- Correlation tests on the observed road link volumes and modelled volumes
- Regression analysis.

Each Gautrain station to station OD pair has its own calibration constant and therefore the model can be calibrated to exactly replicate the base year patronage. Backward predictions are done in order to calibrate the sensitivity of the model. The network sensitivity factor is altered and the calibration and assignment process is repeated. This process is explained in more detail in the following section.

3.5 Calibration and validation of the model

In order to calibrate the model, the following factors are included in the utility cost functions:

- Calibration Factors
- Sensitivity Factor
- Seasonal Factor.

The utility functions as mentioned in section 3.3.2 are as follows:

$$[U_p] = CF\big(VOT_p[P_{tt}] + [P_{JC}]\big)SF$$
$$[U_c] = (VOT_c[C_{tt}] + VOC_c[C_{td}])SF$$

Where the probability functions are:

$$\left[P_{public}\right] = \frac{e^{Up}}{e^{Up} + e^{Uc}}$$
$$[P_{car}] \quad = \frac{e^{Uc}}{e^{Up} + e^{Uc}}$$

The calibration and sensitivity factors each have a different function and effect on the probability function. The probability function utilised for this research project is a Logit probability function. The calibration factors are only multiplied with the public transport's utility functions and the sensitivity factor is multiplied to both utility functions. Figure 8 illustrates the principle of each factor and also indicates the impact thereof.

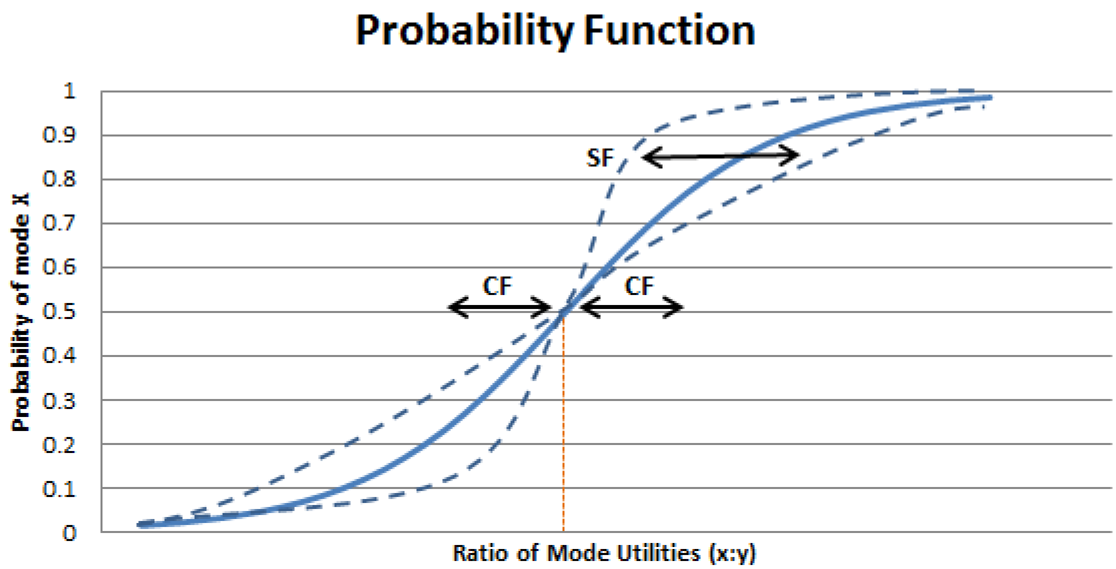


Figure 8: Probability Function

Because the calibration factors are only applied to one mode's utility functions, it will alter the ratio between the two utilities and therefore tend to shift the graph (in Figure 8 above) to the left or right. The sensitivity factor is multiplied to both of the modes' utilities and because of the exponential term will therefore change the shape of the graph, making it steeper or flatter. A steep curve implies that a small difference between the modes' utility will cause a larger modal shift, whereas a flatter curve implies that a smaller modal shift will result from the same imbalance in the utilities.

The Base year scenario is for the month of February 2013, whereas the other two scenarios are for the months of August 2012 and 2013. With the incorporation of a seasonal factor, the model can be calibrated for any month of the year (also see section 4.8). The weather conditions in the Gauteng region are more favourable for public transport during February than in August. The following table provides some seasonal attributes for the City of Johannesburg that indicates the difference in weather conditions for August and February. Source: (Time and Date AS, 1997) and (Canty and Associates LLC, 1999).

Table 6: Seasonal attributes for Johannesburg

Seasonal Attributes for Johannesburg City		
Attributes	**February**	**August**
Time of Sunrise the first day of the month	05:43	06:47
Time of Sunset the first day of the month	19:00	17:42
Average monthly temperature	20°C	13°C
Average low temperature	15°C	7°C
Average wind speed	14 km/h	16 km/h

The change in the observed patronage is not only due to the change in the input parameters, but also includes seasonal variation. The seasonal variation is not limited to weather conditions and may also include factors such as the reduction in transport demand during the end of year festive season and other school holidays. Stated differently, if all the input parameters for the various years stayed unchanged, there would have been a certain amount of variation in the observed patronage. The method of the model is such that change in the patronage must be a consequence of the alterations in the input parameters. The seasonal factor is not included in the proposed utility functions and is a factor separate from the mode choice process. After the theoretical patronage is determined, the total patronage is adjusted with a seasonal factor. Calibrating the model whilst the modelled patronage is not a function of seasonal variation will render the model predictions inaccurate. During the sensitivity calibration phase, the modelled data must be adjusted to incorporate the assumed seasonal factor.

Determining the actual seasonal effect is impossible, as the true magnitude of the seasonal effect is hidden in the normal change in patronage. There is no method to establish what proportion of the growth or decline in patronage is due to the change in input parameters or the effect of various seasonal factors. During this research project, some attempts and assumptions are however made to support the assumed seasonal factors.

4. DEVELOPMENT OF THE GAUTRAIN MODEL

This chapter provides more detail on the development, validation and calibration of the proposed mode choice model. The following steps are detailed in this section:

- Exporting the EMME model into PTV
- Customising the PTV model
- Assigning zones to stations
- Calculating various input matrices
- Calculate input parameters for utility functions
- Validation of the model's logic
- Iterative calibration process
- Model independent seasonal factor validation.

4.1 Exporting EMME model into PTV

The GITMP25 year EMME model was utilised to extract the network layout, zone structure and trip generation rates as used in the GITMP25 study.

The Gauteng EMME model consists of 903 zones, and covers the majority of the Gauteng province. The PTV software licence available has a capacity of 600 zones, thus some of the zones falling on the outskirts of the model had to be removed in order to accommodate the model in the PTV environment. The study area was determined visually in order to incorporate as many zones as possible into the model, where the use of the Gautrain might be an alternative to the private vehicle. The zones that were removed proved to be far away from the Gautrain stations because it is highly unlikely that someone with an origin or destination in any of these zones would utilise the Gautrain. Therefore these zones can be eliminated from the research area with high confidence. Figure 9 illustrates the extent of the EMME model and the study area that was extracted from the EMME model.

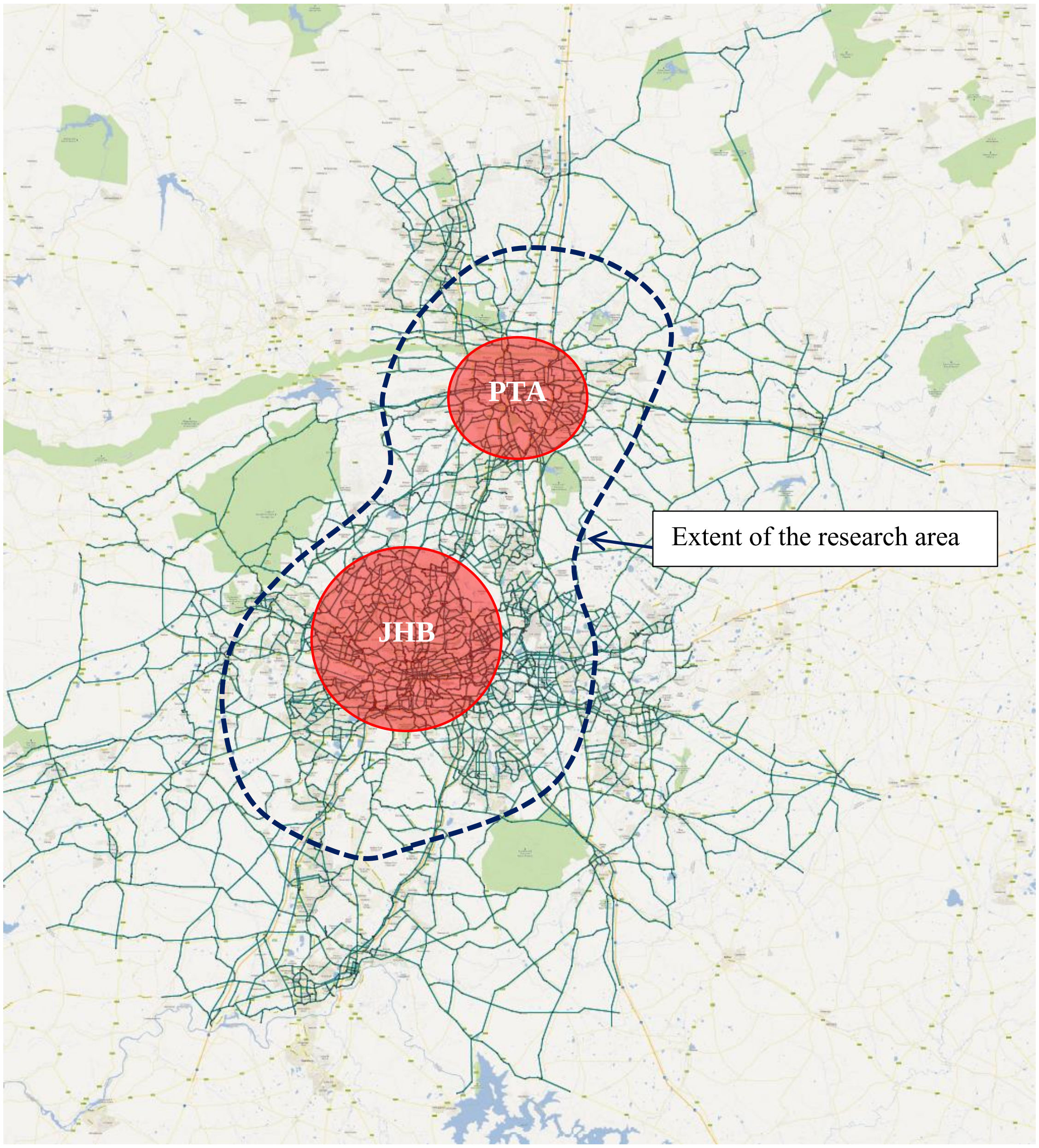


Figure 9: Extent of the EMME network

4.2 Customising the PTV model

The model was enhanced in the PTV environment in order to customise the model and make it more applicable for the particular research project. The following alterations were done to the model:

1. Insert additional zones representing the Gautrain stations
2. Connect Gautrain stations with Gautrain rail links

3. Code the bus routes on the links
4. Add walk connectors from the zones to the bus routes
5. Assigning zones to stations (see section 4.3 for more detail).

4.3 Assigning zones to stations

Each zone in the network is assigned with an attribute that indicates which station is the closest to that particular zone. The station allocation is done taking the direct distance to the various stations into consideration. Figure 10 indicates the boundaries of the probability lines over the network.

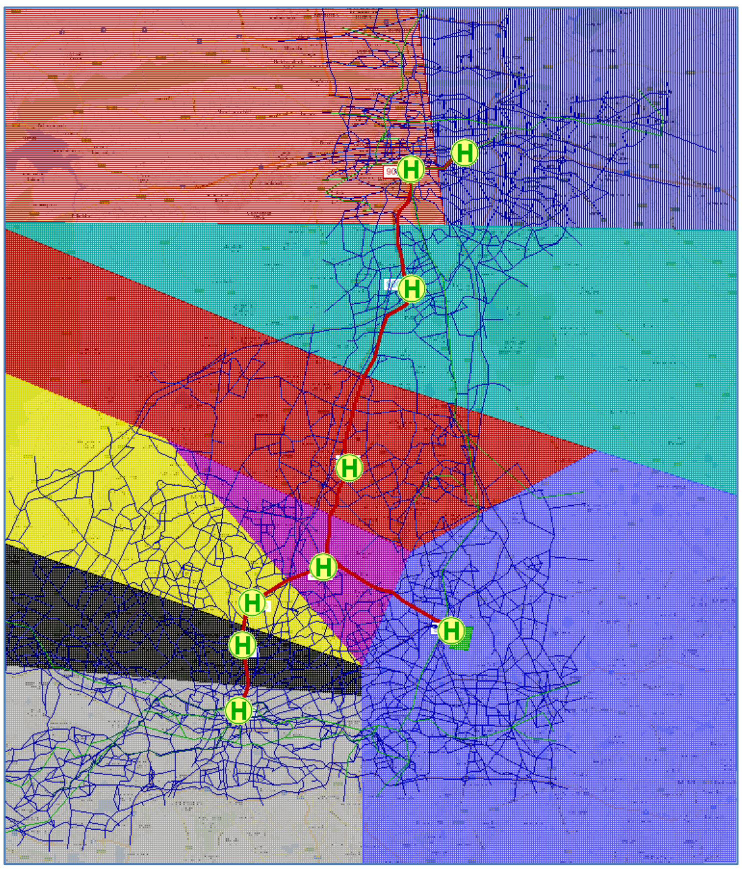


Figure 10: Gautrain Station Probability Areas

It is important to note that the probability of utilising the Gautrain decreases drastically as the distances to the nearest station increase. The effect of incorrectly assigning a zone to a station decreases as the distance from the zone to the closest station increases. This allows for a very coarse allocation of zones to stations. The assignment was done visually by incorporating the direct distance from the network zones to Gautrain stations. The zone centroids were used as a reference point to determine the closest station to that zone. It was ensured that the zones in close proximity to a Gautrain bus route are assigned to that particular station.

The purpose of assigning each zone to a station is necessary in order to calculate the trip travel time and travel cost for the journeys to and from the Gautrain stations.

Any journey that utilises the Gautrain can be broken down into three segments:

1. The journey from the origin zone to the nearest Gautrain station (access journey)
2. The journey on the Gautrain from the station closest to the origin zone to the station closest to the destination zone
3. The journey from the Gautrain station closest to the destination zone to the destination zone (egress journey)

Theoretically these access and egress journeys can be a combination of various modes of transport including the following:

- Walking
- Cycling
- Gautrain buses
- Metro buses
- Private vehicle (park & ride as well as being dropped off or picked up at the station).

4.4 Calculating various input matrices

The zone numbering structure for the model is such that the network zones are numbered from 1 to 500 and the Gautrain stations are numbered from 501 to 510. Figure 11 illustrates the general layout of all the matrices used in this research project. Area “A” contains all the direct network zone to zone values, area “B” contains all the access journey values, area “C” contains all the egress journey values and area “D” contains all the station to station values.

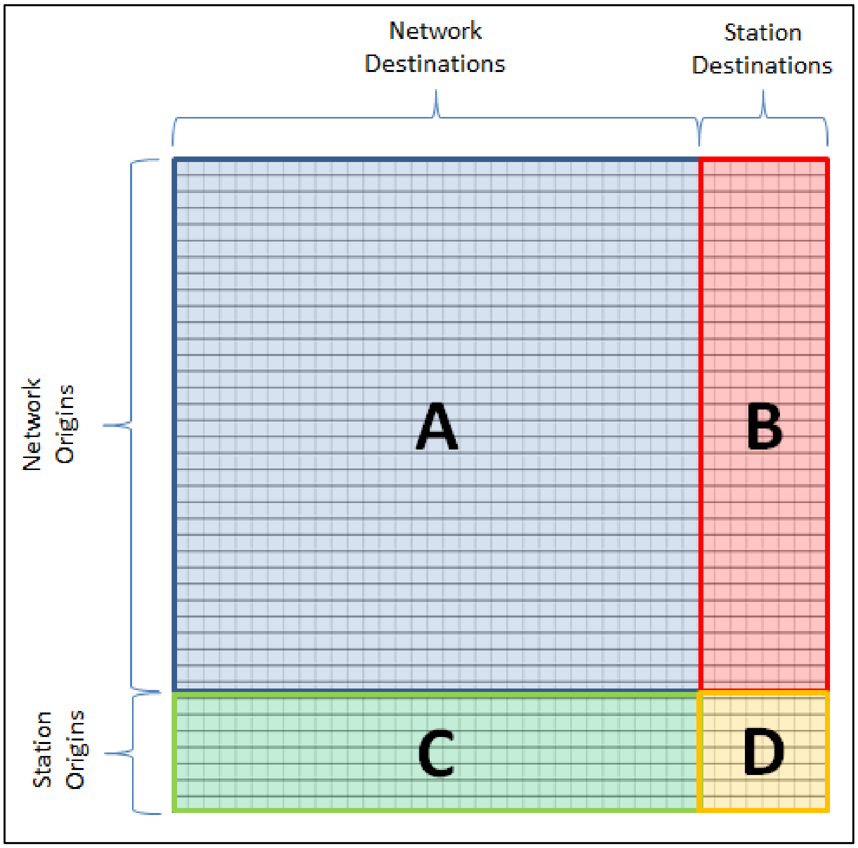


Figure 11: Matrix Structure

The approach was to develop a model in PTV Visum in order to determine the various trip characteristic matrices. These matrices will serve as inputs in the calculations to determine the utility function between each OD pair. The matrices calculated for this research project are -

1. Trip distribution OD matrices
2. Private vehicle journey distances
3. Private vehicle journey times
4. Public transport journey times
5. Public transport journey distances.

Each of these matrices are calculated in PTV and then exported to Excel. Note that the above mentioned matrices will include the values for the access, egress and station to station journeys. Provided that no major network alterations were done during the analysis period, the trip distribution matrices are the only input matrices that change between the various scenarios.

4.5 Calculating input parameters for the utility functions

The scenario specific input parameters are the input parameters that change between the various scenarios. These parameters contribute to the change in the mode choice between the scenarios.

The following input parameters are required as inputs to the utility functions:

1. Scenario information
2. Peripheral public transport input parameters
   - Bus fare
   - Gautrain frequency and transfers times
   - Parking cost at the Gautrain station
3. Station specific input parameters
   - Gautrain station to station fares and travel times
   - Access and egress mode split percentages
4. Private vehicle input parameters
   - Vehicle operating cost of a private vehicle
5. Value of time for public transport and private vehicle users.

4.5.1 Scenario information

All the costs for the various input parameters have to be adjusted to the nominal rand value of the base year scenario. The specified dates of the scenarios are used to calculate the number of months between the various scenarios. The number of months in conjunction with a consumer price index (CPIX) of 6% per annum is used to calculate the February 2013 nominal values of all the input parameters.

4.5.2 Peripheral public transport input parameters

The peripheral public transport input parameters relate to the input parameters for the access and egress journeys.

Table 7 provides a summary of the peripheral public transport input parameters.

Table 7: Summary of the Public Transport Input Parameters

	Historic Scenario August 2012 Real Value	Base Scenario February 2013 Real Value	Future Scenario August 2013 Real Value
Bus Fare (R)	R 6.00	R 6.00	R 6.00
Parking (R)	R 10.00	R 12.00	R 15.00
Gautrain Frequency (min)	12	12	12
Transfer Time (min)	5	5	5

The rest of this section provides more detail on the parameters depicted in Table 7.

4.5.2.1 Gautrain bus fare

The costs of utilising the Gautrain bus service is a flat fare and is independent of the journey length. The cost of a single bus journey for all three scenarios is R6.

4.5.2.2 Parking cost at the Gautrain stations

The Gautrain parking cost structure consists of a flat fare for parking between 15min to 24 hours. It is assumed that the majority of the people utilising the Gautrain during the morning peak does so for the purpose of commuting to work and will therefore return within 24 hours.

4.5.2.3 Gautrain frequency and transfer times

The Gautrain frequency is the time intervals between consecutive trains during the morning peak period. The Gautrain frequency for all three scenarios is 12min. The transfer time is the average time it takes to move between the various modes. The assumed average transfer time is 5min.

4.5.3 Station specific input parameters

The station specific input parameters are divided into station to station specific parameters and single station parameters. The station to station parameters include the travel times and fares between stations. The single station specific parameters include the mode split percentages for the access and egress journeys.

4.5.3.1 Gautrain station to station fares and travel times

The ticket prices and time tables for each station to station journey were obtained from the Gautrain website (Gautrain Management Agency, 2009). The time tables provided are used to calculate the journey times between the various stations. Table 8 and Table 9 depict the ticket prices for the base year scenario and journey times respectively.

Table 8: Ticket Price for Gautrain Journey February 2013.
Source: (Gautrain Management Agency, 2009)

Ticket Price for Gautrain Journey February 2013									
	Hatfield	**Pretoria**	**Centurion**	**Midrand**	**Marlboro**	**Sandton**	**Rosebank**	**Park**	**Rhodesfield**
Hatfield	R 0	R 20	R 25	R 37	R 43	R 46	R 49	R 52	R 49
Pretoria	R 20	R 0	R 23	R 31	R 40	R 43	R 46	R 49	R 46
Centurion	R 25	R 23	R 0	R 25	R 31	R 38	R 40	R 43	R 41
Midrand	R 37	R 31	R 25	R 0	R 23	R 25	R 28	R 31	R 29
Marlboro	R 43	R 40	R 31	R 23	R 0	R 20	R 22	R 25	R 23
Sandton	R 46	R 43	R 38	R 25	R 20	R 0	R 20	R 22	R 27
Rosebank	R 49	R 46	R 40	R 28	R 22	R 20	R 0	R 20	R 29
Park	R 52	R 49	R 43	R 31	R 25	R 22	R 20	R 0	R 31
Rhodesfield	R 49	R 46	R 41	R 29	R 23	R 27	R 29	R 31	R 0

Table 9: Travel Time for Gautrain Journey February 2013, calculated from time table.
Source: (Gautrain Management Agency, 2009)

Travel Time (minutes) for Gautrain Journey February 2013									
	Hatfield	**Pretoria**	**Centurion**	**Midrand**	**Marlboro**	**Sandton**	**Rosebank**	**Park**	**Rhodesfield**
Hatfield	0	7	14	23	29	34	38	42	37
Pretoria	7	0	7	16	22	27	31	35	30
Centurion	14	7	0	9	15	20	24	28	23
Midrand	23	16	9	0	6	11	15	19	14
Marlboro	29	22	15	6	0	5	9	13	8
Sandton	34	27	20	11	5	0	4	8	12
Rosebank	38	31	24	15	9	4	0	4	16
Park	42	35	28	19	13	8	4	0	20
Rhodesfield	37	30	23	14	8	12	16	20	0

4.5.3.2 Access and egress mode split percentages

For this research project it is assumed that only the Gautrain buses and private vehicles are used for the access and egress journeys. Current data indicates that there is very little intermodal usage between the metro buses and the Gautrain.

Figure 12 illustrates the various modes considered for the various segments of a journey completed with the Gautrain.

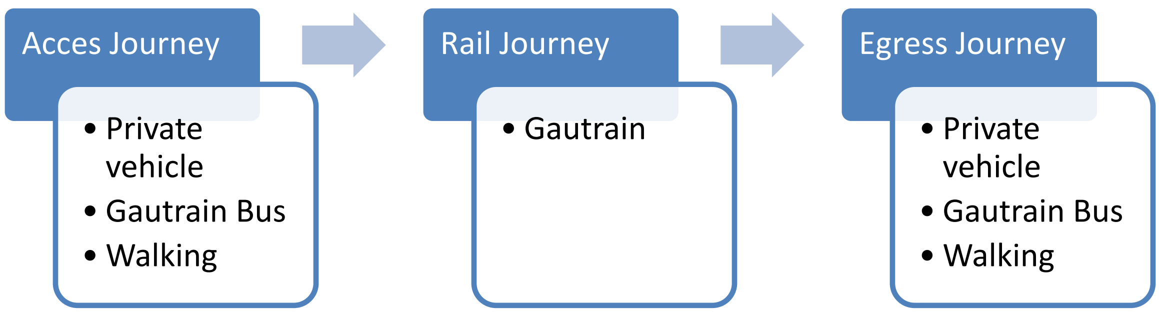


Figure 12: 3 Segments of the Gautrain Journey

Given the implementation of an integrated EFC system, the percentage splits between the different modes are calculated from data obtained from the Gautrain buses and the park & ride facilities. The same card is used to pay for the bus journey, parking and Gautrain fare. It is thus possible to determine the origin, destination, access mode and egress mode for each individual traveling on the Gautrain.

The access and egress percentage splits are calculated from the high detailed information available on every journey. The data available can indicate if a person utilised the parking facility or the bus service during their access or egress journeys. It is assumed that the passengers that did not utilise any of the above mentioned services, where dropped off or picked up at the station. The number of passengers walking to the stations is incorporated within the drop and go passengers. The passengers walking from the stations are incorporated within the pickup and go passengers.

Table 10 illustrates the percentage splits during the morning peak period for the base year scenario.

Table 10: Percentage Mode Split for Access and Egress Journeys

Station Specific Percentage Mode Split						
	Access Journeys			Egress Journey		
	Drop-&-Go	Parking	Bus	Pickup-&-Go	Park	Bus
Hatfield	50%	40%	10%	59%	2%	39%
Pretoria	59%	28%	13%	45%	2%	53%
Centurion	46%	45%	9%	66%	1%	33%
Midrand	56%	29%	15%	22%	1%	77%
Marlboro	80%	20%	0%	94%	6%	0%
Sandton	59%	16%	25%	68%	2%	30%
Rosebank	60%	31%	9%	61%	3%	36%
Park	77%	19%	4%	59%	1%	40%
Rhodesfield	46%	46%	7%	74%	7%	20%

4.5.4 Private vehicle input parameters

The vehicle operational cost is a function of the fuel price, fuel efficiency and wear and tear on the vehicle.

The price of fuel is probably the biggest factor that influences passenger mode choice. It is a direct out of pocket cost to the user, and with the frequent increases in the fuel price, drivers are increasingly conscious of changes in the fuel price.

The fuel factor is a factor used by the Automobile Association of South Africa (AA) to convert the fuel price to a cent per kilometre value. This factor is an indication of the average fuel efficiency of vehicles. For the purpose of this research project the average vehicle is considered to have an engine capacity of between 1500cc and 1800cc.

The wear on the vehicle is a factor used by the AA to incorporate the service cost and tyre wear into the vehicle's operational cost.

The calculated vehicle operational costs for the various scenarios are listed in Table 11 (Automobile Association of South Africa).

Table 11: Private Vehicle Operational Cost

	Historic Scenario August 2012 Real Value	Base Scenario February 2013 Real Value	Future Scenario August 2013 Real Value
Private Vehicle operational cost R/km	R 1.27	R 1.36	R 1.45

4.5.5 Value of time

For the purpose of this research project the nominal value of time is assumed to be R60/hour and is derived from the value of time for high income users utilised in the GITMP25.

4.6 Validation of the models logic

The general working and logic of the model is validated by running the model for the base year (February 2013) scenario whilst incorporating a universal calibration constant. A couple of sensitivity factors are assumed for the purpose of this exercise. The differences between the observed and modelled patronage (delta patronage) are calculated by subtracting the observed patronage from the modelled patronage. The calibration factors and sensitivity factor both influence the mode split probability function (see section 3.5) and therefore a couple of sensitivity factors are assumed. Figure 13 illustrates the effect of the universal calibration factor on the delta patronage. The minimum number of modelled passengers is zero and the total observed patronage for the base year scenario is 7 558 passengers. The delta patronage is calculated by subtracting the observed patronage from the modelled patronage and therefore there is a minimum asymptote at -7 558 passengers.

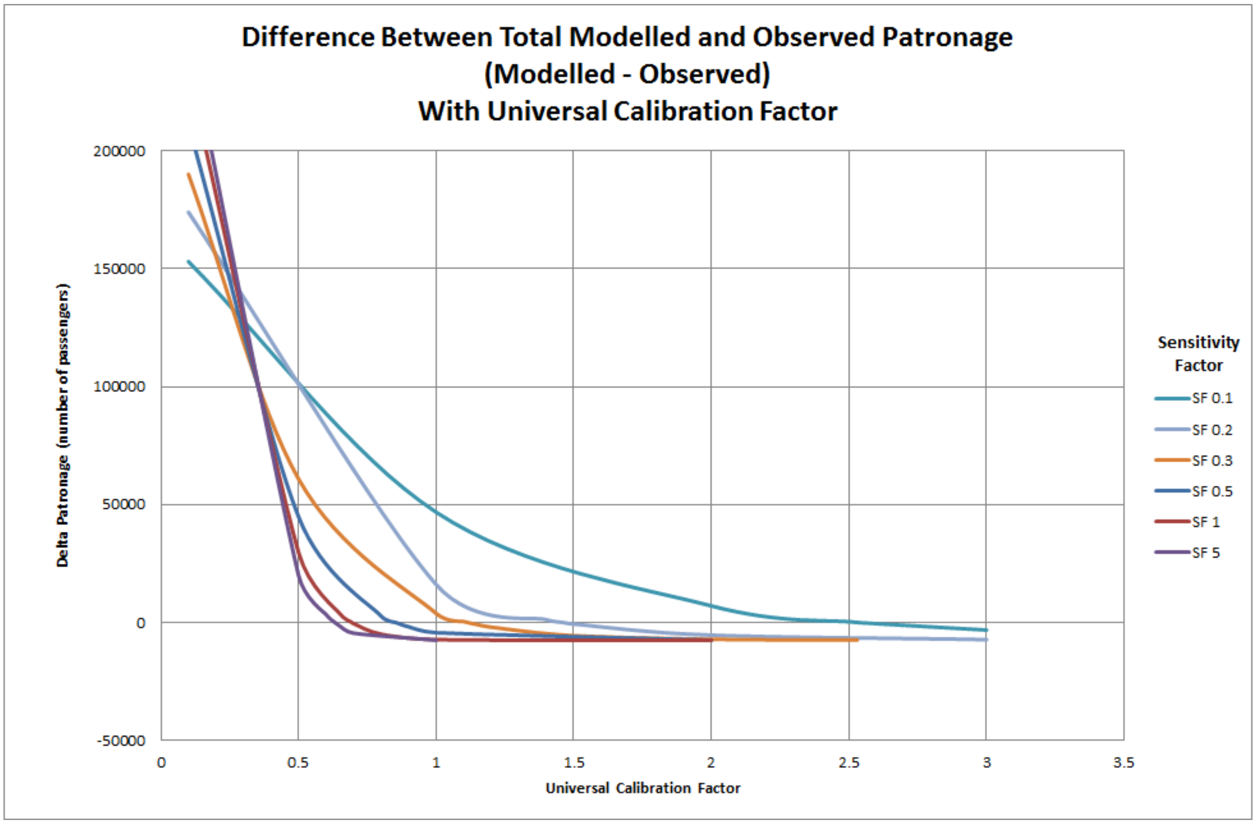


Figure 13: Delta Patronage with Universal Calibration Factor

The points at which the graphs intercept with the zero y-axis (delta patronage of zero) indicate the point at which the total observed patronage is equal to the total modelled patronage. Figure 14 provides a more detailed view (of Figure 13) around the zero delta patronage axis. Note that the gradients of the functions provide a good indication of the sensitivity of the functions towards the change in calibration factors. A smaller network sensitivity factor renders the model less sensitive towards change in the calibration factor than a larger sensitivity factor.

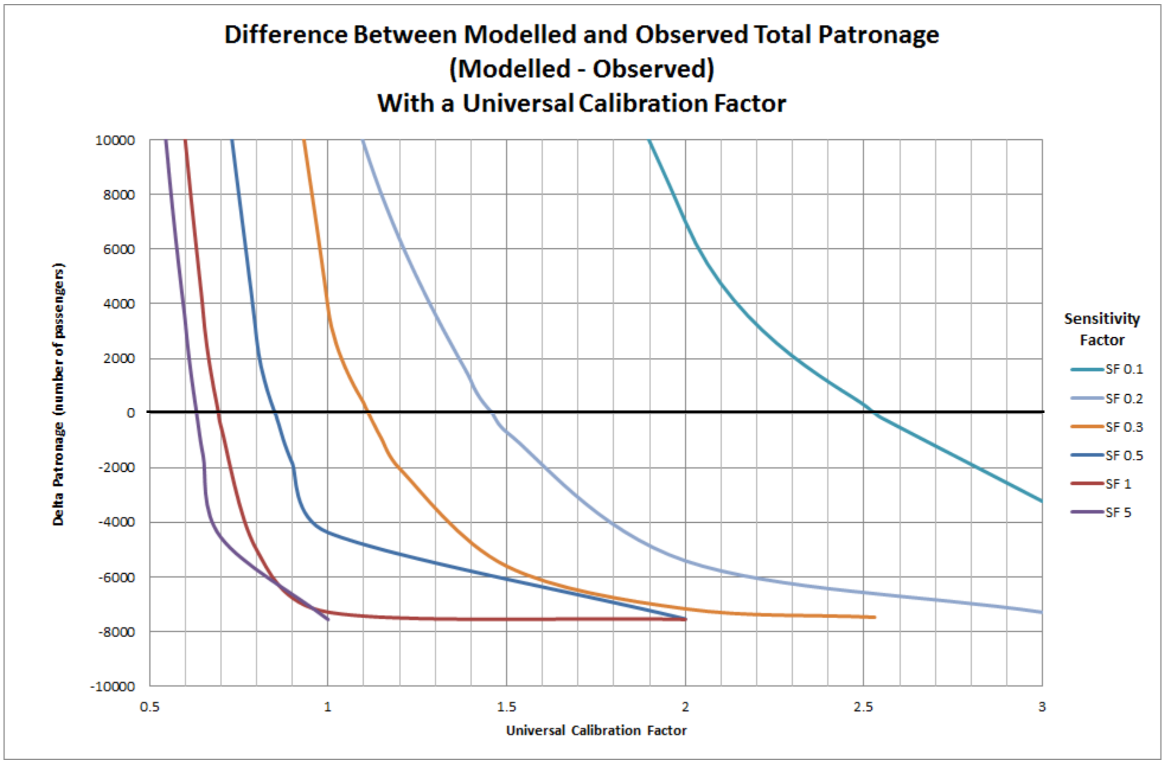
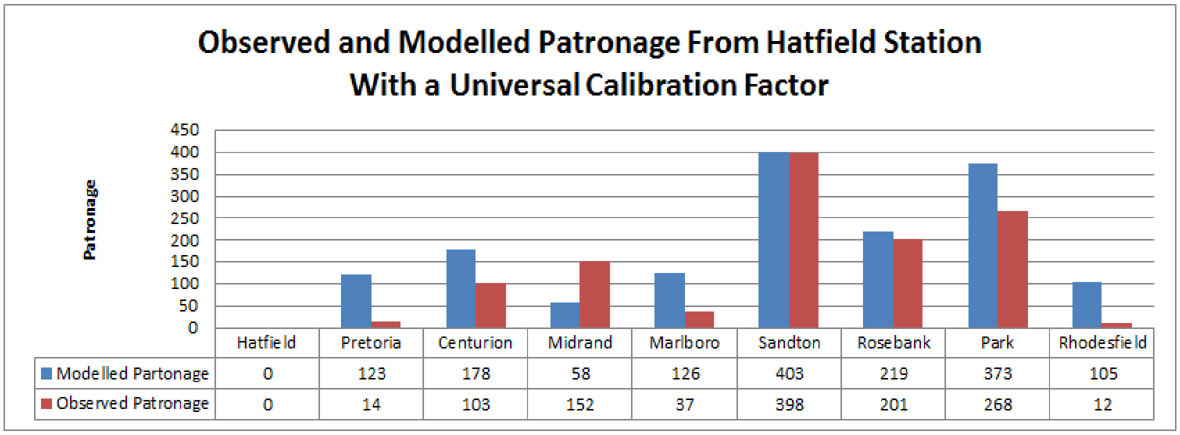


Figure 14: Delta Patronage with Universal Calibration Factor (Zoomed)

The model is calibrated in such a way that, for each sensitivity factor, the total modelled patronage is equal to the total observed patronage.

The approach of utilising a single universal calibration factor assumes that all the input parameters are correct and that there is a homogeneous decision making process throughout the network. As mentioned, this approach validates the model logic and mode split algorithms. If there are some irregularities in the model logic or the mode choice algorithms, the variance between the observed and modelled values will be substantial. Figure 15 provides an example of the analysis output. In this figure all the journeys with an origin at Hatfield station are combined. See Appendix A for more station specific data.



	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	0	123	178	58	126	403	219	373	105
Observed Patronage	0	14	103	152	37	398	201	268	12

Figure 15: Comparison between Observed and Modelled Patronage (Hatfield)

The use of a single sensitivity and calibration factor proves that the logic and general working of the model is adequate and acceptable.

4.7 Iterative Calibration process

After the model validation process has confirmed that the model logic and methods applied are acceptable, the model calibration process follows. The model calibration entails an iterative calibration process between calibrating the station to station calibration factors, the network sensitivity factor and the seasonal factor.

The first iteration assumes that there is no seasonal effect on the patronage. This approach first determines the effect of the network sensitivity factor on the model. Note that this is different from the model validation steps, because individual station specific calibration factors are used instead of a single universal factor.

The model is calibrated using a certain sensitivity factor. A backward and forward prediction is done using the calculated calibration factors and the specified sensitivity factor. Figure 16 illustrates the effect of the sensitivity factor on the modelled patronage.

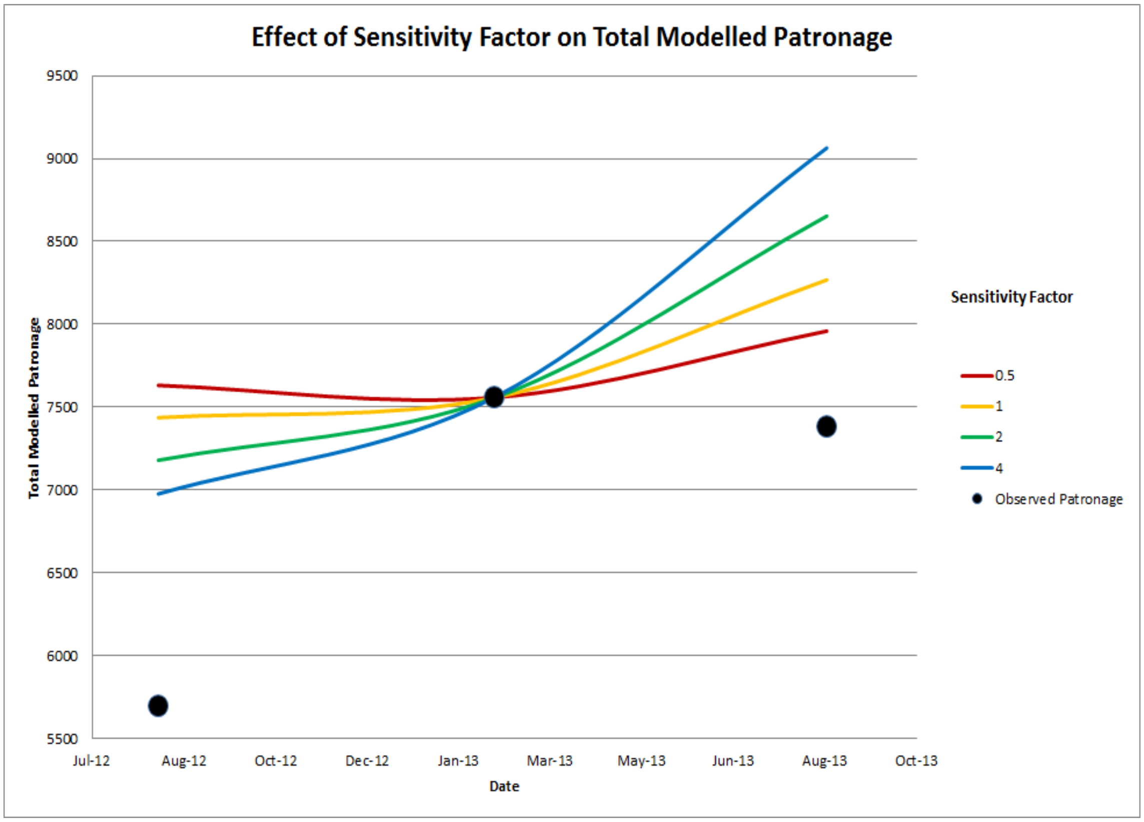


Figure 16: Effect of Sensitivity Factor on the Modelled Patronage

The observed patronage for the various scenarios was:

- August 2012 – 5 691 passengers
- February 2013 – 7 558 passengers
- August 2013 – 7 380 passengers.

Note that the observed patronage for August 2013 is lower than the observed patronage for February 2013. The modelled output however models a higher patronage for the August 2013 scenario, regardless of the sensitivity factor. The backward predicted patronage for August 2012 is also much more than the observed patronage.

The assumption is that the over estimation of the modelled patronage is predominantly because of the difference in weather conditions between the months of February and August. The model is calibrated for the base year scenario and therefore represents the decision making attributes of travelers during the month of February. The seasonal effect is expressed as a percentage change in the patronage relative to February. Both the backward and forward prediction scenarios are for the month of August, and will therefore have the same seasonal factor. Figure 17 shows the effect of the seasonal adjustment on the modelled patronage for a fixed sensitivity factor of 4.

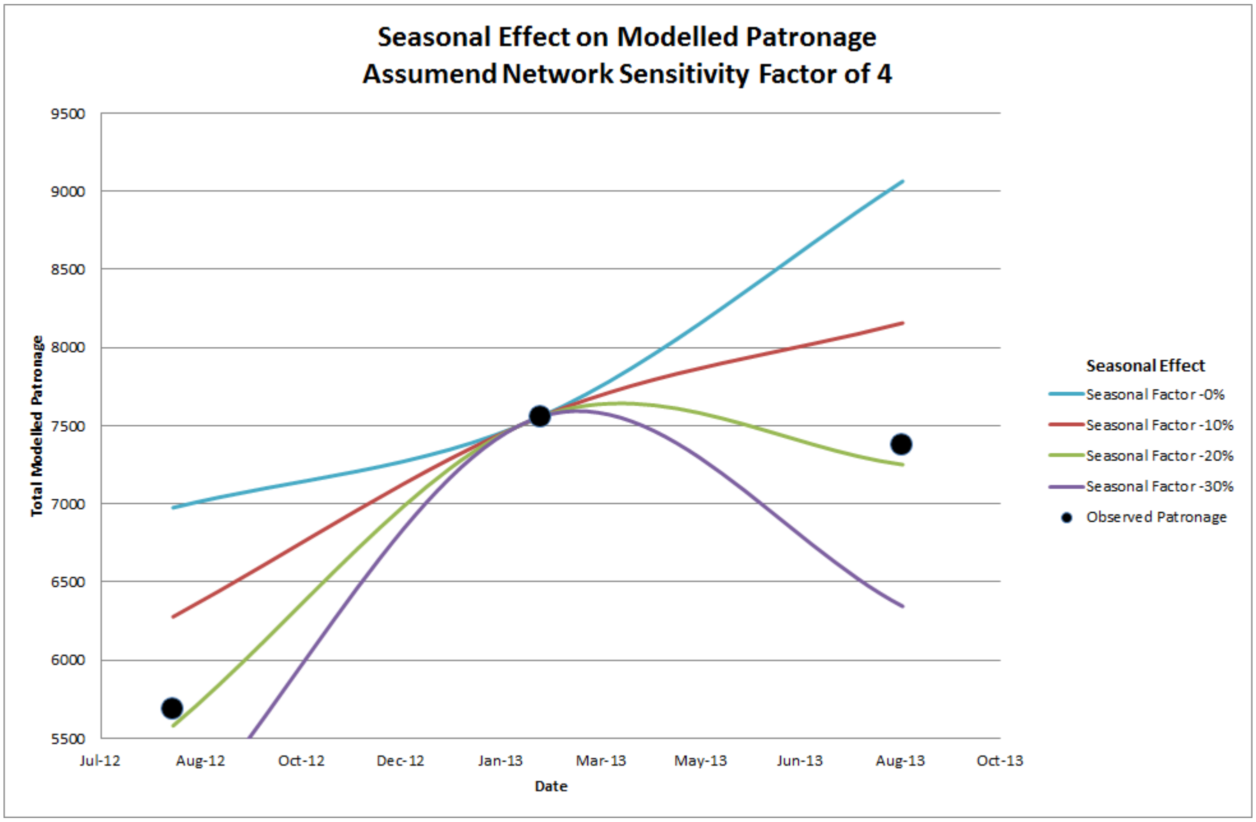


Figure 17: Seasonal Effect on Modelled Patronage

The iterative calibration process produced the following factors for the calibrated model:

- Network sensitivity factor 3.9
- Seasonal effect -18.4%.

Figure 18 illustrates the modelled outputs for the calibrated model.

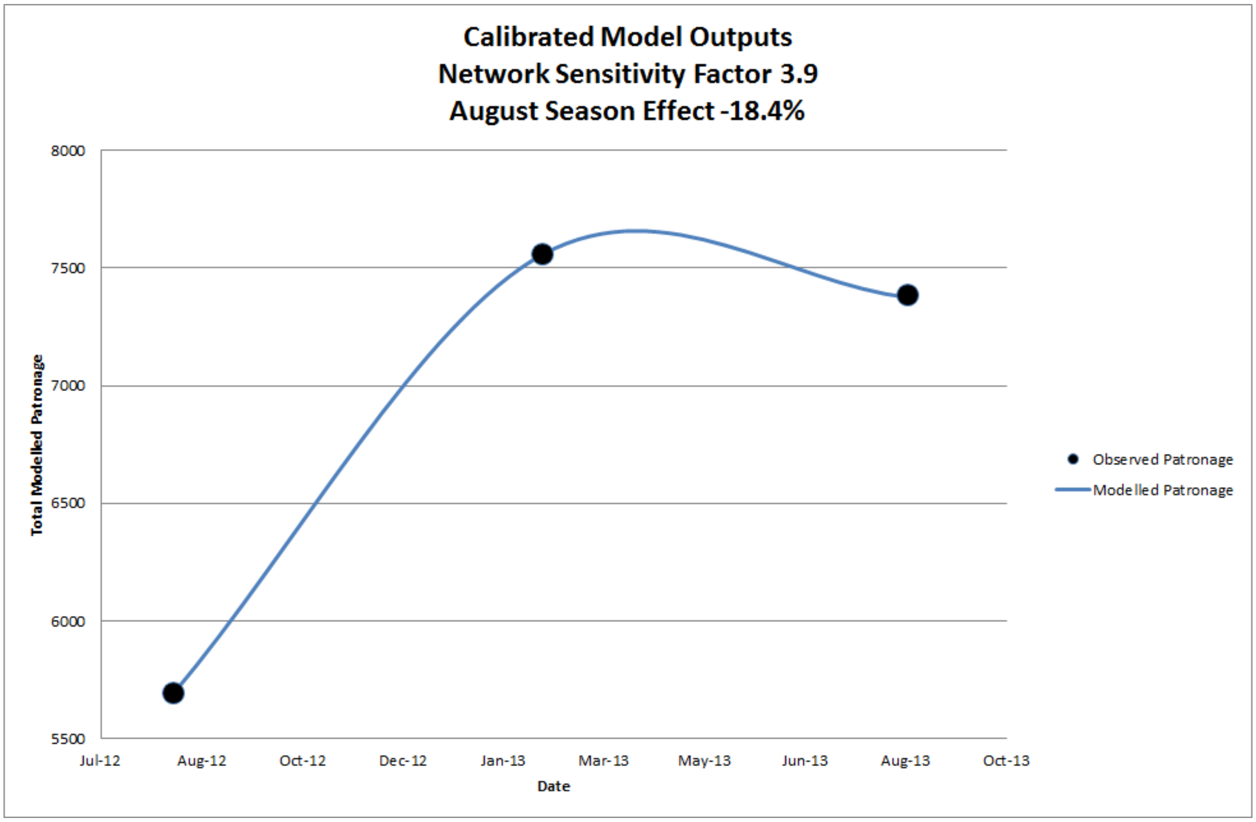


Figure 18: Calibrated Model Outputs

Table 12 provides a summary of the calibrated model outputs.

Table 12: Summary of Calibrated Model outputs

Scenario	Observed Patronage	Modelled Patronage
Scenario 1	5 691	5 691
Scenario 2	5 380	5 375

4.8 Model independent seasonal factor validation

The seasonal factor in the previous section is not derived from any surveyed data, but is rather just a percentage incorporated to compensate for the monthly variations. As mentioned, the purpose of this research project is to develop a model that is a simplified representation of the traditional modelling approach. The simplified mode choice model must still produce the desired results without the incorporation of many additional assumptions derived from surveyed data.

The estimation of a seasonal factor from model independent data is merely to verify that the calculated seasonal effect used in the model calibration process is in line with what is expected.

To roughly estimate the seasonal effect on the Gautrain patronage the approach is to analyse the average total morning patronage for an entire year. It is assumed that the annual growth is linear and the change in patronage between August 2012 and August 2013 is not as a result of seasonal factors, but that it is solely as a result of the change in journey characteristics. Figure 19 indicates the average morning peak patronage from August 2012 to August 2013. Note that the data used for this process includes the airport passengers and differs slightly from the total patronage used in the rest of this research project. The patronage is calculated as the average patronage observed for the typical weekdays for each month. A typical weekday is considered to be all the Mondays, Tuesdays, Wednesdays and Thursdays in a month, but excludes any public holidays. It is commonly accepted in the transport industry to exclude Fridays from any analysis.

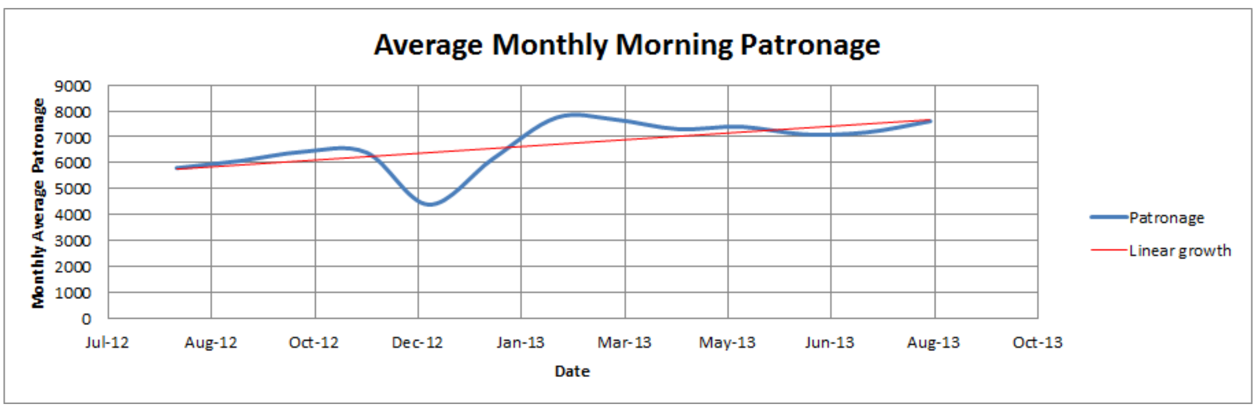


Figure 19: Average Monthly Patronage between August 2013 and August 2013

The relative linear growth for all the values are removed from the above mentioned data in order to set the August 2012 patronage equal to the August 2013 patronage. Figure 20 indicates the flattened patronage curve with the annual growth removed.

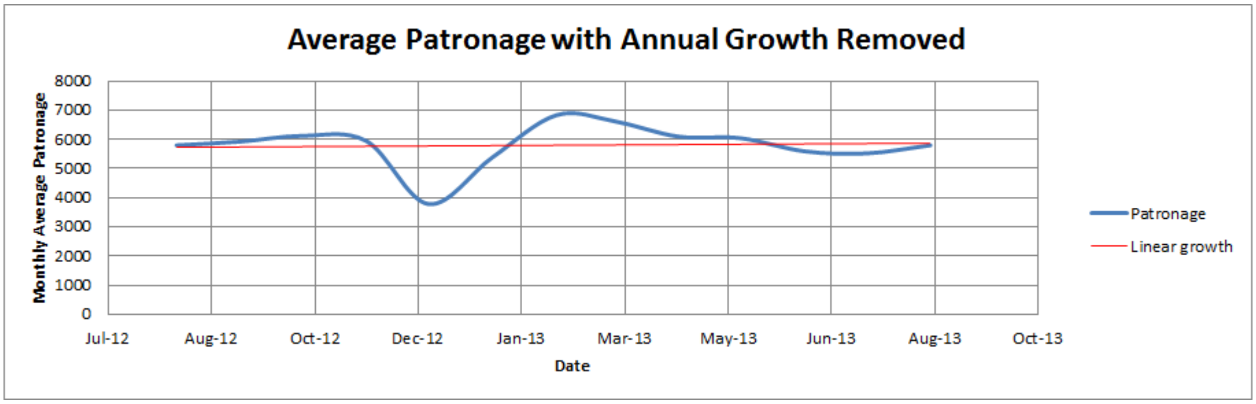


Figure 20: Average Patronage with Annual Growth Removed

According to Figure 20, February is the month with the highest value and is therefore considered as the ideal month with no reduction in patronage due to seasonal effects. The seasonal effect for the other months is calculated as a percentage reduction in patronage from the ideal month of February. It is also apparent that the weather is not the only factor involved in the monthly variation and factors such as the end of year holiday period also have an effect. Table 13 lists the calculated percentage reductions for the various months.

Table 13: Monthly Percentage Reductions

Month	% Reduction
January	21.2%
February	0.0%
March	3.1%
April	10.8%
May	11.6%
June	18.1%
July	19.2%
August	15.2%
September	13.3%
October	10.4%
November	13.0%
December	44.7%

The monthly percentage reduction for the month of August is 15.2% and is in the same order of magnitude as the seasonal factor of 18.6% as calculated by the iterative calibration process.

4.9 Summary

Considering the data presented in this section it is concluded that the simplified approach to model the Gautrain patronage is adequate for the purpose of this research project. The total observed patronage can be replicated with the model and the total predicted patronage is an adequate representation of reality. A more detailed analysis on the model's prediction capabilities is presented in the following chapters.

5. GAUTRAIN MODEL SENSITIVITY ANALYSIS

A sensitivity analysis is done in order to determine the effect of individual input parameters on the modelled patronage. The model's sensitivity towards certain parameters is analysed by altering individual input parameters whilst the rest of the parameters is kept unchanged. Each of the parameters is increased and reduced by 10% and 20% respectively. Figure 21 illustrates the model's sensitivity towards the change in the various input parameters. A positive gradient indicates that an increase in the value of an input parameter will result in an increase in the modelled patronage, like the fuel price graph shows. A negative gradient indicates that an increase in the value of that particular input parameter will lead to a reduction in the modelled patronage. The factors resulting in a negative gradient include the Gautrain ticket prices, transfer times, train frequencies, station parking costs and the Gautrain bus fares. The steepness of the gradient indicates the model's sensitivity towards that particular input parameter, where the greater the gradient the more sensitive the model.

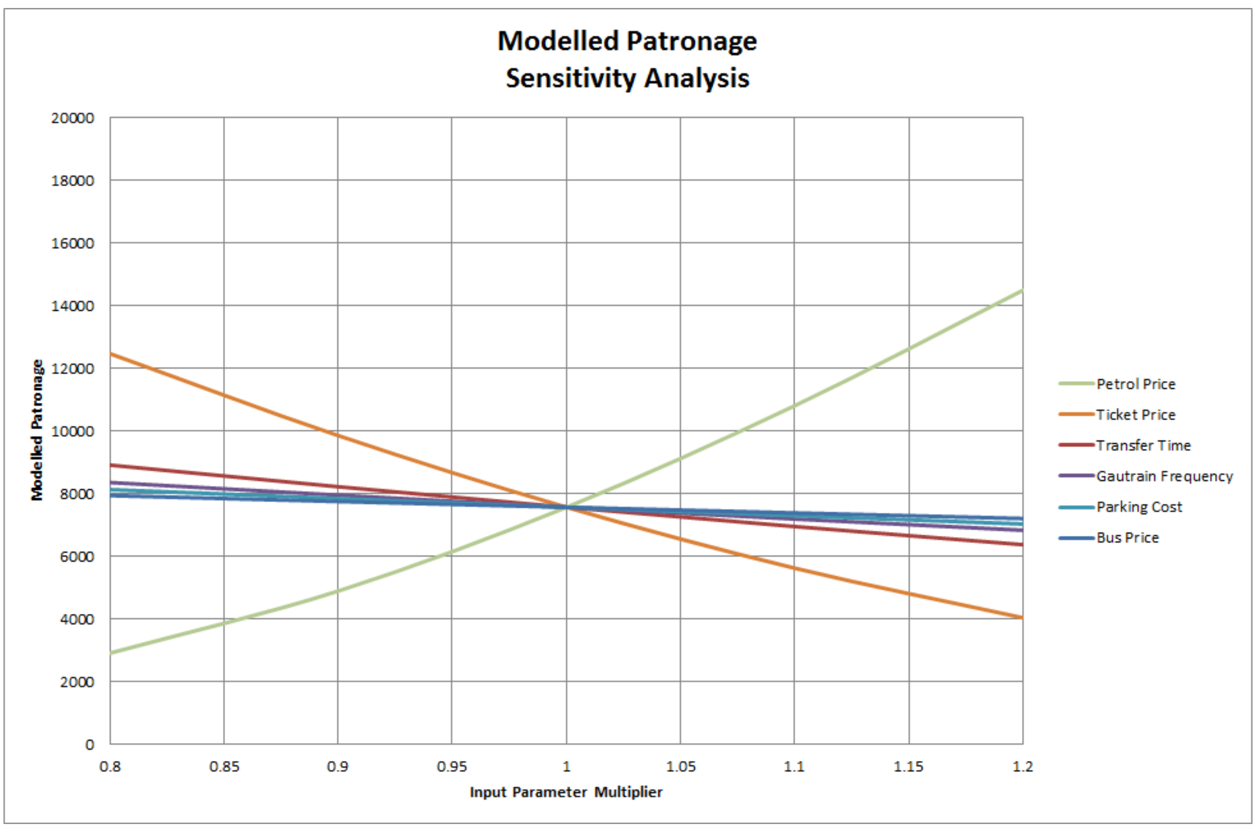


Figure 21: Sensitivity Analysis on the Change in Input Parameters

Figure 21 is summarised in Table 14.

Table 14: Sensitivity Analysis on the Change in Input Parameters

Change in Input parameter	Petrol Price	Ticket Price	Transfer Time	Gautrain Frequency	Parking Cost	Bus Price
-20%	-62%	65%	18%	10%	7%	5%
-10%	-35%	30%	9%	5%	4%	2%
No Change	0%	0%	0%	0%	0%	0%
+10%	43%	-26%	-8%	-5%	-4%	-2%
+20%	92%	-47%	-16%	-10%	-7%	-5%

The two major contributing factors to the mode choice model are the fuel price and the Gautrain ticket prices. These input parameters are a direct out of pocket cost and contribute to the largest portion of the total perceived journey cost. It is considered as realistic that a 10% increase in the fuel price will result in a 43% increase in Gautrain patronage. Note that for the purpose of the sensitivity analysis, all the other input parameters are kept unchanged and only the single parameter being tested is altered. It is however unlikely that in reality the fuel price alone will increase with 10% whilst the rest of the input parameters remains unchanged.

The relatively high percentage changes in the parameters are used to test the robustness of the model. Most of the cost related input parameters will change according to the CPIX of 6% per annum, whereas the fuel price increases on average by 10.68% per annum (Automobile Association of South Africa).

The parameters effected by Gautrain operations, such as the transfer times and train frequencies, only have an impact on the access and egress times. The percentage change of these parameters will only have an impact on a portion of the overall journey time. The effect that these parameters have on the patronage is therefore expected to be less than the major contributing factors of fuel and ticket prices. The parking cost at the stations and the bus fares also only contribute to a small percentage of the overall perceived journey cost and has therefore a smaller impact on the overall patronage.

The sensitivity of the model is within reasonable expectation and responds to the change in input parameters as anticipated.

6. RESULTS

A detailed analysis is done on the overall prediction capabilities of the model. The model is calibrated as described in section 4.7. Scenario 1 is the backward prediction scenario for August 2012 and scenario 2 is the forward prediction scenario for August 2013. The overall correlation between the modelled and observed values is 0.961 and 0.975 for scenarios 1 and 2 respectively.

In order to visually present the accuracy and predictability of the model, the modelled values for each OD pair are plotted against the observed values. For each data point the “x” value is the observed value, while the “y” value is the modelled value. A desired output trend line is added with a slope of 1 and a “y” intercept at zero. The desired output trend line has a function of “y = x”, where “x” represents the observed value and “y” represents the modelled value. If the modelled values are equal to the observed values, the data point will be on the desired output trend line. As the variation between the modelled and observed values increase, so too will the distance between the data point and the desired output trend line. The total patronage to and from each station is also calculated in order to determine the general predictability for each individual station.

The outputs of each of the scenarios are presented in the following sections and are followed by a discussion thereof at the end of the chapter. The detailed modelled and observed station to station OD matrices are presented in Appendix B.

6.1 Scenario 1

The overall correlation between the observed and modelled data is 0.980 (R-squared value of 0.961) and the backward prediction capabilities are considered as adequate. The total observed and modelled patronage for August 2012 is 5691 passengers. Figure 22 indicates the individual OD pair's passenger volumes as a combination of the modelled and observed values.

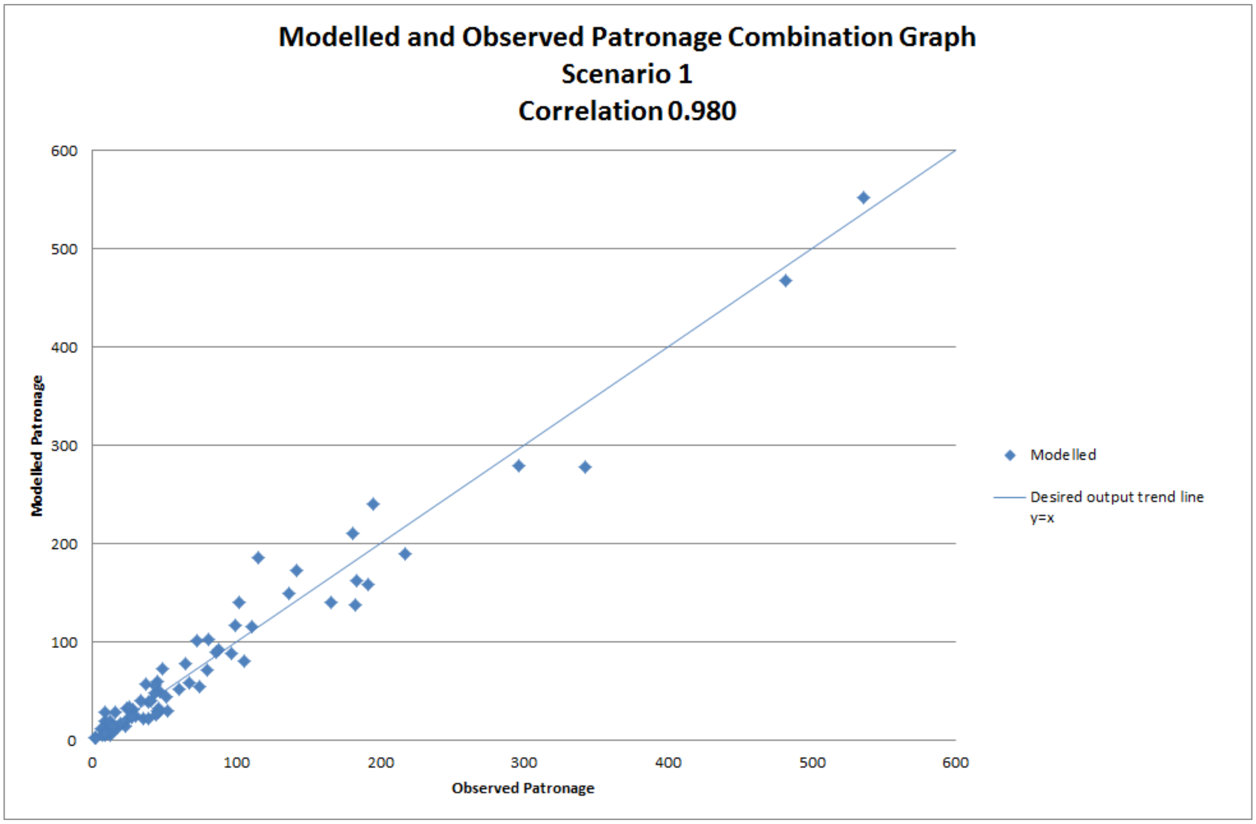


Figure 22: Scenario 1 Combination Graph

The total modelled patronage of each individual station is analysed separately in order to determine the predictability of the model on a station level. The total number of passengers to and from each station is calculated for the observed and modelled patronage. Table 15 illustrates the per station analysis. Also see Appendix C for more detailed station specific comparisons.

Table 15: Scenario 1 Station Specific Summary

Station	Scenario 1			
	Total Observed Patronage to and from the station	Total Modelled Patronage to and from the station	Difference in Patronage (Modelled – Observed)	Percentage Error
Hatfield	1471	1395	-76	-5%
Pretoria	1016	1086	70	6%
Centurion	1714	1718	5	0%
Midrand	1156	1160	4	0%
Marlboro	363	322	-41	-13%
Sandton	2194	2167	-27	-1%
Rosebank	996	902	-95	-11%
Park	1452	1589	137	9%
Rhodesfield	1043	1043	0	0%
Average/ Weighted Average	**1267**	**1265**	**49**	**4.0%**

The average number of passengers observed and modelled per station is 1267 and 1265 passengers respectively. The weighted average of the absolute difference is 49 passengers per station and the weighted average of the absolute percentage error is 4.0%.

6.2 Scenario 2

The overall correlation between the observed and modelled data is 0.987 (R-squared value of 0.975). The total observed and modelled patronage is 7380 and 7375 passengers respectively. Figure 23 indicates the individual OD pair's passenger volumes as a combination of the modelled and observed values.

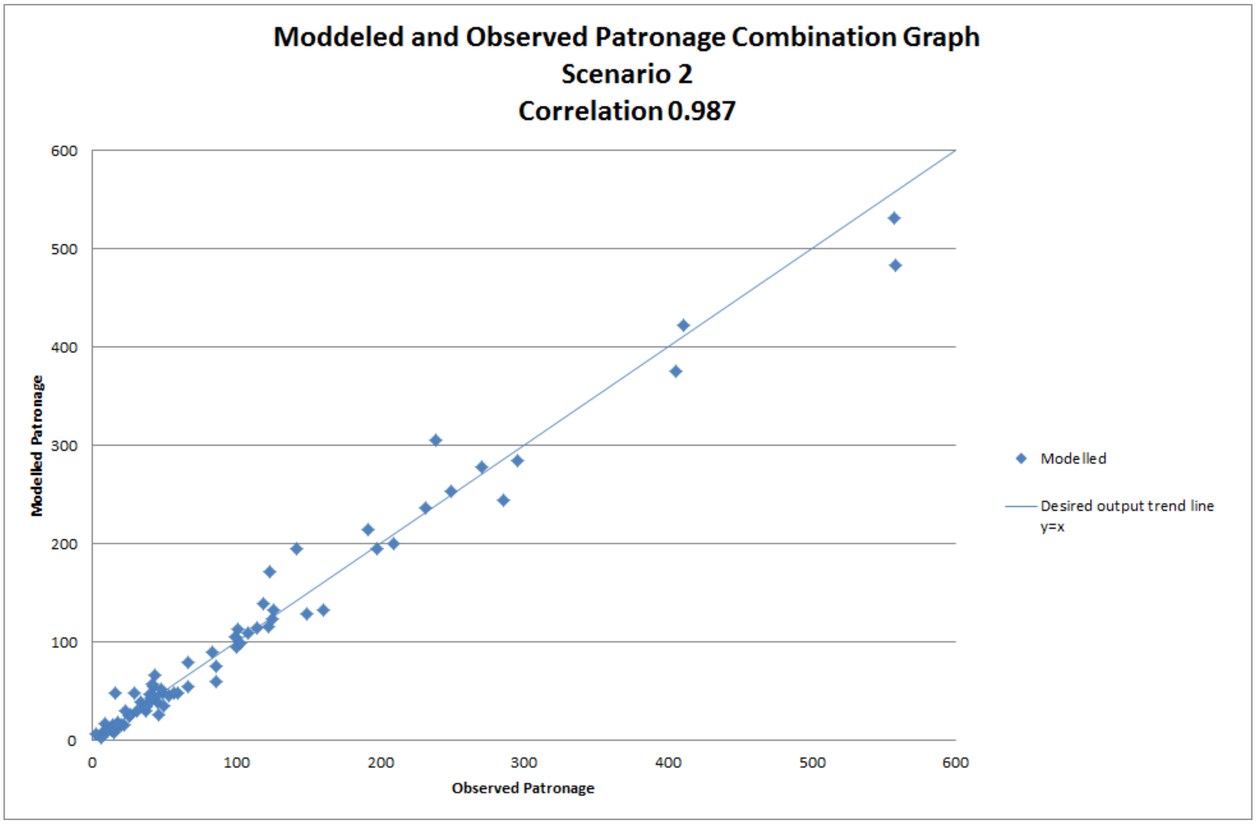


Figure 23: Scenario 2 Combination Graph

The overall forward prediction capability of the model is considered as adequate. Each of the individual stations is analysed in order to determine the model's behaviour per station. The total number of passengers to and from the stations is calculated for the observed and modelled patronage. Table 16 illustrates the per station analysis. Also see Appendix D for more detailed station specific comparisons.

Table 16: Scenario 2 Station Summary

Station	Scenario 2			
	Total Observed Patronage to and from the stations	Total Modelled Patronage to and from the stations	Delta Patronage	Percentage Error
Hatfield	1912	1954	43	2.2%
Pretoria	1333	1493	160	10.7%
Centurion	2140	2135	-5	-0.2%
Midrand	1538	1614	76	4.7%
Marlboro	469	423	-46	-11.0%
Sandton	2722	2665	-56	-2.1%
Rosebank	1155	1192	37	3.1%
Park	2321	2268	-52	-2.3%
Rhodesfield	1172	1007	-165	-16.4%
Average/ Weighted Average	**1640**	**1639**	**65**	**5.9%**

The average number of passengers observed and modelled per station is 1640 and 1639 passengers respectively. The weighted average of the absolute difference is 65 passengers per station and the weighted average of the absolute percentage error is 5.9%.

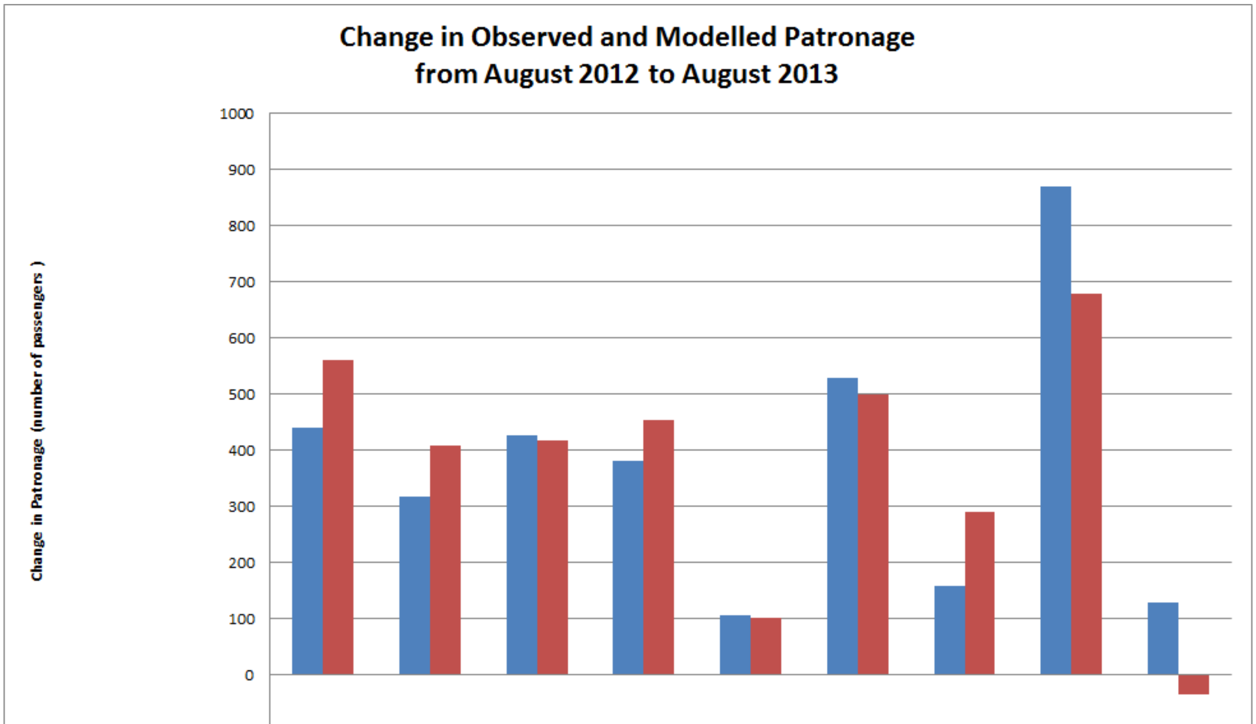
6.3 Discussion

The overall model prediction capabilities are considered as adequate with both the scenarios exhibiting an R-squared value in excess of 0.9. The model's predictability performance per station should be determined by considering both the percentage error as well as the absolute difference between the modelled and observed patronage.

The percentage error may be misleading if the total number of passengers is low. If the number of observed passengers is small, a small difference in observed and modelled patronage will result in a large percentage error. An example of this is the percentage error for Marlboro station in Scenario 1. For Scenario 1, Marlboro Station exhibits the largest percentage error of 13%. The total observed patronage for Marlboro station is 363 passengers and is considered as small in comparison to the network average of 1267 passengers per station. The difference of 41 passengers between the modelled and observed patronage at Marlboro station is in line with the average network difference of 49 passengers, but produces a large error percentage.

The absolute difference in patronage may also be incorrectly interpreted if the percentage error is not considered. If the total number of observed passengers is large, a relatively large difference between the modelled and observed patronage will result in only a small percentage error.

In addition to the analysis of the total modelled and observed patronage for the two scenarios separately, the relative change in patronage between the two scenarios is also analysed. Figure 24 illustrates the change in observed and modelled patronage between the two scenarios (2013 – 2012 Patronage)



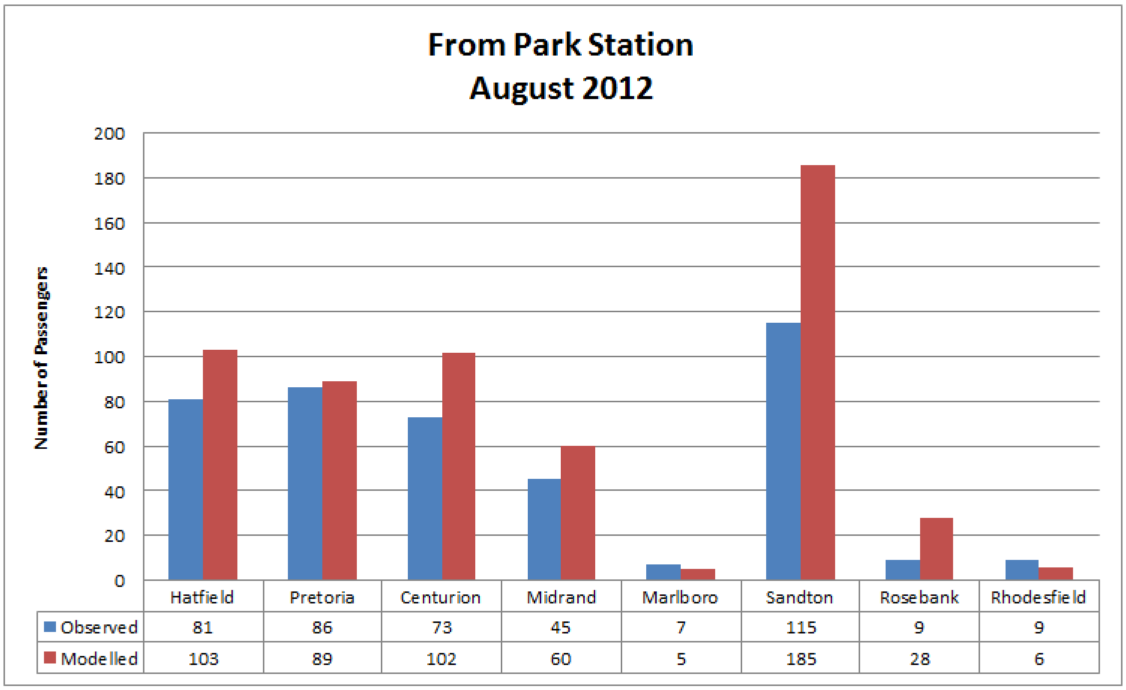
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Change in observed Patronage	440	317	426	382	106	528	159	869	129
Change in Modelled Patronage	559	407	416	454	100	498	290	679	-36

Figure 24: Change in Patronage between August 2012 and August 2013

When the outputs of Figure 24 are considered in conjunction with Table 15 and Table 16 the following conclusion is made: With the exception of Rosebank, Park and Rhodesfield, all the remaining stations are modelled with an acceptable degree of accuracy.

6.3.1 Park Station Analysis

For Scenario 1, Park station exhibits the largest absolute difference between the observed and modelled patronage. The total modelled patronage for Park station is 137 passengers more than the observed patronage and results in a 9% over estimation of the modelled passenger demand. The difference between the modelled and observed values for Park station for Scenario 2 is 52 passengers with a 2.3% error. Figure 25 illustrates a more detailed analysis of Park station during August 2012.



	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Rhodesfield
Observed	81	86	73	45	7	115	9	9
Modelled	103	89	102	60	5	185	28	6

Figure 25: Patronage from Park Station August 2012

The modelled patronage is notably higher than the observed patronage. The most plausible explanation for the overestimation is that the link to Park station had only been operational for 3 months by August 2012. In some instances the patronage on a newly implemented service may be less than what would have been observed had the service been operational for a longer period. Some people might not yet have considered the newly implemented service as a mode of choice, because they may still be unfamiliar with it. As the service matures these people will become familiar with the service and they will then start to consider the service as an option in their mode choice decisions making process. This increase in patronage is not due to the change in mode choice decision making factors, but rather due to the increase of the percentage of the population that considers the specific mode or service as a feasible choice. This occurrence is referred to as "ramp up". It is assumed that as a system matures, the above mentioned percentage will not change substantially between the various scenarios and can be ignored. The model will overestimate the passenger demand if a backward prediction is done on a link that has not yet matured and experienced a certain degree of ramp up.

The difference of 190 passengers (Figure 24) between the two scenarios for Park station is therefore attributed to the above mentioned effect of ramp up.

6.3.2 Rhodesfield Station Analysis

A detailed analysis of Rhodesfield station is done in order to determine the possible cause of the variation between the modelled and observed patronage. Figure 26 illustrates the detailed patronage for Rhodesfield for the two scenarios.

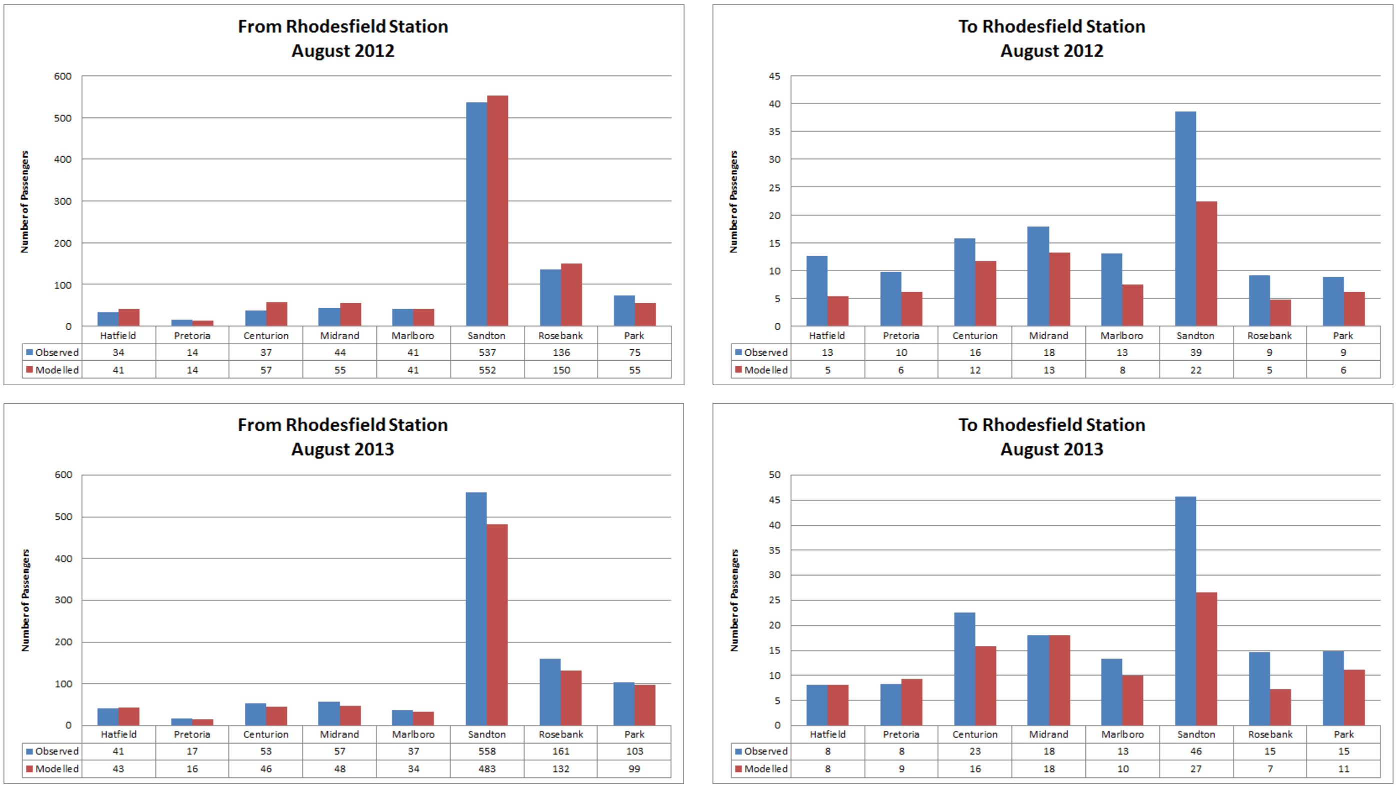


Figure 26: Rhodesfield Station Detailed Patronage for Scenarios 1 and 2

The overall patronage predictions for Rhodesfield station seem to be adequate, with the exception of the OD pair of passengers travelling from Sandton station to Rhodesfield station. The overall modelled values to Rhodesfield during August 2012 and 2013 are also predominantly lower than the observed values.

No clear explanation for the variation exists and a couple of possible causes are mentioned at the end of the following section.

6.3.3 Rosebank

For Scenario 1, the modelled patronage is 95 passengers less than the observed patronage and results in an 11% error. For Scenario 2, the difference between the modelled and observed patronage is 37 passengers with a 3.1% error. Figure 27 illustrates the detailed patronage for Rosebank for the two scenarios.

Figure 27: Rosebank Station Detailed Patronage for Scenarios 1 and 2

No clear explanation for the variation exists.

Considering all, the overall performance of the model is regarded to be adequately accurate. The probable cause for any difference between the observed and modelled patronage may be as a result of the following:

- Inaccurate land use data utilised during the trip generation step
- The simplified and coarse approach followed to determine a universal trip distribution pattern for the model
- The use of a single network sensitivity factor
- Difference in population demography around various stations
- Natural variance and random human behaviour.

7. CONCLUSION

With the development of the simplified mode choice model, it is concluded that the Gautrain patronage can be modelled with acceptable accuracy. The overall correlation of the model proves to be much higher than what is typically accepted with traditional models. The proposed simplified approach is therefore considered as a feasible solution to modelling public transport ridership with fewer resources and in a shorter time frame. The simplification of the model is achieved by incorporating the following techniques:

- Include all the measurable parameters that have accurate data available into the utility functions
- If no data is available and the parameters are deemed less important, do not assume values for the parameters, but rather exclude them from the mode choice model
- Exclude the assumed or estimated coefficients, typically calculated from regression analysis, from the model
- Exclude all intangible factors, such as safety and comfort from the utility function
- Include a single calibration factor for each station to station OD pair
- Incorporate all the non-measurable parameters into the above mentioned calibration factors
- Only focus on the applicable public transport mode.

The accuracy of the modelled patronage achieved for this research project is only feasible because of the high accuracy of the data derived from the Intelligent Transport Systems (ITS) implemented on the Gautrain network. Most of the public transport systems planned to be implemented in the future will have a high degree of ITS implementation on the networks and accurate traveller information will be available. It is anticipated that in future high detailed traveller information will become the norm rather than the exception because of the rapid growth in digital technology.

The negative effect of utilising roughly estimated trip generation and trip distribution patterns are reduced by calibrating each individual OD pair of the base year to correlate accurately with the observed values. The relatively short period of six months between the various scenarios also ensures that the effect of the predicted and assumed land use changes are relatively small and will therefore only have a small effect on the modelled patronage. It is however expected that with the incorporation of incorrect land use data and the extension of the forecast period, the outcomes of the model will not be as accurate as the results from the tested scenarios.

With the simplified model only focusing on the public transport portion of the network, the model requires much less surveyed information than traditional models. Because the model is geared to model only the public portion of the demand this approach cannot be used to model the private vehicle portion of the network.

With the forthcoming implementation of all the various BRT systems in South Africa, there is a need to develop a mode choice model for each of these systems. There are therefore other applications to which this model can be applied. The BRT system operates similar to a rail service and can therefore be modelled in a similar manner.

One of the major shortcomings of the proposed modelling approach is that it can only be developed after a system has become operational and cannot be used to predict the patronage of a system that is not yet operational.

Another shortcoming of the proposed model is that the model does not include any capacity constraints and only calculates the theoretical demand on the public transport network. During the analysis period, the Gautrain was not operating at capacity and therefore no capacity constraints were incorporated. It is important to note that the approach utilised for this research project will probably not produce accurate results for a system that is operating at, or close to, capacity. Future studies can incorporate capacity constraints into the model in order to model mode choice behaviour for systems operating near capacity.

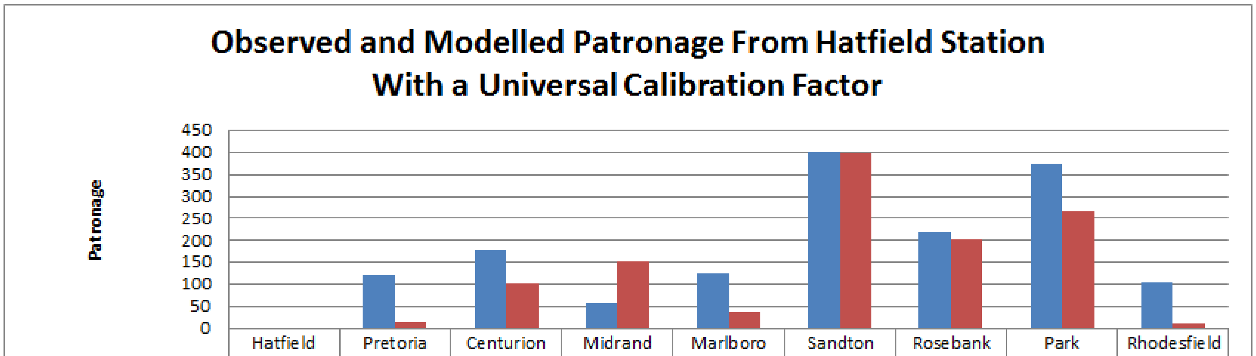
The availability of useful data will aid the further development of simplified models. High detailed traveller information is becoming more readily available through the implementation of various technological devices. The information from these devices may be implemented purely for transport related applications (such as network and vehicle monitoring devices), or may be derived from data that is generated from technologies outside the transport infrastructure (such as mobile phones).

The trip distribution pattern utilised for the purpose of this research project was calculated using the gravity-based model. This is a high level approach to developing a trip distribution pattern. With the ever increasing availability of usable data, it is recommended that the possibility of developing a high detailed trip distribution pattern for the Gauteng region utilising mobile device data should be investigated.

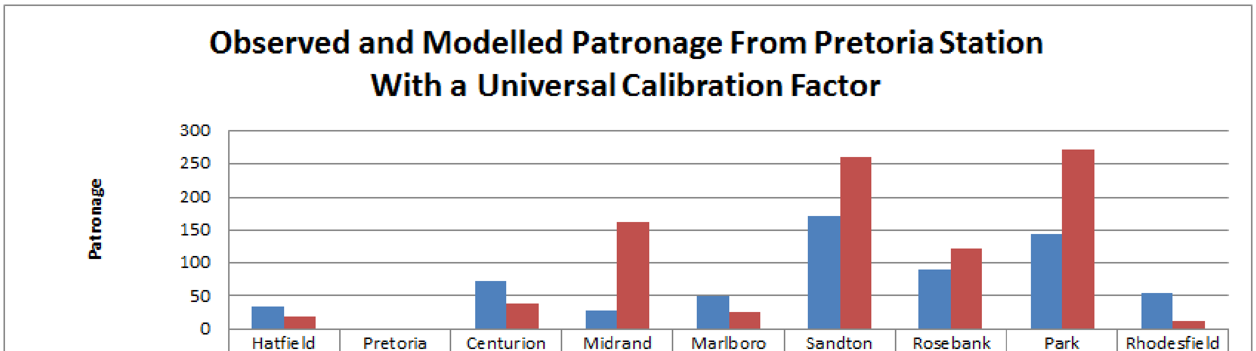
The possibility of including other modes of transport into the current model should also be investigated. As the various public transit systems are becoming operational, so should the model's capability to accommodate more than one public transport mode.

APPENDIX A

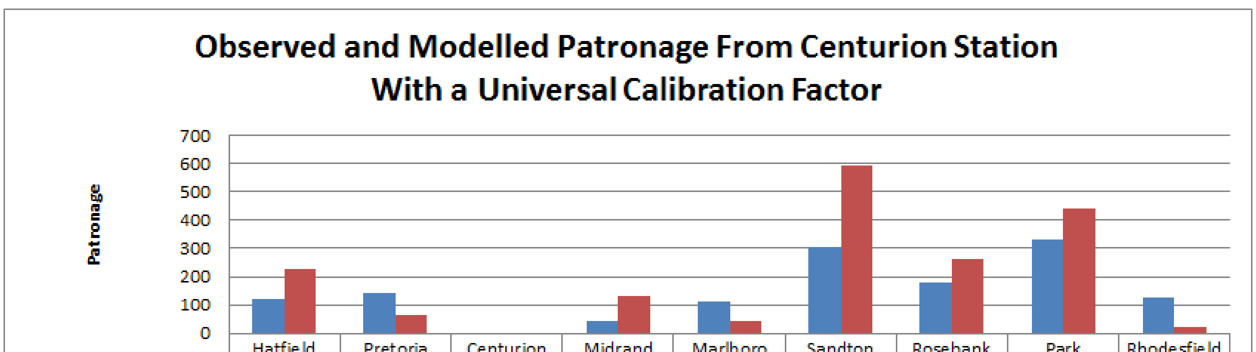
Appendix A contains the analysis outputs of the calibration process with a Universal Calibration Factor.



	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	0	123	178	58	126	403	219	373	105
Observed Patronage	0	14	103	152	37	398	201	268	12

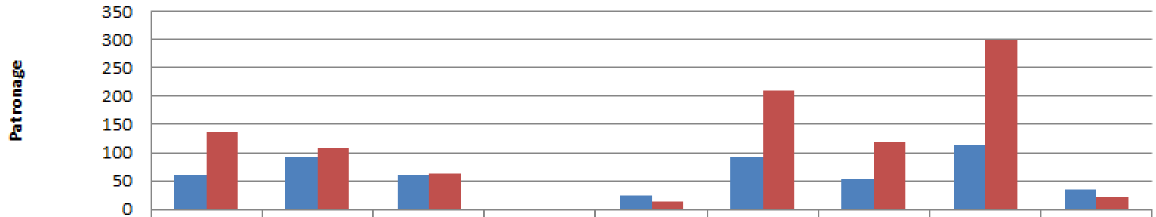


	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	34	0	72	28	49	171	91	144	55
Observed Patronage	19	0	39	163	26	262	121	272	12



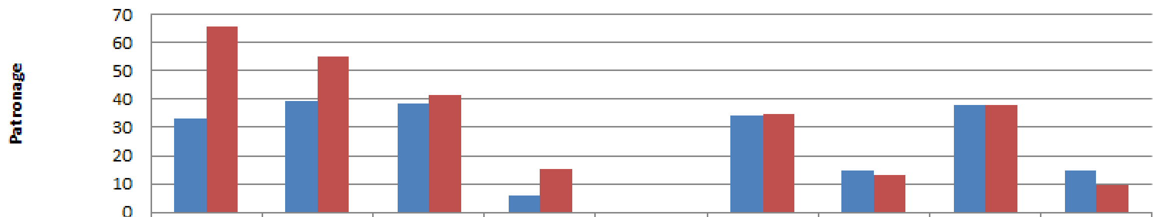
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	123	144	0	45	112	305	179	328	125
Observed Patronage	228	62	0	134	46	591	261	440	21

Observed and Modelled Patronage From Midrand Station With a Universal Calibration Factor



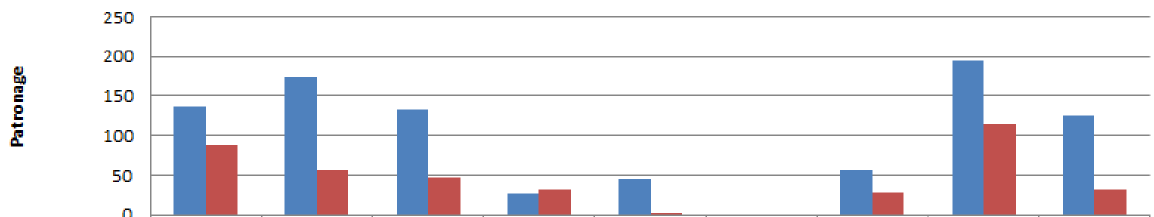
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	61	91	61	0	25	91	53	113	35
Observed Patronage	138	107	64	0	15	211	118	299	21

Observed and Modelled Patronage From Marlboro Station With a Universal Calibration Factor

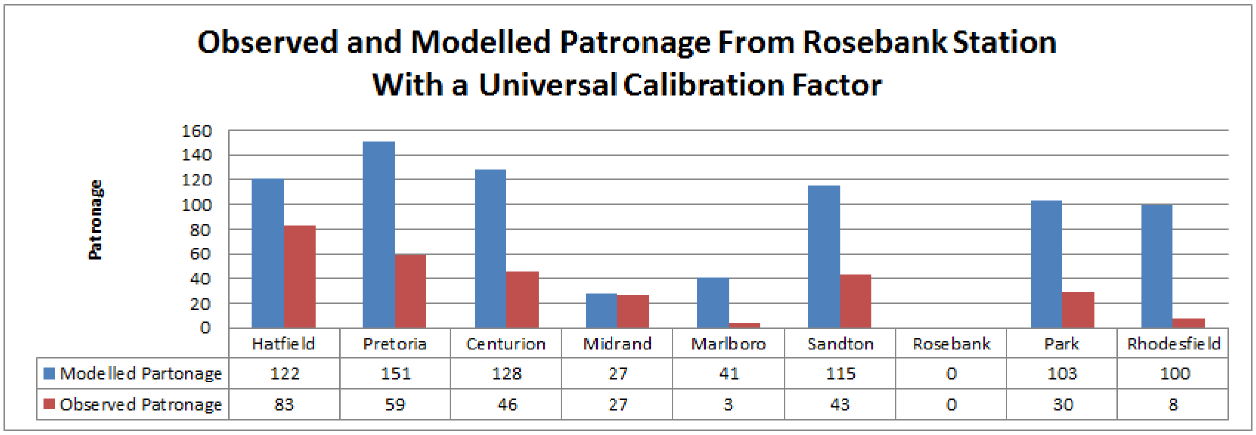


	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	33	39	38	6	0	34	15	38	15
Observed Patronage	65	55	42	15	0	35	13	38	10

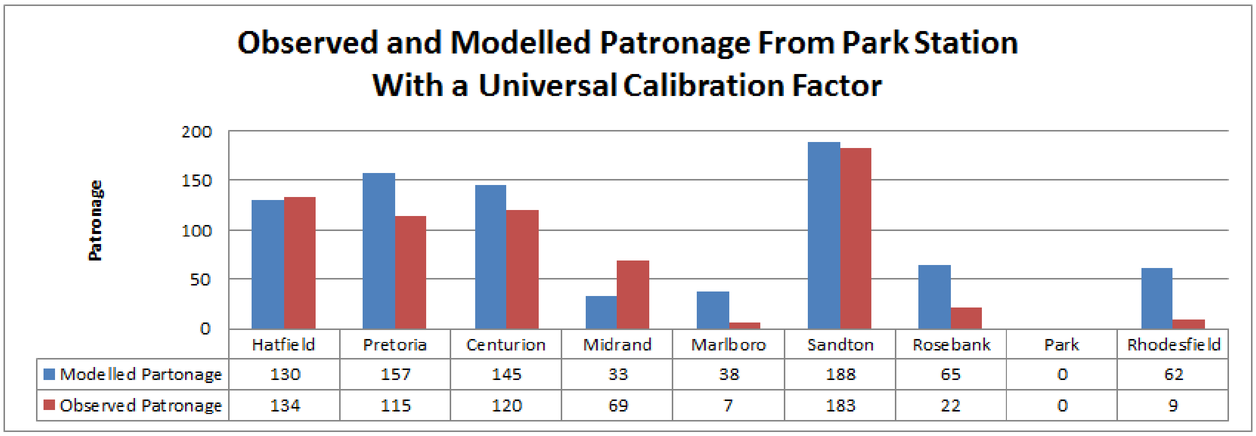
Observed and Modelled Patronage From Sandton Station With a Universal Calibration Factor



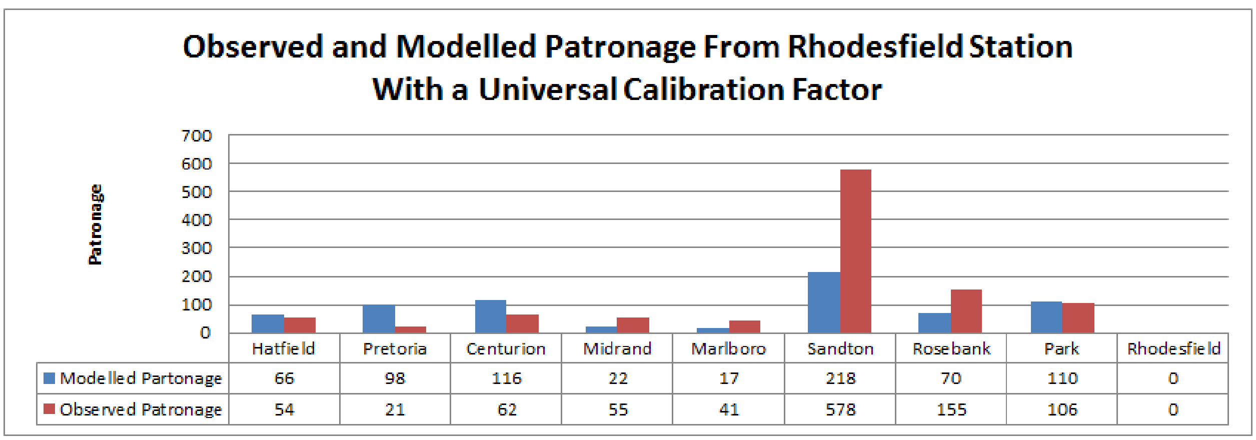
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	138	175	134	27	46	0	57	195	127
Observed Patronage	89	57	47	33	3	0	29	114	32



	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	122	151	128	27	41	115	0	103	100
Observed Patronage	83	59	46	27	3	43	0	30	8



	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	130	157	145	33	38	188	65	0	62
Observed Patronage	134	115	120	69	7	183	22	0	9



	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Modelled Partonage	66	98	116	22	17	218	70	110	0
Observed Patronage	54	21	62	55	41	578	155	106	0

APPENDIX B

Appendix B contains the various observed and modelled OD matrices.

August 2012 Observed Patronage									
	Hatfield	**Pretoria**	**Centurion**	**Midrand**	**Marlboro**	**Sandton**	**Rosebank**	**Park**	**Rhodesfield**
Hatfield	0	6	65	111	36	343	183	191	13
Pretoria	9	0	23	102	23	181	97	195	10
Centurion	142	52	0	100	46	482	217	296	16
Midrand	88	79	47	0	10	166	105	183	18
Marlboro	43	39	29	11	0	30	7	25	13
Sandton	67	46	45	28	2	0	16	49	39
Rosebank	61	52	44	19	2	26	0	12	9
Park	81	86	73	45	7	115	9	0	9
Rhodesfield	34	14	37	44	41	537	136	75	0

February 2013 Observed Patronage									
	Hatfield	**Pretoria**	**Centurion**	**Midrand**	**Marlboro**	**Sandton**	**Rosebank**	**Park**	**Rhodesfield**
Hatfield	0	14	103	152	37	398	201	268	12
Pretoria	19	0	39	163	26	262	121	272	12
Centurion	228	62	0	134	46	591	261	440	21
Midrand	138	107	64	0	15	211	118	299	21
Marlboro	65	55	42	15	0	35	13	38	10
Sandton	89	57	47	33	3	0	29	114	32
Rosebank	83	59	46	27	3	43	0	30	8
Park	134	115	120	69	7	183	22	0	9
Rhodesfield	54	21	62	55	41	578	155	106	0

August 2013 Observed Patronage									
	Hatfield	**Pretoria**	**Centurion**	**Midrand**	**Marlboro**	**Sandton**	**Rosebank**	**Park**	**Rhodesfield**
Hatfield	0	9	99	124	37	411	198	286	8
Pretoria	16	0	42	142	26	239	114	296	8
Centurion	192	42	0	119	45	558	231	405	23
Midrand	123	100	86	0	18	209	108	270	18
Marlboro	67	49	34	20	0	50	14	40	13
Sandton	86	60	42	37	6	0	27	101	46
Rosebank	67	48	43	23	3	44	0	29	15
Park	149	125	126	84	11	250	31	0	15
Rhodesfield	41	17	53	57	37	558	161	103	0

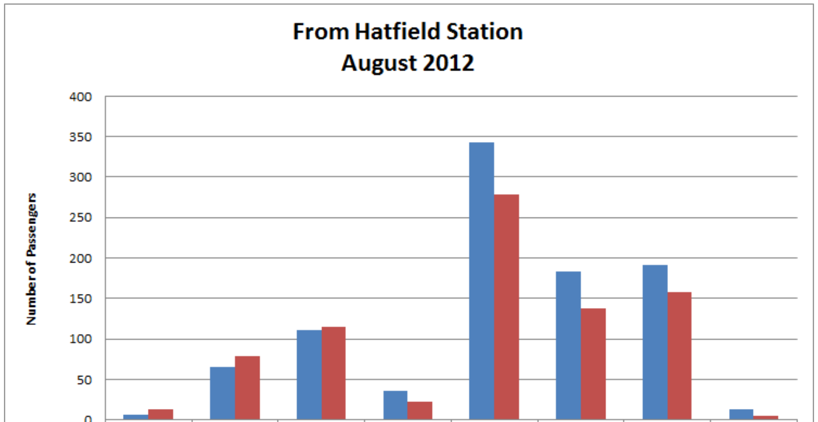
August 2012 Modelled Patronage									
	Hatfield	**Pretoria**	**Centurion**	**Midrand**	**Marlboro**	**Sandton**	**Rosebank**	**Park**	**Rhodesfield**
Hatfield	0	12	78	115	22	278	138	158	5
Pretoria	19	0	32	140	14	211	88	241	6
Centurion	173	44	0	117	29	467	190	279	12
Midrand	93	72	50	0	10	140	81	163	13
Marlboro	48	39	31	11	0	25	12	22	8
Sandton	58	33	30	23	2	0	29	73	22
Rosebank	52	29	26	17	3	35	0	19	5
Park	103	89	102	60	5	185	28	0	6
Rhodesfield	41	14	57	55	41	552	150	55	0

February 2013 Modelled Patronage									
	Hatfield	**Pretoria**	**Centurion**	**Midrand**	**Marlboro**	**Sandton**	**Rosebank**	**Park**	**Rhodesfield**
Hatfield	0	14	103	152	37	398	201	268	12
Pretoria	19	0	39	163	26	262	121	272	12
Centurion	228	62	0	134	46	591	261	440	21
Midrand	138	107	64	0	15	211	118	299	21
Marlboro	65	55	42	15	0	34	13	38	10
Sandton	89	57	47	33	3	0	29	114	32
Rosebank	83	59	46	27	3	43	0	30	8
Park	134	115	120	69	7	183	22	0	9
Rhodesfield	54	21	62	55	41	579	155	106	0

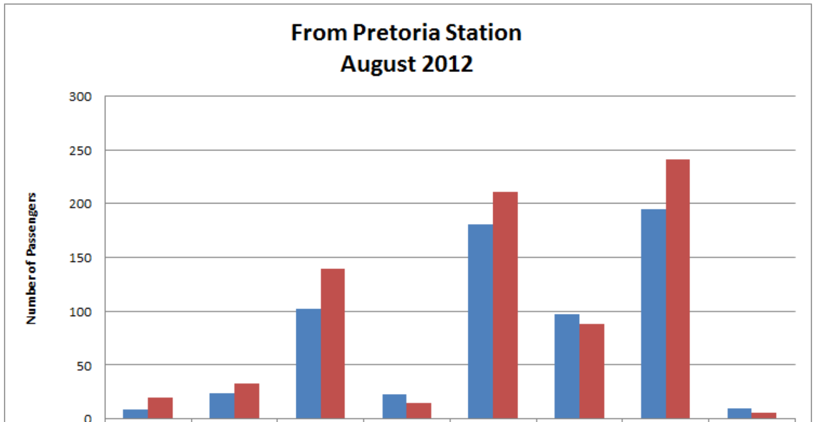
August 2013 Modelled Patronage									
	Hatfield	**Pretoria**	**Centurion**	**Midrand**	**Marlboro**	**Sandton**	**Rosebank**	**Park**	**Rhodesfield**
Hatfield	0	16	106	172	30	422	194	245	8
Pretoria	49	0	55	195	25	305	114	284	9
Centurion	215	57	0	139	39	531	237	376	16
Midrand	116	95	60	0	13	201	110	278	18
Marlboro	55	48	39	16	0	35	15	47	10
Sandton	76	48	42	35	2	0	26	113	27
Rosebank	79	52	44	30	7	67	0	48	7
Park	129	123	133	89	9	254	30	0	11
Rhodesfield	43	16	46	48	34	483	132	99	0

APPENDIX C

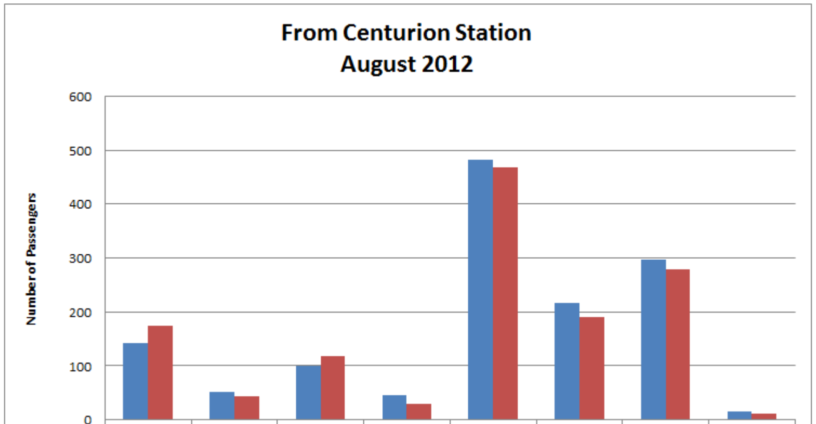
Appendix C contains the various station specific patronage values for Scenario 1.



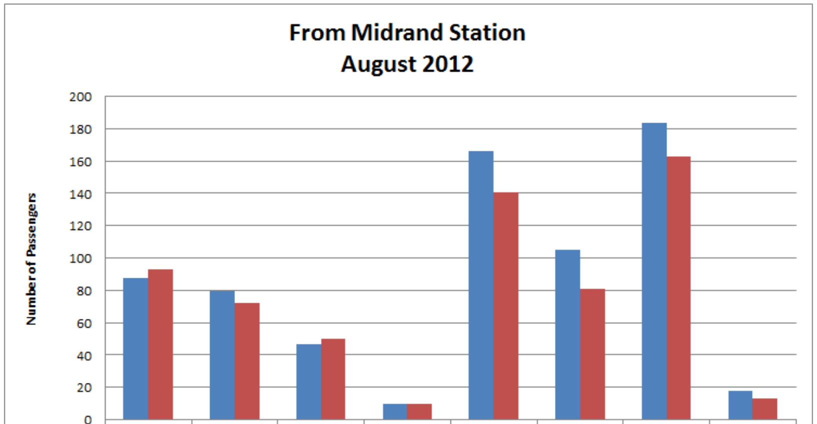
	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	6	65	111	36	343	183	191	13
Modelled	12	78	115	22	278	138	158	5



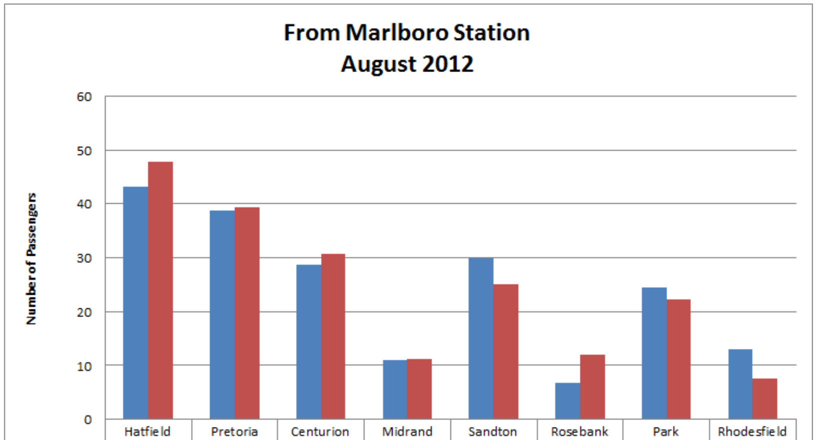
	Hatfield	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	9	23	102	23	181	97	195	10
Modelled	19	32	140	14	211	88	241	6



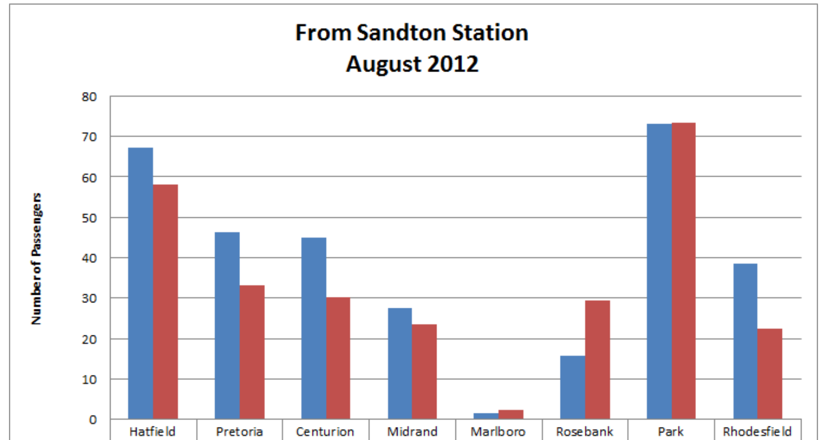
	Hatfield	Pretoria	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	142	52	100	46	482	217	296	16
Modelled	173	44	117	29	467	190	279	12



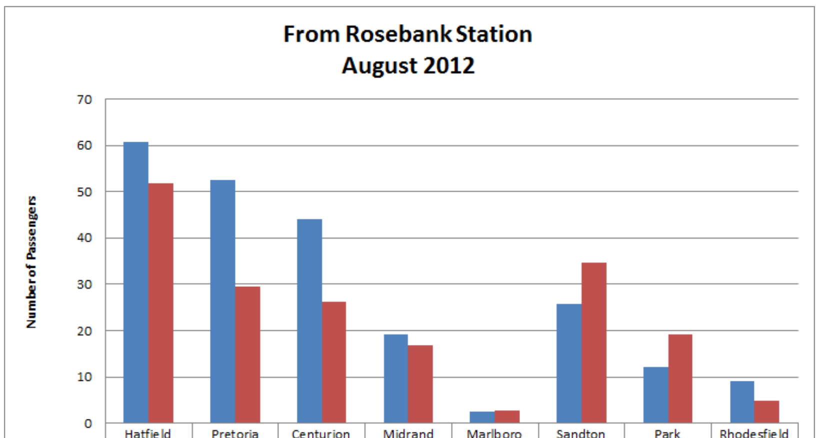
	Hatfield	Pretoria	Centurion	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	88	79	47	10	166	105	183	18
Modelled	93	72	50	10	140	81	163	13



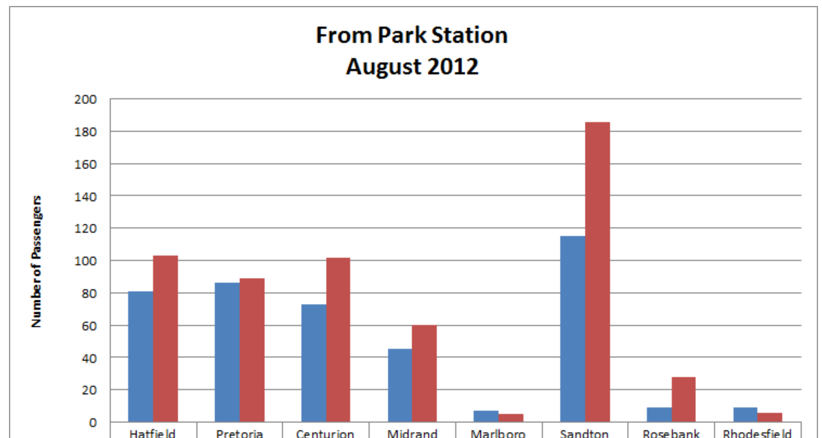
	Hatfield	Pretoria	Centurion	Midrand	Sandton	Rosebank	Park	Rhodesfield
Observed	43	39	29	11	30	7	25	13
Modelled	48	39	31	11	25	12	22	8



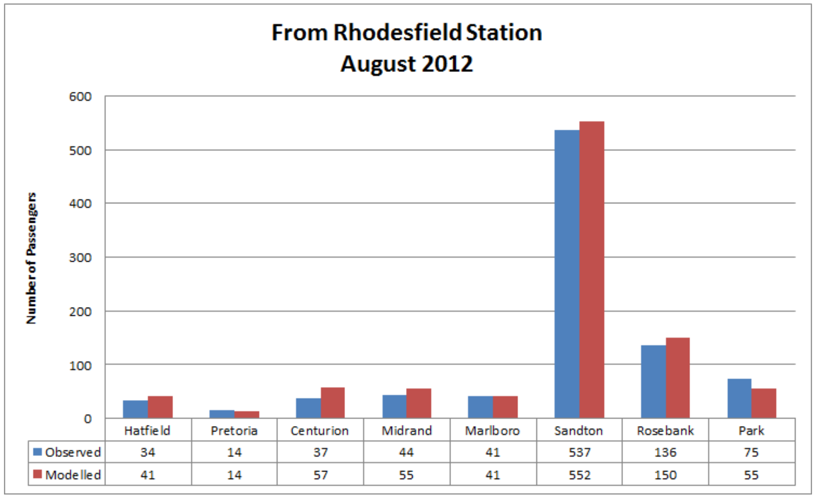
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Rosebank	Park	Rhodesfield
Observed	67	46	45	28	2	16	73	39
Modelled	58	33	30	23	2	29	73	22



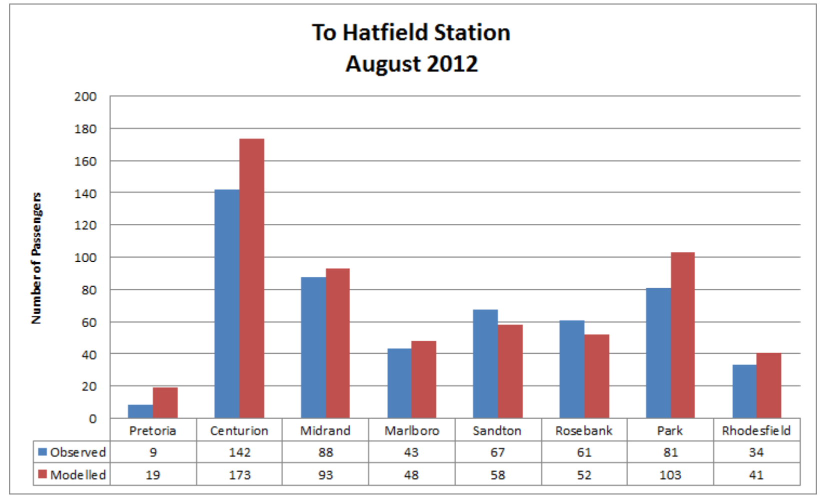
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Park	Rhodesfield
Observed	61	52	44	19	2	26	12	9
Modelled	52	29	26	17	3	35	19	5



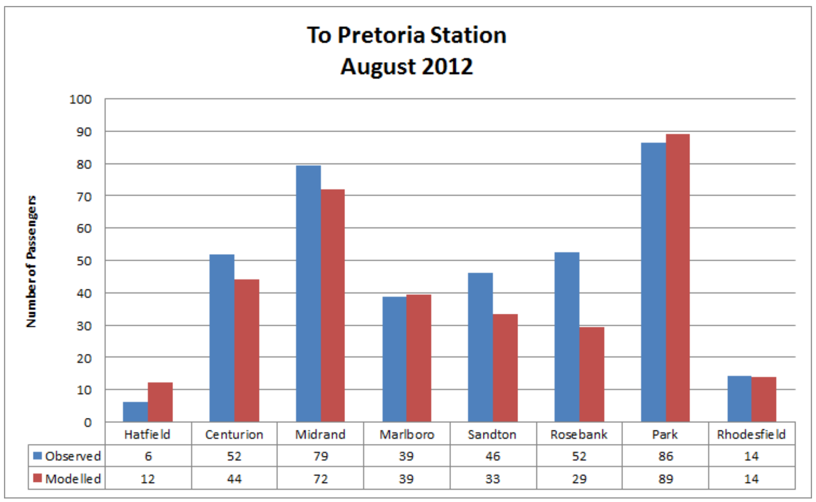
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Rhodesfield
Observed	81	86	73	45	7	115	9	9
Modelled	103	89	102	60	5	185	28	6



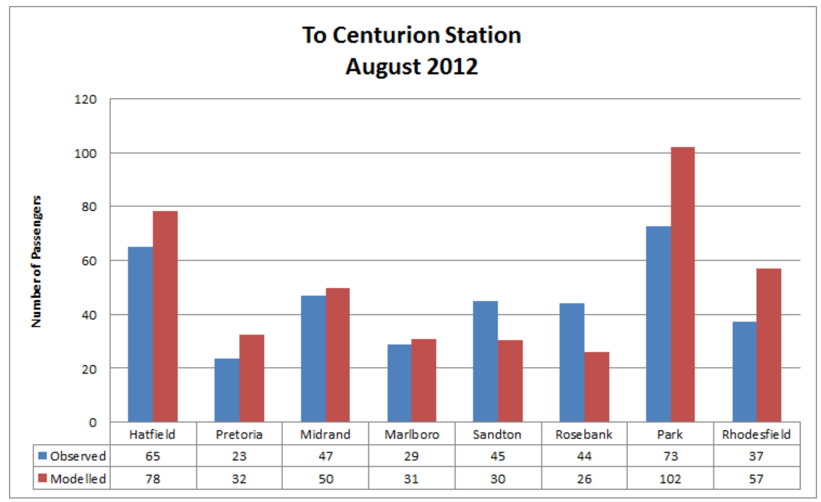
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park
Observed	34	14	37	44	41	537	136	75
Modelled	41	14	57	55	41	552	150	55



	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	9	142	88	43	67	61	81	34
Modelled	19	173	93	48	58	52	103	41

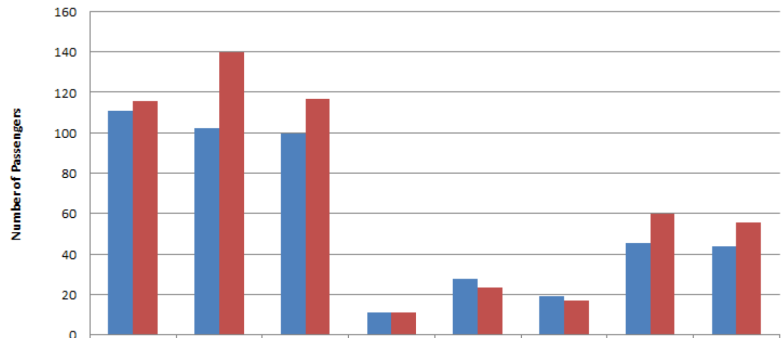


	Hatfield	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	6	52	79	39	46	52	86	14
Modelled	12	44	72	39	33	29	89	14



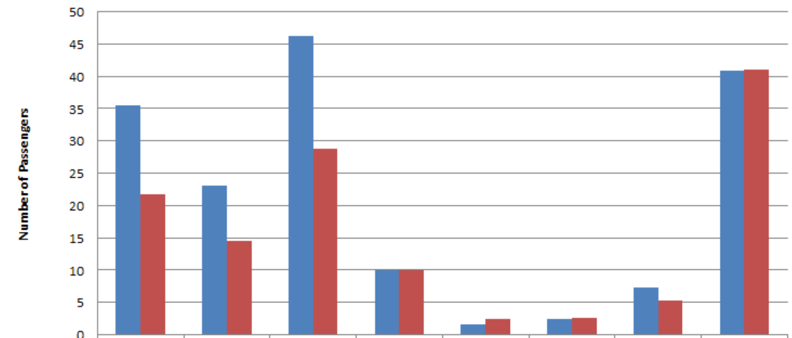
	Hatfield	Pretoria	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	65	23	47	29	45	44	73	37
Modelled	78	32	50	31	30	26	102	57

To Midrand Station
August 2012



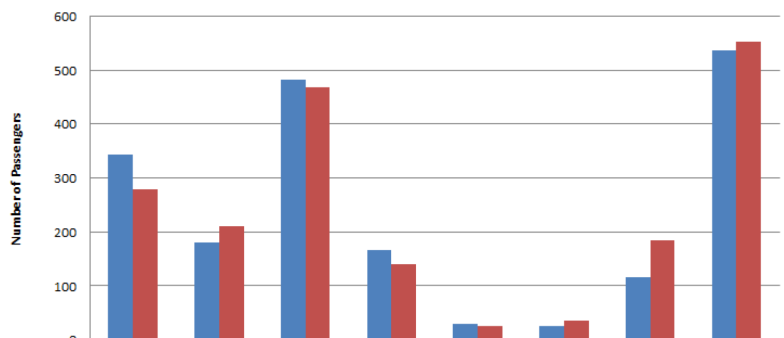
	Hatfield	Pretoria	Centurion	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	111	102	100	11	28	19	45	44
Modelled	115	140	117	11	23	17	60	55

To Marlboro Station
August 2012



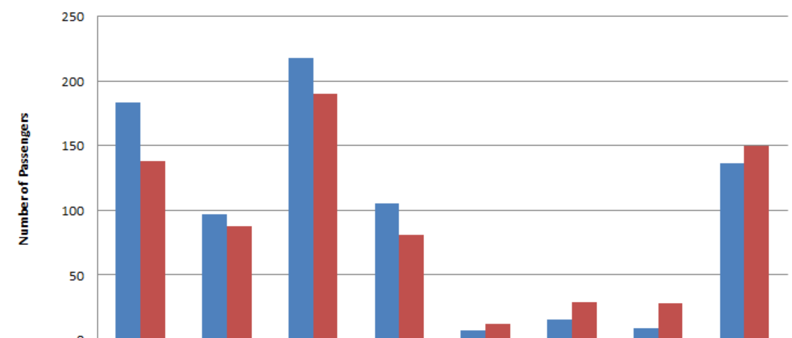
	Hatfield	Pretoria	Centurion	Midrand	Sandton	Rosebank	Park	Rhodesfield
Observed	36	23	46	10	2	2	7	41
Modelled	22	14	29	10	2	3	5	41

To Sandton Station
August 2012

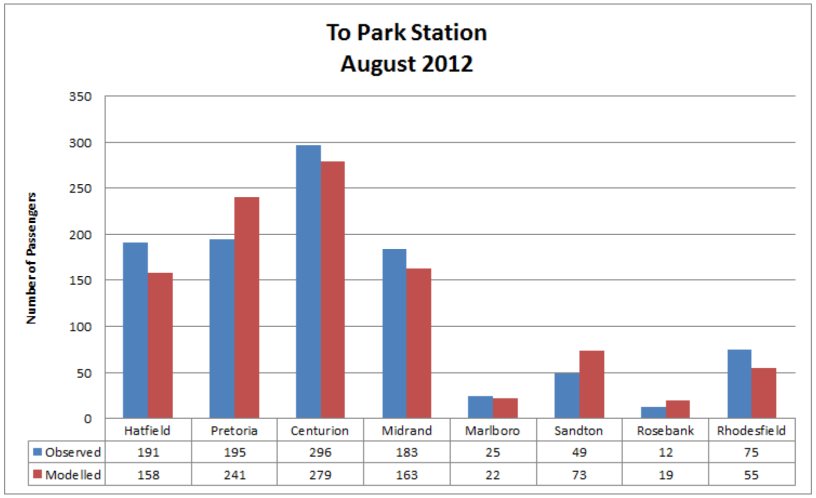


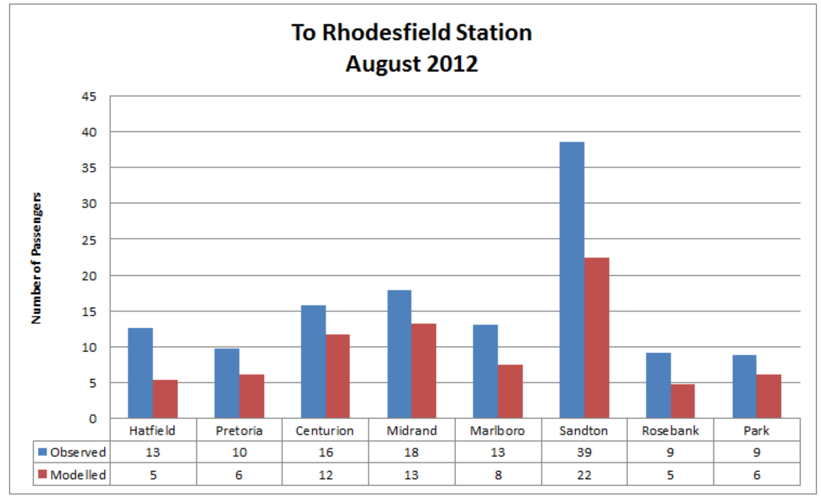
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Rosebank	Park	Rhodesfield
Observed	343	181	482	166	30	26	115	537
Modelled	278	211	467	140	25	35	185	552

To Rosebank Station
August 2012



	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Park	Rhodesfield
Observed	183	97	217	105	7	16	9	136
Modelled	138	88	190	81	12	29	28	150

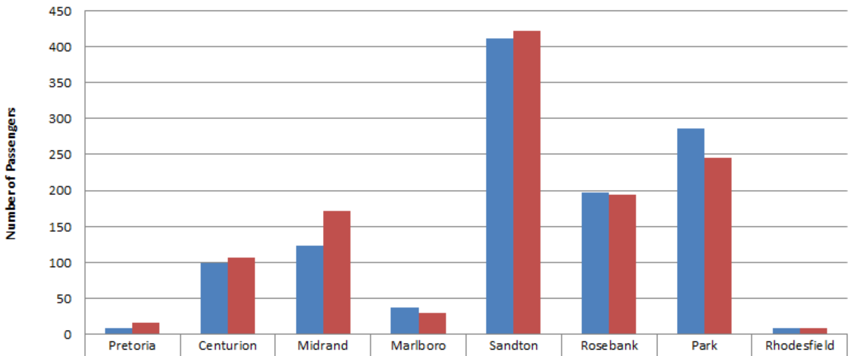

To Park Station
August 2012
Number of Passengers
350
300
250
200
150
100
50
0
Hatfield Pretoria Centurion Midrand Marlboro Sandton Rosebank Rhodesfield
Observed 191 195 296 183 25 49 12 75
Modelled 158 241 279 163 22 73 19 55


To Rhodesfield Station
August 2012
Number of Passengers
45
40
35
30
25
20
15
10
5
0
Hatfield Pretoria Centurion Midrand Marlboro Sandton Rosebank Park
Observed 13 10 16 18 13 39 9 9
Modelled 5 6 12 13 8 22 5 6

APPENDIX D

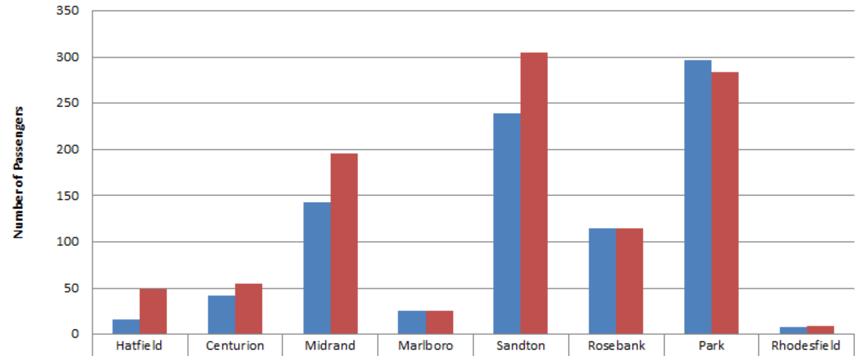
Appendix D contains the various station specific patronage for Scenario 1.

From Hatfield Station
August 2013



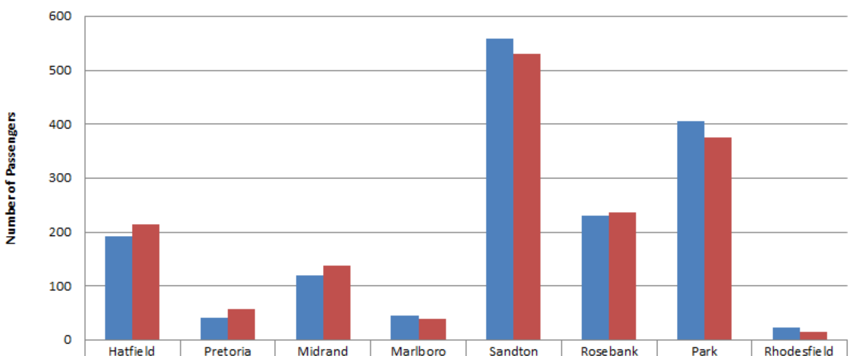
	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	9	99	124	37	411	198	286	8
Modelled	16	106	172	30	422	194	245	8

From Pretoria Station
August 2013



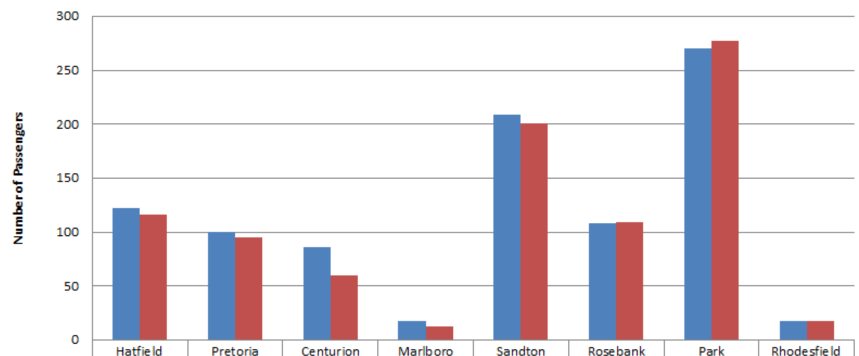
	Hatfield	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	16	42	142	26	239	114	296	8
Modelled	49	55	195	25	305	114	284	9

From Centurion Station
August 2013

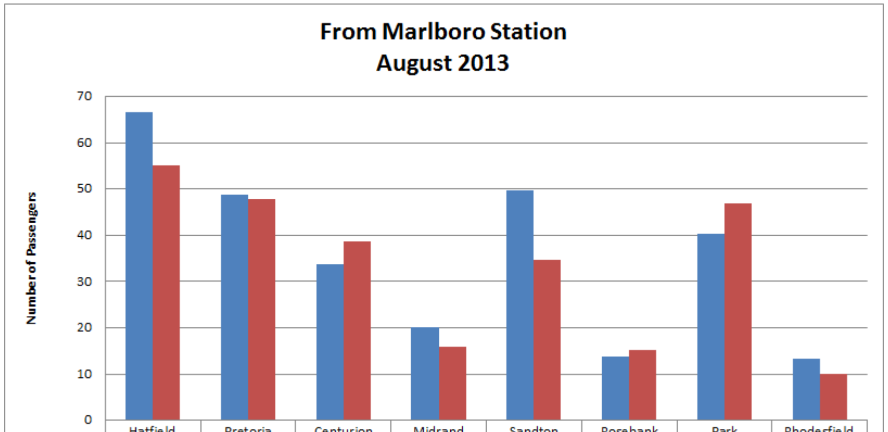


	Hatfield	Pretoria	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	192	42	119	45	558	231	405	23
Modelled	215	57	139	39	531	237	376	16

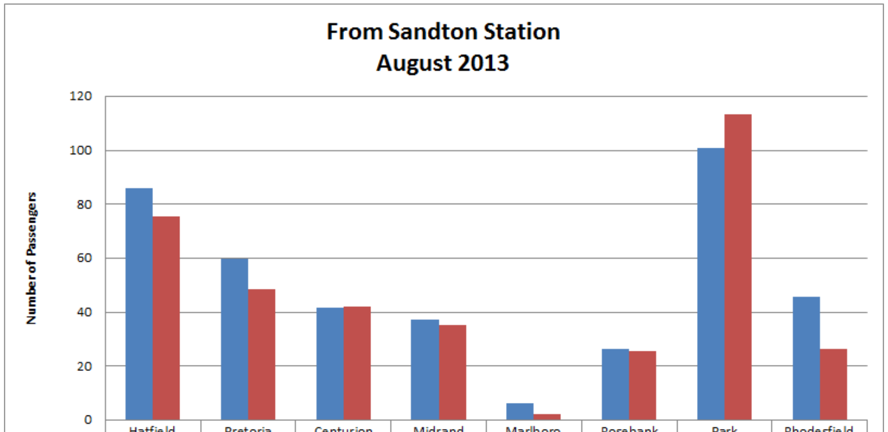
From Midrand Station
August 2013



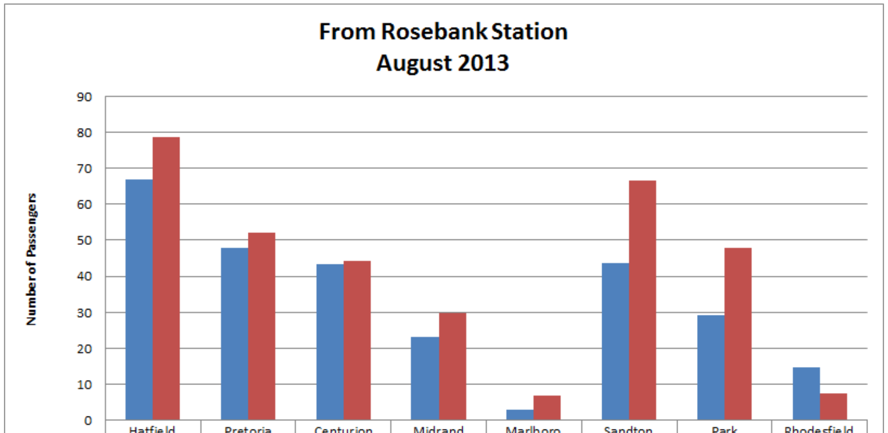
	Hatfield	Pretoria	Centurion	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	123	100	86	18	209	108	270	18
Modelled	116	95	60	13	201	110	278	18



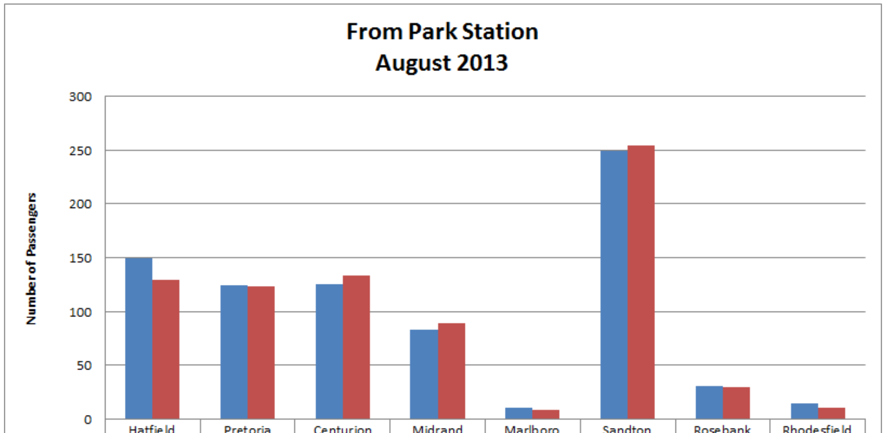
	Hatfield	Pretoria	Centurion	Midrand	Sandton	Rosebank	Park	Rhodesfield
Observed	67	49	34	20	50	14	40	13
Modelled	55	48	39	16	35	15	47	10



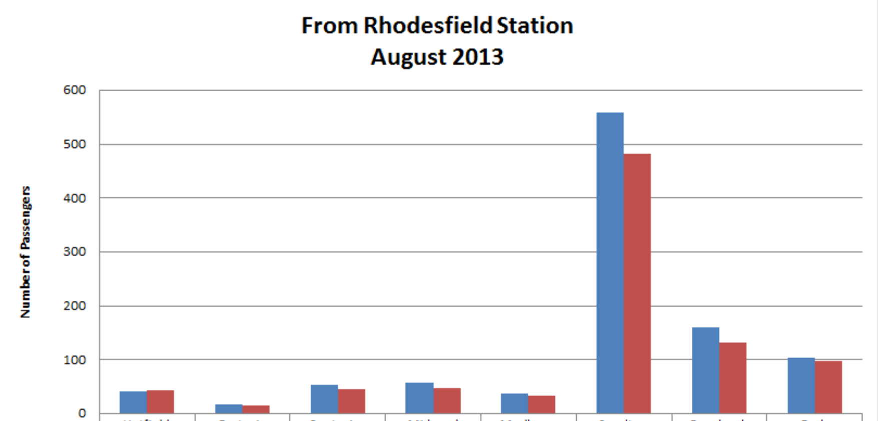
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Rosebank	Park	Rhodesfield
Observed	86	60	42	37	6	27	101	46
Modelled	76	48	42	35	2	26	113	27



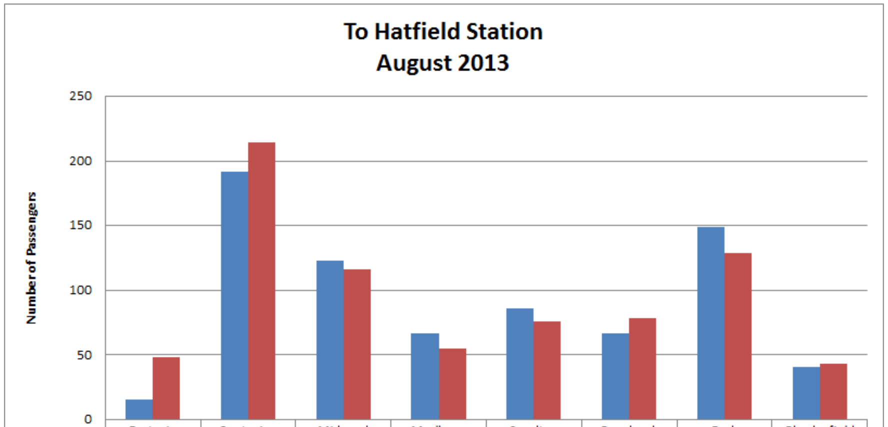
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Park	Rhodesfield
Observed	67	48	43	23	3	44	29	15
Modelled	79	52	44	30	7	67	48	7



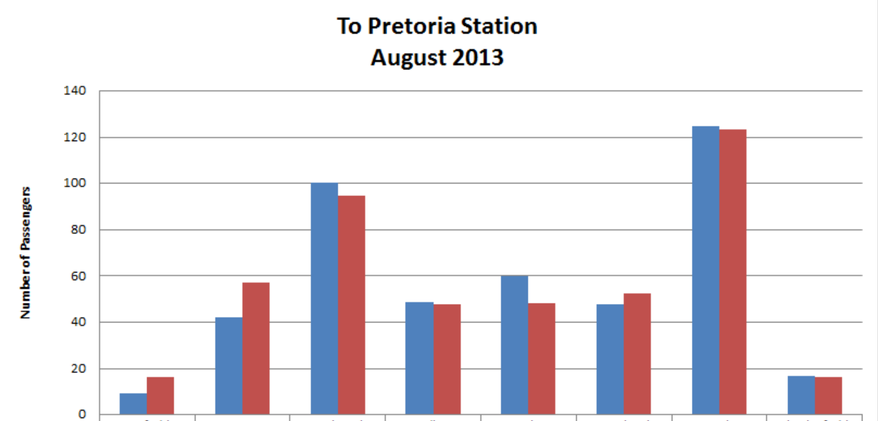
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Rhodesfield
Observed	149	125	126	84	11	250	31	15
Modelled	129	123	133	89	9	254	30	11



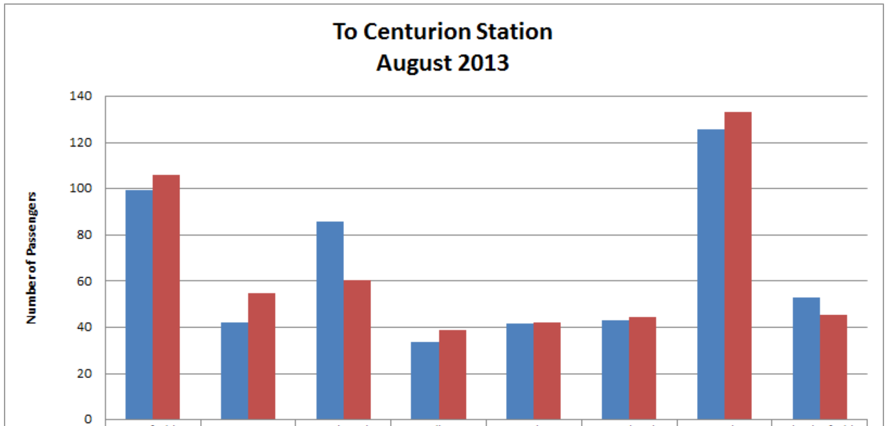
	Hatfield	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park
Observed	41	17	53	57	37	558	161	103
Modelled	43	16	46	48	34	483	132	99



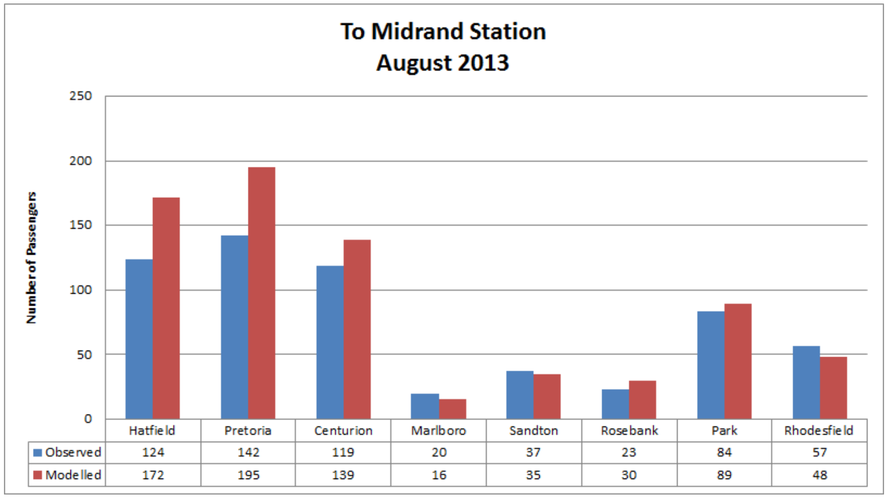
	Pretoria	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	16	192	123	67	86	67	149	41
Modelled	49	215	116	55	76	79	129	43

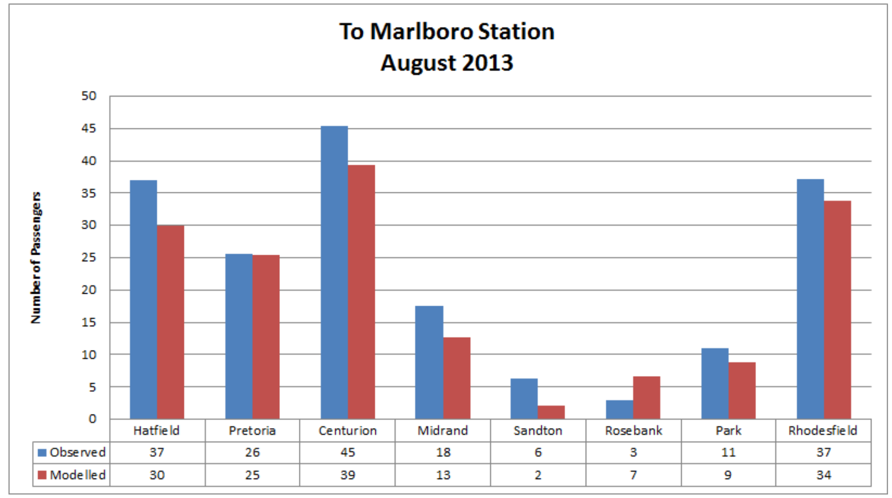


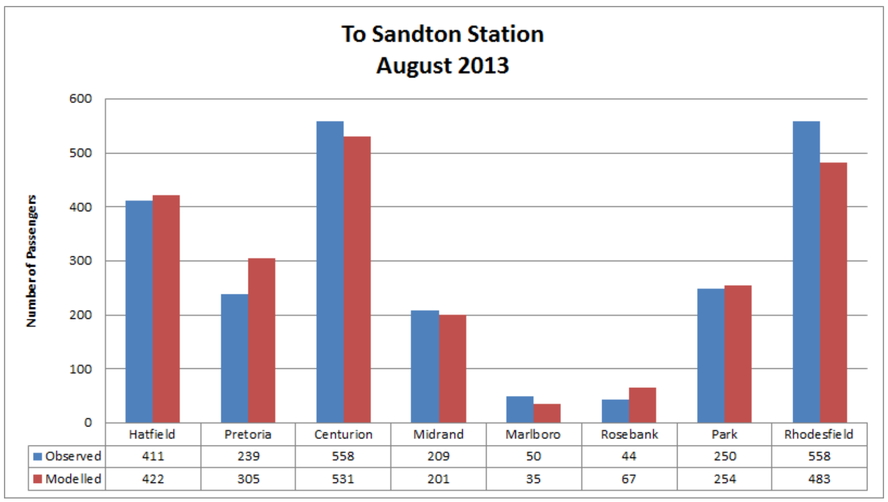
	Hatfield	Centurion	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	9	42	100	49	60	48	125	17
Modelled	16	57	95	48	48	52	123	16

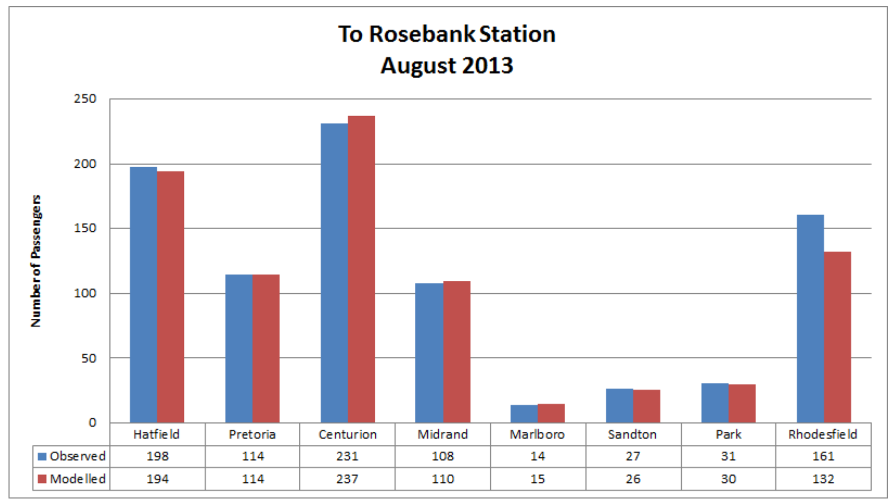


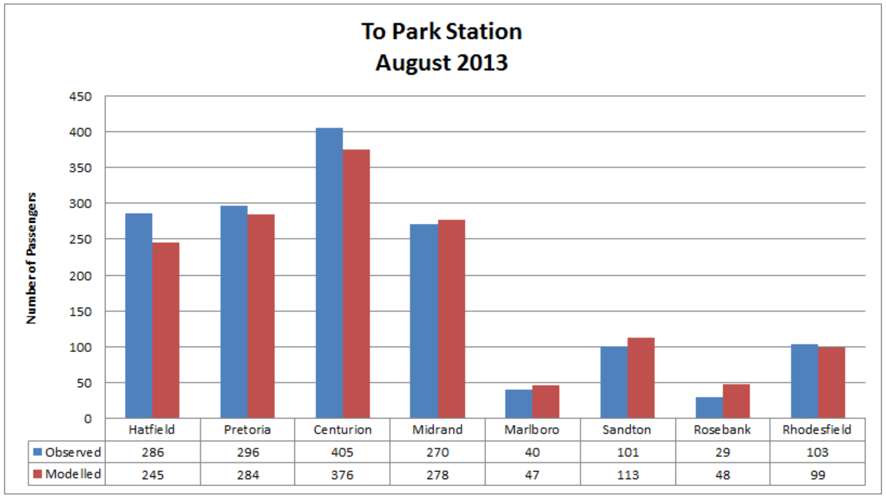
	Hatfield	Pretoria	Midrand	Marlboro	Sandton	Rosebank	Park	Rhodesfield
Observed	99	42	86	34	42	43	126	53
Modelled	106	55	60	39	42	44	133	46

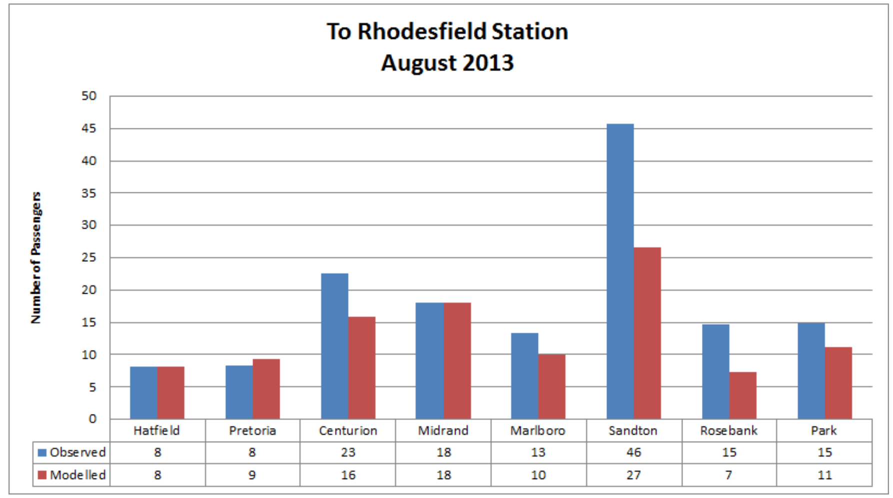

To Midrand Station
August 2013
Number of Passengers
250
200
150
100
50
0
Hatfield Pretoria Centurion Marlboro Sandton Rosebank Park Rhodesfield
Observed 124 142 119 20 37 23 84 57
Modelled 172 195 139 16 35 30 89 48


To Marlboro Station
August 2013
Number of Passengers
50
45
40
35
30
25
20
15
10
5
0
Hatfield Pretoria Centurion Midrand Sandton Rosebank Park Rhodesfield
Observed 37 26 45 18 6 3 11 37
Modelled 30 25 39 13 2 7 9 34


To Sandton Station
August 2013
Number of Passengers
600
500
400
300
200
100
0
Hatfield Pretoria Centurion Midrand Marlboro Rosebank Park Rhodesfield
Observed 411 239 558 209 50 44 250 558
Modelled 422 305 531 201 35 67 254 483


To Rosebank Station
August 2013
Number of Passengers
250
200
150
100
50
0
Hatfield Pretoria Centurion Midrand Marlboro Sandton Park Rhodesfield
Observed 198 114 231 108 14 27 31 161
Modelled 194 114 237 110 15 26 30 132


To Park Station
August 2013
Number of Passengers
450
400
350
300
250
200
150
100
50
0
Hatfield Pretoria Centurion Midrand Marlboro Sandton Rosebank Rhodesfield
Observed 286 296 405 270 40 101 29 103
Modelled 245 284 376 278 47 113 48 99


To Rhodesfield Station
August 2013
Number of Passengers
50
45
40
35
30
25
20
15
10
5
0
Hatfield Pretoria Centurion Midrand Marlboro Sandton Rosebank Park
Observed 8 8 23 18 13 46 15 15
Modelled 8 9 16 18 10 27 7 11

8. BIBLIOGRAPHY

Automobile Association of South Africa. (n.d.). *AA*. Retrieved August 2013, from http://www.aa.co.za

Ben-Akiva, M. (2002). *Hybrid Choice Models: Progress and Challenges.* Kluwer Academic Publishers.

Ben-Akiva, M. E. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* Massachusetts: MIT Press.

Bierlaire, M. (1995). *Mathematical models for transportation demand analysis.* Universitaires Notre-Dame de la Paix de Namur: Facultes des Sciences Departement de Mathematique.

Bin Miskeen, M. A., Alhodairi, A. M., & Bin O.K. Rahmat, R. A. (2013). Modeling of Intercity Travel Mode Choice Behavior for Non-Business Trips within Libya. *Research Journal of Applied Sciences, Engineering and Technology 7(3): 442-453, 2014*, 442-453.

Canty and Associates LLC. (1999). *Weather Base*. Retrieved September 2013, from http://www.weatherbase.com/

Celikoglu, H. B. (2007). *Calibration of Logit Modal Split Models with Feed Forward Back-Propagation Neural networks.* Istanbul: Technical University of Istanbul, Faculty of Civil Engineering, Division of Transportation.

Furnish, P., & Wignall, D. (2009). *Making the Most of Models.* Wellington: Victoria Transport Policy Institute.

Gautrain Management Agency. (2009). *Gautrain*. Retrieved 2013, from Gautrain: http://www.gautrain.co.za/

Guo, Z., Wilson, N. H., & Rahbee, A. (2007). *The Impact of Weather on Transit Ridership in Chicago.* Cambridge: Massachusetts Institute of Technology.

Gupta, J., & Shah, N. H. (2012). Origin Destination Transportation Models: Methods. *International Journal of Mathematical Sciences & Applications*, 819-825.

Jewett, S. (2004). *Physics for Scientists and Engineers with Modern Physics.* California: Thomson, brooks and Cole.

Khan, O. (2007). *Modelling Passenger Mode Choice Behavior Using Computer Aided Stated Preference Data.* Queensland: Queensland University of Technology.

Kohn, H. (2000). *Factors Affecting Urban Transit Ridership.* Canada.

Koppelman, F. S., & Bhat, C. (2006). *A Self Instructing Course in Mode Choice Modeling: Mutinomal and Nested Logit Models.*

Maitra, B., Ghosh, S., Das, S. S., & Boltze, M. (2013). Effect of model specification on valuation of travel attributes:. *Journal of Transport Literature*, 8-28.

McNally, M. G. (2007). *The Four Step Model.* Irvine: Institute of Transportation Studies University of California.

Metropolitan Washington Council of Goverments. (n.d.). Retrieved October 2013, from Metropolitan Washington Council of Goverments: http://www.mwcog.org/

Ortuzar, J., & Willumsen, L. G. (1990). *Modelling Transport Second Edition.* Chichester: Wiley.

Time and Date AS. (1997). *Time and Date*. Retrieved September 2013, from http://www.timeanddate.com/

Wegmann, F., & Everett, J. (2012). *Minimum Travel Demand Model Calibration and Validation Guidelines for the State of Tennessee Update 2012.* Knoxville: University of Tennessee (Center for Transportation Research).

Zhao, F., Li, M.-T., Chow, L.-F., Gan, A., & Shen, D. (2002). *FSUTMS Mode Choice Modelling: Factors Affecting Transit use and Access.* Florida: University of South Florida.