A STUDY TO MEASURE THE ACCURACY OF SPEED DATA REPORTED BY FLOATING CAR DATA IN RURAL AREAS

by

Antony Mugambi Kiautha

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

Supervisor: Mrs. Megan Bruwer

March 2023

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 (save to the extent explicitly otherwise stated), 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 2023

ABSTRACT

Intelligent Transportation Systems (ITS) technologies are becoming routine worldwide, offering newer methods of obtaining traffic data. Recently, on-board vehicle navigation devices have offered Global Positioning System (GPS) data spanning an entire road network referred to as Floating Car Data (FCD). The accuracy of FCD is critical to evaluate traffic information and inform stakeholders involved in transportation planning in different regions. This study was conducted to measure the accuracy of speed data reported by commercial FCD in rural areas.

The study compared benchmark speed data and TomTom historical speeds on 12 rural routes and 7 urban routes in the Western Cape in South Africa. The benchmark data was provided by the Western Cape Department of Transport, collected through loop inductance. Data collected was analysed for weekdays (Monday to Friday) in the month of February 2019 and aggregated over 1-hour time intervals. This was done to give a true representation of traffic patterns and analyse traffic patterns before the COVID-19 pandemic.

To evaluate accuracy, probe penetration rate, signed error bias (SEB), average absolute (AASE) and signed error were determined. The minimum probe penetration rate considered for accuracy measurements was 4% and the maximum allowable errors for SEB, AASE and signed error were ±7.5 km/hr, ±10 km/hr and ±10 % respectively. The SEB and AASE accuracy measures showed that FCD reported speeds were consistently lower in the urban areas while some routes in the rural areas recorded higher FCD speed estimate than the benchmark data. Probe penetration rates for both urban and rural areas indicated that a high probe penetration rate is not directly proportional to a high level of speed data accuracy.

This study is relevant because evaluation of commercial FCD accuracy has not been examined in rural areas and it is an informative research to assist in improving the transportation industry across the globe and especially in developing countries.

OPSOMMING

Intelligente vervoerstelsels (ITS)-tegnologieë word wêreldwyd benut, en bied nuwe metodes aan om verkeersdata te bekom. Voertuignavigasietoestelle het redelik onlangs Global Positioning System (GPS) data begin verskaf wat oor die hele padnetwerk beskikbaar is: Floating Car Data (FCD). Die akkuraatheid van FCD is van kritieke belang om verkeersinligting te evalueer vir gebruik van belanghebbendes wat betrokke is by vervoerbeplanning. Hierdie studie is uitgevoer om die akkuraatheid van spoed-data, wat deur kommersiële FCD in landelike gebiede gerapporteer word, te meet.

Die studie het maatstafspoeddata en TomTom-historiese snelhede op 12 landelike roetes en 7 stedelike roetes in die Wes-Kaap in Suid-Afrika vergelyk. Die maatstafdata is verskaf deur die Wes-Kaapse Departement van Vervoer, ingesamel deur induksie-lusse. Data wat ingesamel is, is ontleed vir weeksdae (Maandag tot Vrydag) in Februarie 2019 en saamgevoeg oor 1-uur tydsintervalle. Dit is gedoen om 'n ware voorstelling van verkeerspatrone te gee en verkeerspatrone voor die COVID-19-pandemie te ontleed.

Om akkuraatheid te evalueer, is toestelpenetrasiekoers, “signed error bias” (SEB), die gemiddelde absolute spoed verandering (AASE), en persentasie vout bepaal. Die minimum toestelpenetrasiekoers, wat vir akkuraatheidsmetings oorweeg is, was 4% en die maksimum toelaatbare foute vir SEB, AASE en persentasie fout was ±7.5 km/h, ±10 km/h en ±10 % onderskeidelik. Die SEB- en AASE-akkuraatheidsmaatreëls het getoon dat FCD-gerapporteerde snelhede laer in stedelike gebiede was, terwyl sommige roetes in die landelike gebiede hoër FCD-spoede as die maatstafdata aangeteken het. Toestelpenetrasiekoerse vir beide stedelike en landelike gebiede het aangedui dat 'n hoë toestelpenetrasiekoers nie direk eweredig is aan 'n hoë vlak van spoeddata-akkuraatheid nie.

Hierdie studie is relevant omdat evaluering van kommersiële FCD-akkuraatheid nie voorheen in landelike gebiede ondersoek is nie. Hierdie insiggewende navorsing kan help met die verbetering van die vervoerbedryf regoor die wêreld en veral in ontwikkelende lande.

DEDICATION

“To my wife, Mrs. Wanja Mugambi”

ACKNOWLEDGEMENTS

I firstly thank God for enabling me to successfully finish my masters research project. For giving me the will and perseverance to work on and attain this major achievement in my life.

A special thanks goes to my family for the constant messages of encouragement and knowing that you had my back at every stage of this research project. My wife, Wanja Mugambi, my parents Mr. Kiautha Arithi and Mrs. Zipporah Kiautha, my parents-in-law Prof. Isaiah Wakindiki and Mrs. Miriam Chabaari, my siblings Brian G. Kiautha and Mrs. Brenda Gikunda, Maureen Fiona Kiautha, Muriithi Chabaari and Muthee Chabaari, I am truly thankful.

I would also like to sincerely thank my supervisor Mrs. Megan Bruwer. Thank you for the guidance and support since the first day I walked into your office. You have played a pivotal role in completion of this research paper.

I also thank the South African Radio Astronomy Observatory (SARAO) a branch of the National Research Fund (NRF) for sponsoring the last year of my master’s programme.

Despite working remotely most of the time due to the hybrid learning that was introduced as a result of the Covid-19 pandemic, I thank the Civil Engineering staff and students for their input in this research. More so the Stellenbosch Smart Mobility Lab (SSML) for making it possible to have access to the TomTom data used in this research project.

Lastly, I would like to thank TomTom and the Western Cape Department of Transport for providing the data used in this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 : INTRODUCTION

1.1 BACKGROUND

Traffic data measurements from conventional in-situ traffic sensor technologies have been influential methods in observing traffic patterns over the past decades. These traditional methods of data collection are categorised into non-intrusive and intrusive methods. Intrusive methods are comprised of a data recorder and a sensor positioned either in or on the road surface. The most common intrusive techniques include magnetic loops and piezoelectric sensors. The non-intrusive methods are techniques based on observing through sensors located on the side of the road. These include, passive and active infrared sensors, microwave radar sensors, passive magnetic sensors, video image detectors and manual counts (Guillaume, 2008). Since these conventional data collection techniques work on a spot-measurement basis (data is collected at a particular spot on the road network), collected traffic data is pertinent to a limited number of locations on the road network and normally collected under established guidelines avoiding geometric constraints and away from intersections.

Intelligent Transportation Systems (ITS) technologies are becoming routine worldwide, offering newer methods of obtaining traffic data. Recently, on-board vehicle navigation devices have offered Global Positioning System (GPS) data spanning an entire road network. The traffic data which can be observed from this expansive source of GPS-based data are referred to as probe vehicle data or Floating Car Data (FCD) (Jurewicz, et al., 2017).

In transportation operations, there is a growing number of drivers who are willing to pay service providers to have access to relevant real-time traffic information. Travel information provided to the drivers via communications and display technologies introduce a new stage in transportation infrastructure management strategies. Such information gives a guidance to the overall performance of the system and can assist drivers in decision-making. Informed road users can make logical decisions regarding time of arrival, route choice and trip length based on travel time reliability and travel time (Cohen & Christoforou, 2015).

According to Kessler, et al. (2018) traffic data providers such as TomTom, Google, INRIX and HERE are utilising floating-car based technologies to offer real-time information such as traffic delays to drivers as well as to traffic authorities. As GPS-enabled devices become more widespread, the numbers of probe vehicles increase and this technology becomes highly considered for traffic management applications. As a result, traffic data is required to be reliable, timely and accurate (Guillaume, 2008).

1.2 RESEARCH PROBLEM

In South Africa, no studies have examined if commercially available FCD is tolerably accurate to measure speeds in rural areas. This study seeks to research the acceptability of FCD speeds in rural areas and provide guidance on its adoption to transportation operations and planning purposes.

The South African road network is made up of approximately 754 600 kilometres of roads (NT, 2006). Road authorities need accurate and reliable traffic data to plan long-term transportation strategies and effectively manage extensive road network. Such data is needed for the design of planned new roads, inform on remedial works of existing roads and general network level management (Smith & Visser, 2001). Table 1 provides a breakdown of the road network division according to the jurisdiction. The majority of the South African road network is rural.

The high number of rural road kilometres and the vastness of the road network in South Africa makes the continuous deployment of in-situ traffic sensors as methods of data collection too costly. Factors contributing to the expense of static sensors are the installation and maintenance cost of the sensors and their communication channels (Yatskiv, et al., 2013). To cover the extensive rural network, it will be greatly beneficial to use FCD which does not require in-situ traffic sensor infrastructure.

Table 1: Approximate length of road networks in South Africa (NT, 2006)

Kilometres	**Length**	**Percentage**
Surfaced National toll and non-toll roads	15 600	2,1%
Surfaced provincial roads	348 100	46,1%
Unproclaimed rural roads	222 900	29,5%
Metropolitan, Municipal and other	168 000	22,3%
Total	**754 600**	**100,0%**

Limited research has been conducted on transportation in South Africa and especially on rural roads with 40% of South Africa’s population dwelling in rural areas (DoT, 2015; Sewell, et al., 2019). The maintenance of rural roads and development of adequate transport infrastructure remains a challenge and it is important to establish a road management system that is effective (Sewell, et al., 2019).

Rural road development projects receive a relatively small share of the country's national budget for road upgrading influenced by circumstances such as lack of accountability, corruption as well as biased policies (Sewell, et al., 2019). With such challenges being prevalent, engineers should make use of intelligent transportation systems to be able to collect traffic data and improve operations on the rural road network (Gwara & Andersen, 2018). Research done by Vanderschuren (2006) on ITS in the South African context, elucidated that ITS technologies can benefit South Africa and can play a dominant role in improving sustainability of traffic operations.

Advancements in technology and provision of various ITS modes of data collection, have provided the gateway to acquire tremendous amounts of traffic data. The FCD method of data collection has proved to be a potential source of obtaining traffic information at a low cost (Chandra, et al., 2014). A concern arises regarding the accuracy and reliability of FCD. The number of probe vehicles, usually quantified as the penetration rate (the proportion of probe vehicles along a traffic link compared to the total traffic volume during an average time interval) is a determining factor for traffic data accuracy in a transportation system (Chandra, et al., 2014). The penetration rate and availability of various sources of unprocessed traffic data are factors for possible variations in the quality of FCD data globally (Gwara & Andersen, 2018).

To utilize FCD in transportation planning and operations in rural areas in South Africa, evaluation of traffic data accuracy and quality has to be conducted. The validation of the traffic data is assessed against reference data also known as benchmark data that is assumed to provide accurate values of speed for comparison. Reference data have been obtained from in-situ traffic sensors, such as those used by provincial departments of transportation (DOTs) on major freeways, highways, and arterials (Adu-Gyamfi, et al., 2017).

In South Africa, most traffic data are gathered via spot speed measurements such as inductance loop detectors and roadside video cameras. Recently, developments in methods of traffic data collection have also been enabled by the utilisation of Open Road Tolling (ORT) on freeways in Gauteng province for example, as an alternative source of reference data (Gwara & Andersen, 2018).

1.3 LIST OF OBJECTIVES

The research objectives considered for this study are as follows:

i) Provide information on evaluating the accuracy of traffic speed data.
ii) Collect speed information from roadside sensors and FCD along a number of rural and urban roads.

iii) Compare speeds obtained from traditional road sensors and FCD to establish the accuracy of FCD reported speeds.

iv) Consider if levels of probe penetration influence accuracy of FCD speed estimates along rural roads.

v) Investigate if FCD is sufficient to provide adequate speed data for transportation planning along rural roads.

1.4 RESEARCH LIMITATIONS

The study area was focused on a limited number of locations selected along higher order roads in the rural areas of the Western Cape Province. The higher order arterial rural roads selected were of high provincial importance as they link urban areas and centres of development. The results concluded that FCD, limited to TomTom FCD, was accurate for the rural area counting stations selected for the study. While this gives an indication that TomTom FCD should be accurate in rural areas of South Africa, the wider acceptance of this accuracy cannot be guaranteed by this research due to the limited number of sites that formed part of the study area. Furthermore, the accuracy of TomTom FCD on lower order rural roads can also not be accepted based on this research. Further research should be conducted on lower order rural roads to determine their FCD accuracy. Acceptable accuracy results in other provinces of South Africa were not derived and a discussion was not included regarding that. Moreover, impact of geometric design on the accuracy results was not evaluated.

The required traffic data for this research was collected from TomTom and traffic stations collecting data through conventional in-situ traffic sensor technology. TomTom historical data was available through a licencing agreement between TomTom Africa (Pty) Ltd and the Stellenbosch Smart Mobility Laboratory (SSML) of Stellenbosch University.

Quotations for purchasing a single loop traffic monitoring logger were requested from Mikros Systems (Pty) Ltd. Mikros Systems (Pty) Ltd is a company based in South Africa that provides affordable and reliable traffic monitoring products. A RAKTEL 8010 universal traffic event logger was considered as the product of choice viable for this study. However, due to financial constraints, it was costly to procure the traffic data logger. The reference data was then requested from the Western Cape Department of Transport who have a network of in-situ sensor technology in rural and urban roads around the province.

1.5 RESEARCH OUTLINE

The research outline for this study is as follows;

- Chapter 1 – Introduction
- Chapter 2 – Literature review
- Chapter 3 – Methodology
- Chapter 4 – Data collection
- Chapter 5 – Data analysis and Results
- Chapter 6 – Conclusion and Recommendations
- References

CHAPTER 2 : LITERATURE REVIEW

2.1 INTRODUCTION

The objective of this chapter is to look at the existing literature related to the evaluation of the accuracy of traffic speed data. The chapter commences by discussing the general functional road classification system. Thereafter, a discussion is done encompassing methods of collecting traffic data and a consequent discussion on the guidelines of assessing the accuracy of traffic speed data. To conclude the chapter, selected research studies with the aim of assessing traffic data quality are presented.

2.2 FUNCTIONAL ROAD CLASSIFICATION

2.2.1 Access and mobility functions

Transport networks play a vital role in the well-being of the society and the economy of the country. Therefore, it is important to provide an integrated transport system which is reliable, efficient and effective that will ensure sustainable social and economic development of the country (COTO, 2012).

A road network makes it possible for the road-based transportation of goods and people from one destination to another. The main functions that a road network serve are access and mobility: access to specific land uses, and mobility between destinations. Access to the road network is provided at intersections, driveways and interchanges that provide ingress to specific roads while mobility is the efficiency of traffic to move at relatively high speeds with minimal interruptions along roads (COTO, 2012).

A number of road classification methods are based on access and mobility. South Africa has adopted the six-class functional road classification system that is used for both urban and rural road networks. This classification system promotes (DoT, 2007):

- Integration of road network system
- An avenue for adequate discussions with respect to ownership and administrative responsibility for the road system among road authorities
- Appropriate division of liabilities for the entire route network
- Getting rid of road backlogs and unclassified roads, and
- Equality in funding roads that is consistent with government policies.

2.2.2 South African Functional Road Classification System

The six-class urban and rural road classification system is tabulated in Table 2. The classification is categorised according to the road functionality as either mobility road or access road.

Table 2: Functional road classification system (COTO, 2012)

Number	Function	Description
Class 1 Class 2 Class 3	Mobility	Principal arterial Major arterial Minor arterial
Class 4 Class 5 Class 6	Access/ Activity	Collector street Local street Walkway

To adequately classify the road network and distinguish between the different classes, the following criteria are employed (COTO, 2012):

i) Functionality of the road. Mobility roads link centres of development in both rural and urban areas while Access streets give access to individual properties and link them to mobility roads, as they distribute traffic from the properties.

ii) Connectivity distance. Mobility roads are required for longer reach of connectivity while Access roads are required to be less than 1km before joining a mobility road. Access roads inhibit speeding especially in urban roads.

iii) Travel stage. Travelling is characterised by three stages, that is, local at the origin, through (destination not reached) and local at the destination. The local stages of travelling are served by access roads while mobility roads serve the "through" part of the trip.

Rural areas are characterised by sporadic developments requiring roads with longer travelling distance. Therefore, rural roads require more higher-class arterials than urban roads. However, both urban and rural roads have the same functional classes but considered at different scales. Table 3 depicts the different classes.

The rural environment consists of natural area and, extensive and intensive agricultural areas. The roads in these areas are ordinarily rural in design, built above the natural ground level to ensure drainage and provide for vehicular movement with limited pedestrian presence. Operating speeds will typically vary from 80 – 120 km/hr in rural areas, depending on the standards applied and the road hierarchy. On the other hand, urban areas are characterised by dense commercial, institutional and residential development within activity nodes. The road grid pattern is quite fine to serve high-level access seeking traffic with operating speeds of about 50 km/hr depending on the road type (EDAT, 2002).

Table 3: Rural and Urban road classes (COTO, 2012)

	Rural Classes		**Urban Classes**
R1	Rural principal arterial*	U1	Urban principal arterial
R2	Rural major arterial*	U2	Urban major arterial
R3	Rural minor arterial*	U3	Urban minor arterial
R4	Rural collector road	U4	Urban collector road
R5	Rural local road	U5	Urban local road
R6	Rural walkway	U6	Urban walkway

*If preferred, the word “arterial” can be substituted by “distributor” for Rural Classes 1 to 3

In the case where a rural road enters an urban area, the rural classification is changed to an urban classification while the class is usually maintained. Changes in speed limits, approaches to major intersections or end of freeway zones are some of the factors that motorists encounter during the transition from a rural to an urban area. In situations where the rural road is maintained as a Bypass or Through-Way, the rural road continues with the same classification over the course of the road (COTO, 2012).

According to the Road Access Guidelines for South Africa (EDAT, 2002), road classification is divided into higher order arterials and lower order roads. The higher order arterials are inclusive of freeways, expressways and primary arterials while the lower order roads encompass the district distributors, local distributors and access roads. Figure 1 has a guideline representation of the movement function.

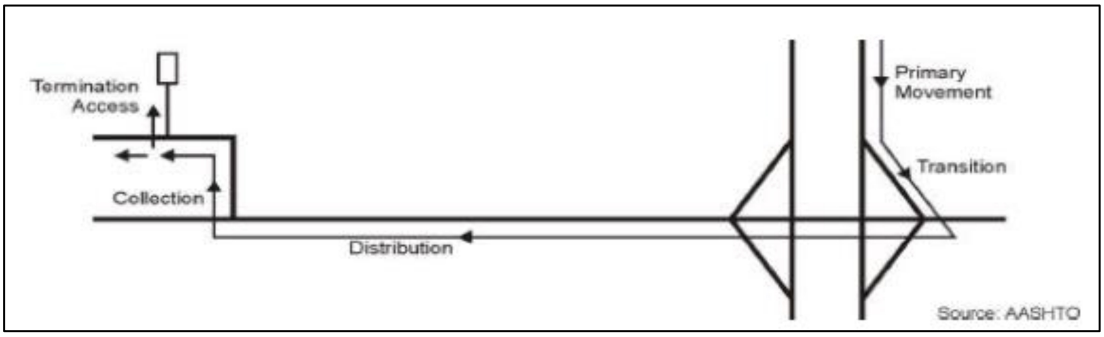


Figure 1: Hierarchy of Movement (EDAT, 2002)

2.3 METHODS OF TRAFFIC DATA COLLECTION

Traffic data is required to plan for traffic management operations, road maintenance, road design, and traffic forecasting among other functions.

Traffic data collection provides the basis for traffic problem identification, earlier traffic studies hypotheses confirmation, quantification of impact of changes for example on speed zones, and determination of the magnitude of improvements required to achieve favourable traffic conditions. Standardised traffic data collection and analysis assists in achieving adequate and reliable data, which is essential to any traffic engineering study (MnDOT, 2007).

Traffic data measurements from conventional in-situ traffic sensor technologies have had an influential role in traffic observations over the past decades. Traffic information gained from in-situ sensors ranges from speed data and traffic count to vehicle classification data. Vehicle classification consists of determining the percentage of vehicles of a particular class at a certain location in the traffic system (MnDOT, 2007). These traditional methods of data collection are categorised into intrusive and non-intrusive methods. Intrusive methods are comprised of a data recorder and a sensor placed either on or in the road surface. The intrusive methods of data collection are generally expensive due to their high installation and maintenance costs (Guillaume, 2008). Some of the widely used methods are presented hereafter.

- Magnetic loops

  Wire loops are installed in the surface of the road mostly in a square configuration. The loops generate a magnetic field that is disturbed by a moving vehicle. Detected information is transmitted to a roadside counting device from which traffic data can be obtained. The magnetic loop technique has been dominant over the years and it is a well-understood technology. Magnetic loops provide basic traffic parameters such as speed and volume while high frequency models of the technique providing vehicle classification data. Some of the disadvantages include possible compromised sensor installation in old pavements, susceptible to damage by road repairs and heavy vehicles and the loops are prone to high maintenance resulting from installation errors (Bennett, et al., 2005).

- Pneumatic road tubes
  The tubes are normally made of rubber and placed across a road lane. A pulse of air is created, recorded and processed by a roadside counter when a vehicle passes. This technique is seldom used since its efficacy is subject to traffic conditions, weather and temperature. Vehicles in low speed flows are hardly detected.

- Piezoelectric sensors
  This method is generally applicable to measurement of vehicle speeds, vehicle count and vehicle weight and classification. Installation of sensors requires grooving of the road surface and mounting the sensors. The sensors consolidate data by changing mechanical energy into electrical energy through deformation of the piezo electric material. This produces a change in the surface charge density for which a change in voltage is recorded between the electrodes.

The non-intrusive methods of data collection are techniques based on remote observations. The techniques do not cause traffic flow interference during installation and operation. However, operating conditions such as bad weather and traffic conditions serve as major drawback for the data measuring technique. The non-intrusive methods include (Bennett, et al., 2005):

- Video image detection
  Traffic data is recorded using video techniques such as trip line and tracking. A camera is mounted depending on the desired lane coverage. The camera utilises a microprocessor for video image analysis. Video image detection facilities are capable of monitoring traffic conditions, sampling traffic parameters such as spot speeds, and detecting and verifying traffic incidents (Turner, et al., 1998).
- Microwave radar
  Radio detecting and ranging (radar) is capable of sensing and detecting vehicles from a distance. A device directs high frequency radio waves and determines the delay of return signals. The signals are categorised as either pulsed, phase-modulated or frequency-modulated. The microwave radar records traffic data for simple vehicle classification, traffic count data and speed (Guillaume, 2008).

- Passive and active infrared

  Passive infrared devices function by measuring the infrared energy radiating from a detection zone while active infrared devices emit a reflective laser beam at the road surface and measure the time that the signal returns to the device. Reduction in time for a signal to return to the measuring device indicates a detected vehicle. Both passive and active infrared detectors can be utilised to record traffic speed, traffic count and vehicle classification data
- Ultrasonic and passive acoustic sensors

  Ultrasonic devices utilise sound energy whereby emitted ultrasonic sound energy hits a passing vehicle and the time is measured for the signal to return to the detection devices. Vehicle detection is actualised if the return signal is received after a shorter duration than the normal road surface background time.

  Passive acoustic sensors on the other hand, make use of sound waves. A series of microphones are aimed at the traffic stream. Sound detected from a passing vehicle is compared to a set of pre-programmed sonic signatures to identify various classes of vehicles. The passive acoustic and ultrasonic devices are employed to collect vehicle classification data, traffic count and traffic speed.

Other forms of data collection classified under the non-intrusive techniques of data collection are manual counting and off-roadway technologies described below.

- Manual observation

  Manual observation techniques entail people counting and recording vehicle classification information. Most of these surveys are done at single points and use hand-held devices for on-site recording. Electronic devices that have the capability of obtaining more information for analysis are available but relatively expensive to acquire. Average daily traffic (ADT) can be calculated from the recorded data.
- Off-roadway Technologies

  The off-roadway technologies are new technologies being considered for traffic data collection. Most common modes include usage of probe vehicles and remote sensing. Probe vehicles are advantageous in collecting travel time data while remote sensing is reported to be on its developmental stages (Bennett, et al., 2005).

A schematic representation of the data collection and traffic monitoring techniques is laid out in Figure 2.

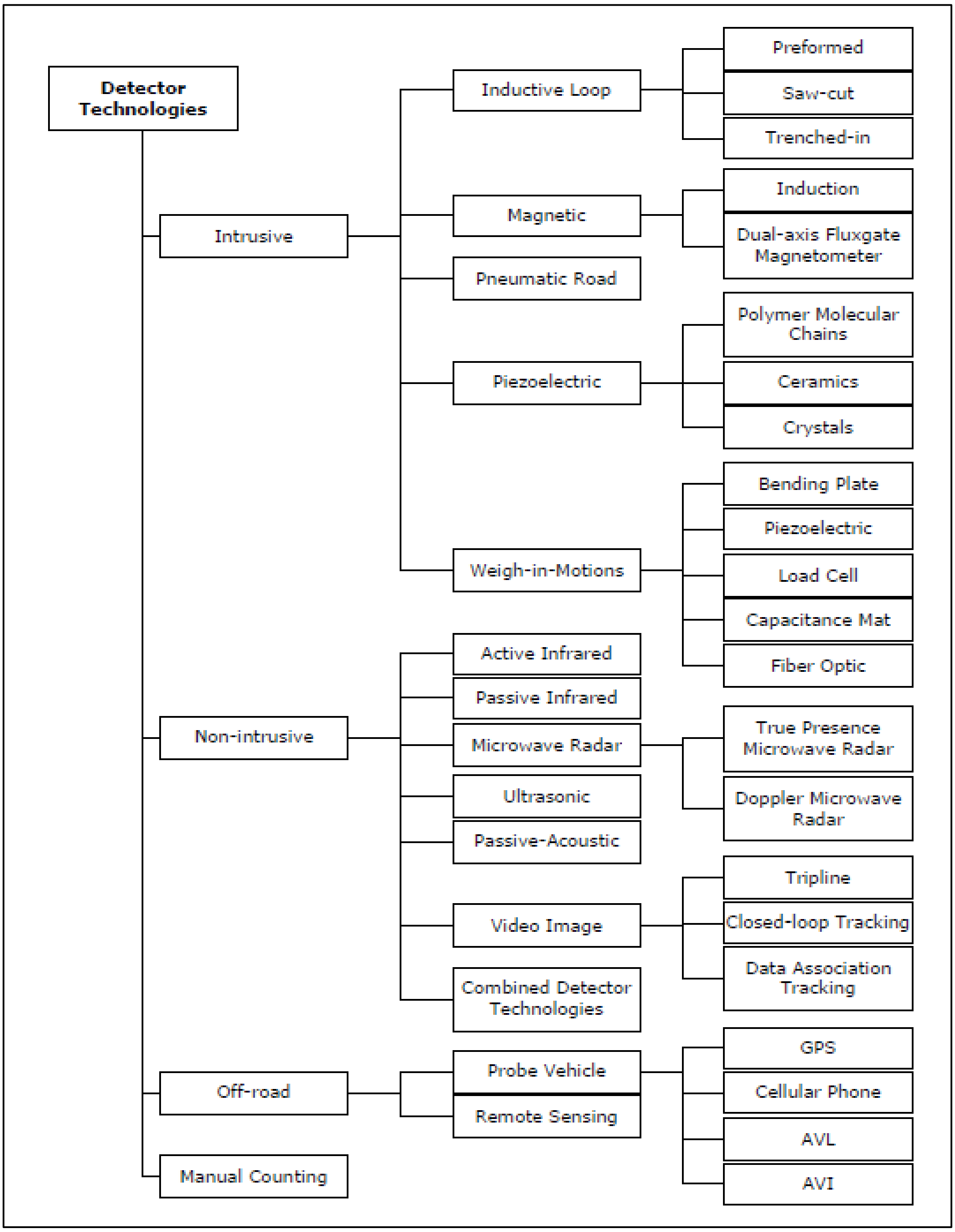


Figure 2: Data collection techniques (Bennett, et al., 2005)

2.4 FLOATING CAR DATA

Floating Car Data (FCD) is a source of traffic data collected in real-time via technological advancements over an entire road network. Traffic data that are collected include speed, direction of travel and car location. The data is conveyed anonymously to a central processing centre where it is extracted, and useful information such as traffic status communicated back to the drivers (FCD users) on the road (Guillaume, 2008).

Through rapid development in technology, there has been an emergence of ITS (intelligent transportation system) probe vehicles designed to collect traffic data instantaneously. Despite their pre-eminent purpose being in real-time traffic operations monitoring, route guidance and incident detection, the probes are often used for collecting travel time data (Turner, et al., 1998). ITS probe vehicles that collect travel time data are sometimes known as “passive” probe vehicles since they are in the traffic stream primarily for a different purpose.

Jurewicz et al. (2017) point out that substantial, accurate GPS probe data or FCD offered through vehicle navigation and GPS-enabled smartphone devices across the entire road network has been recorded over recent years. Their study proceeds to mention that there is a notable employment of FCD globally in road transport management for purposes of monitoring network performance and congestion, and in evaluating traffic flow improvement projects.

According to Kessler et al (2018), traffic data providers such as TomTom, Google, INRIX and HERE are utilising floating-car based technologies to offer real-time information such as traffic delays. With increasing number of probe vehicles, this technology is being highly considered for traffic management applications.

FCD are obtained from GPS devices and cellular (smartphones) probe data. These sources of FCD are discussed in the following sections.

2.4.1 GPS-based FCD

The GPS is a satellite-based navigation system consisting of 24 satellites, placed by the Department of Defence of the United States of America, that circle the earth in a precise orbit as they transmit signal information. A comparison is done for the time a satellite transmits a signal and when it is received. The time difference indicates the distance of the satellite from the receiver. Sequential distance measurements from a couple of satellites record the positional information of a user and display it on the unit’s electronic map (Jia, et al., 2013).

Vehicles with on-board GPS devices can log in signals and relay information such as the vehicle’s position and the time a location is passed along the road network. These signals can also be used to monitor location, speed and direction worldwide (Turner, et al., 1998). Figure 3 illustrates a typical trilateration process involved in GPS signal transmission.

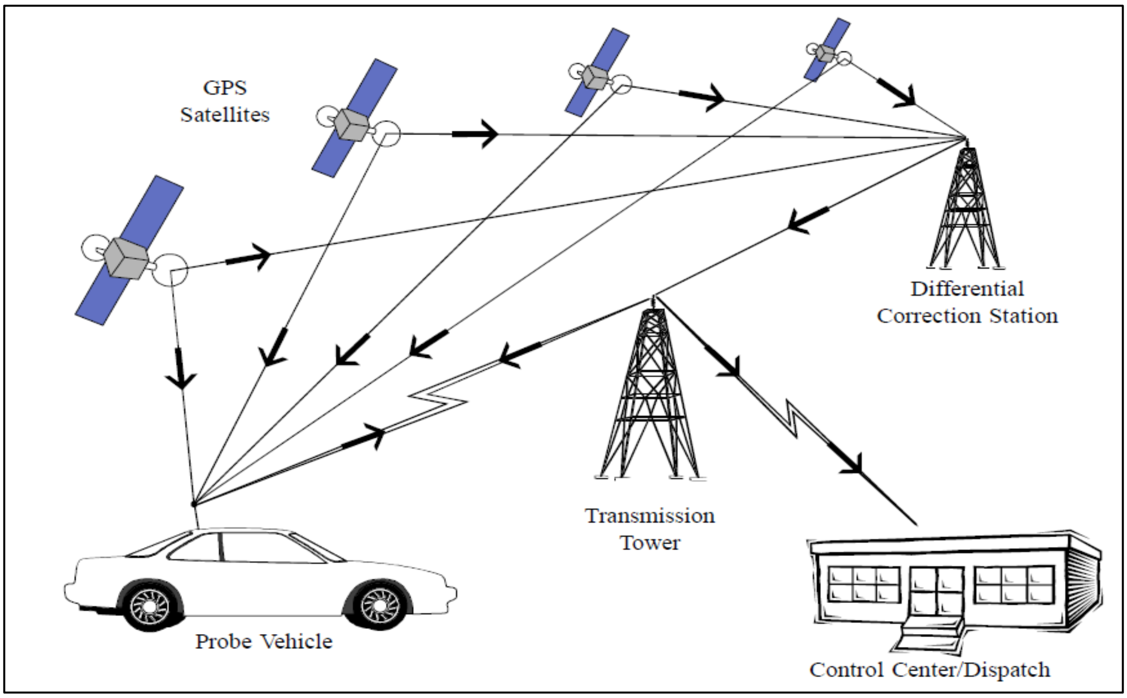


Figure 3: Configuration of a satellite-based ITS probe vehicle technique (Turner, et al., 1998)

The GPS based technique is widely used as a source of floating car data. This technique suffers limitation brought about by high equipment costs and a limited number of well-equipped vehicles compared to cellular probe data (Guillaume, 2008).

2.4.2 Cellular phone-based FCD

Cellular phone-based FCD relies on mobile phones present in a moving vehicle. The phones act as data providing sensors. Evaluation of positional information is done through triangulation techniques. It is worthwhile to note that the cellular phones should be turned on to transmit data. The road network is divided into segments and all observations from different cellular phones on a segment are computed to approximate the travel time for that specific segment (Bar-Gera, 2007).

Traffic data are collected continually hence offering a high coverage capability. No specific infrastructure is required to be built along the road network and neither is a special device required in the vehicle. The floating cellular data technique is less costly compared to other conventional detectors but requires complex algorithmns to extract accurate and meaningful traffic data that can be relayed to road users (Guillaume, 2008).

Common cellular phone systems include GPRS (General Packet Radio Service), CDMA (Code Division Multiple Access), UMTS (Universal Mobile Telecommunications System) and GSM (Global System for Mobile communications).

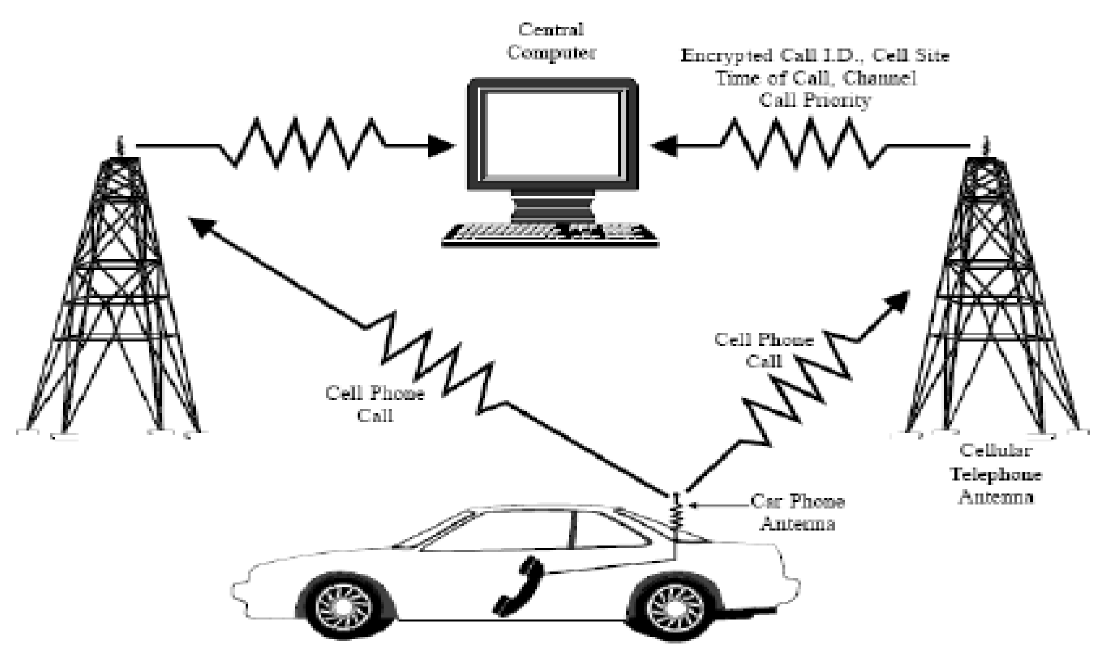


Figure 4: Mobile phone based FCD communication (Guillaume, 2008)**.**

Guillaume (2008) states that rural areas and motorways are favoured by traffic data collected from private vehicles or trucks while presence of a large number of taxi fleets and their on-board communication systems become advantageous sources in the case of urban regions. FCD has progressively become a complement source of large amounts of traffic data. Companies such as TomTom, INRIX and Be-Mobile have succeeded in providing FCD commercially (Altintasi, et al., 2019).

Despite vast amount of data being available from intelligent transport systems, uncertainty over data quality presents as a concern to ITS data. To try to reduce the discrepancies associated with the issue of different methodologies and procedures in evaluation, guidelines have been developed to assess data quality from various data sources. Data from ITS is often compared to other traditional data sources (Gwara & Andersen, 2018).

Gwara and Andersen (2017) mention some of the available guidelines for assessing traffic data quality. The guidelines are inclusive of*, “Traffic information benchmarking guidelines”, “Guidelines for evaluating the accuracy of travel time and speed data”, “Traffic data quality measurement”* and “*Travel time data collection handbook”*. The guidelines are not prescriptive and assist in giving guidance information for fair and consistent evaluation procedures producing comparable results.

Most available guidelines are based on evaluating the quality of real-time traffic data. Recent studies conducted by the Texas Transportation Institute (2012) and Gwara and Andersen (2018) have administered the same guidelines and have shown that guidelines for assessing real-time traffic data are adequate for the assessment and evaluation of historical traffic data quality.

2.5 TRAFFIC DATA QUALITY

In defining data quality, the researcher borrows from Turner (2002) in his defining and measuring traffic data quality report. Turner (2002) defines data quality as "*the fitness of data for all purposes that require it. Measuring data quality requires an understanding of all intended purposes for that data.*"

For decision makers to use traffic data in confidence, it should be sufficiently accurate. ITS infrastructure provides a platform to collect immense amount of data for immediate application in operations and for analytical purposes through Archived Data Management Systems (ADMS).

To understand the assessment of data quality, it is key to consider the different consumers and users, intended application of traffic data and the source of traffic data. The framework for traffic data quality measurements identifies the following three main types of data;

- Original data – the initial and actual data collected from various traffic data collection devices. The data could be real-time or archived.
- Archive data – Data contained in a storage database. It is derived from original source data and can be processed or present in its original raw state.
- Traveller information- Data provided as information to travellers. It is derived from original source data and often provided as real-time information (i.e. processed or aggregated).

The type of data to be used is dependent on the type of user and the application required.

Table 4 provides a tabulation of various traffic data consumers, data types and their applications. This is beneficial to comprehend since the type of data and its application helps in determining the methods required to calculate the quality measures and the thresholds for data quality evaluation.

Table 4: Distinctive applications of different data types and data users (Battelle, 2004)

Data Consumers	Source of Data	Applications
Traffic operators	Original data Archived data	Managing traffic and incidents
Archived data administrators	Original data	Administrating the database
Archived data users	Original data Archived data Archived processed data	Planning, modelling and analysis
Traffic data collectors	Original data Archived data	Planning on data collection, equipment calibration and traffic monitoring
Information Service Providers	Original real-time data	Circulating traveller information
Travellers	Traveller information	Pre – trip planning

Traffic data quality is expected to change as the original source data undergoes transformation from its collection phase to consumption by the end user. Data quality measures are employed and can vary according to different calculation procedures for various applications and users. Turner (2002) recommends the measures listed below and considers them the global fundamental measures for data quality assessment in traffic data applications;

- Accuracy

  Accuracy is the measure of the degree of agreement between a set of data and a reference source (Siripollawat, 2012). Accuracy measures the proportion to which estimated values and benchmark values are relative to each other. Accuracy measures are selected based on the evaluation scenario. Evaluation scenarios include link speeds, route travel time and congestion category (Turner, et al., (2011); Gwara and Andersen, (2018)). More discussion on accuracy measures is presented in section 2.6.1.8.
- Completeness

  Completeness is the degree of presence of data values in attributes such as traffic speed and traffic volume (Samuelson, 2011). Completeness is also referred to as availability and data is typically presented as number of values or items expressed in percentage

(Turner, et al., 2011). It is the number of available data values quantified as a percentage to the total number of expected data values.

- Validity

  Validity is the degree to which data values fall within the respective domain of acceptable values (Battelle, 2004). Data validation criteria is dependent on data application and it is often contrasting since the criteria can range from simple to complex rules (Gwara & Andersen, 2018). A common way to express data validity is to quantify the percentage of data values either passing or failing data validity criteria.
- Timeliness

  Timeliness is the measure of the degree to which data values are provided at the required time and can be expressed in relative or absolute terms. Timeliness can be beneficial in relaying real time traffic data from Traffic Management Centres to end users for incident location and updates. However, it is challenging to provide percipient information from timeliness in situations involving historical archived data in the analysis (Gwara & Andersen, 2018).
- Coverage

  Coverage is the degree to which data values in a sample accurately describe the entire collection that has been measured (Siripollawat, (2012), Battelle, (2004)). It gives an understanding to the proportion of the system being measured. In traffic data quality, coverage is reported in relative terms as a percentage and in absolute terms as kilometres of coverage.
- Accessibility

  Accessibility is interchangeably referred to as usability. Accessibility is the easiness with which data is obtained and processed by data consumers to meet their desired goals. Battelle (2004) defines accessibility as the needful choice by data consumers to the effectively retrieve and manipulate data. It is notable that accessibility is the only accuracy measure that can be described in quantitative (average time for data retrieval by data consumers) and qualitative (description of mechanisms through which data is obtained) terms.

In this study, the researcher looks to determine the accuracy of FCD. Therefore, focus is directed towards evaluating accuracy in FCD traffic data, which is discussed in the following section.

2.6 GUIDELINES FOR EVALUATING SPEED DATA ACCURACY

The intention of the guidelines for evaluating traffic data accuracy is to provide guidance on handling the required evaluation parameters. The guidelines ensure that the procedures for evaluating accuracy are consistent to provide comparable evaluation results. Fair and consistent results are helpful to government agencies to provide reliable and accurate traffic information.

The need to deliver quality traffic data has prompted the Federal Highway Administration (FHWA) to hold workshops primarily to devise an action plan to address problems on the quality of traffic data. The workshops resulted in a framework structured as a sequentially to calculate the data quality measures and assess the data quality (Battelle, 2004).

The researcher adopts the guidelines presented by Turner *et.al* (2011) since they contain more information compared to other relatable documents on this topic. Other closely related documents include *"Traffic Information Benchmarking Guidelines"*, *"USDOT Section 1201 Requirements" and* "*Traffic Data Quality Measurement"*.

Turner *et al.* (2011) clearly depicts three main steps involved in the process of accuracy evaluation:

1. Developing a data accuracy evaluation plan
2. Collecting and reducing reference data
3. Computing and reporting accuracy measures

The following sections detail each of these steps with reference to Turner *et al.* (2011).

2.6.1 Developing an evaluation plan

An evaluation plan serves as the initial phase in data quality assessment. A well-developed plan defines the scope and context and seeks to answer: "What information is being evaluated?" (Turner, et al., 2011). The evaluation plan should include methodological processes to calculate data quality measures taking into consideration data collection parameters such as accuracy measures, time intervals, links to evaluate, benchmark data collection, time periods, road segmentation, duration of evaluation, and evaluation transparency and objectivity.

Proper definition of the context and scope is pivotal in the evaluation plan as it has an effect on saving vital evaluation resources like time and money. It is important to note that there can be several preferred solutions for choosing evaluation parameters since they are reliant on the scope and context (Turner, et al., 2011).

Table 5, amended from Turner *et al.* (2011), presents a logical development of the evaluation plan. Specific details are provided adjacent to the corresponding process to assist in making decisions about the evaluation parameters.

Table 5: Overview of data accuracy assessment process (Turner, et al., 2011)

Process	**Consideration details**
Determination of evaluation scenarios and available resources	• The type of traffic information for evaluation • The type of available resources such as equipment and budget
Identification of applicable roadway segmentation	• Considers how traffic information is currently provided
Time interval selection	• Choose a time interval for evaluating traffic data • Consider longer time intervals for inadequate sample sizes
Route selection	• Choose road network according to facility type such as arterials • Choose routes according to level of expected variance
Time periods for evaluation	• Consider peak periods for collecting traffic data • Consider non-recurring events in evaluating data
Benchmark data collection method	• For statistically valid evaluations, consider re-identification method • Consider using test vehicle for initial screening or other limited situations
Accuracy measures selection	• Selection of accuracy measures is based on the evaluation scenario • Consider multiple accuracy measures for a detailed and comparative evaluation
Determining frequency and duration of evaluations	• Re-identification method increases duration of evaluation. This gives a better possibility of evaluating traveller information service (TIS) accuracy in non-recurrent congestion
Transparency and objectivity of evaluation	• A purchaser of TIS data hires a third party for evaluation • Publicly avail the evaluation method and results

2.6.1.1 Evaluation scenario

The following named scenarios are the most common when considering evaluation of travel time data and link speed accuracy based on traveller information service (TIS); congestion category, link speed (Chandra, et al., 2014) and route travel time. Table 6 depicts travel information provision for each evaluation scenario.

Table 6: Evaluation scenarios with corresponding travel information required

Evaluation Scenario	**Travel information**
Congestion category	Defined range of travel speed values
Link speed	Numeric speed values for specified roadway links
Route travel time	Travel times for specified travel routes provided as integer values

In a case where more than one scenario is included in an evaluation, adjustment is required on some of the evaluation parameters. This accounts for the multiple scenarios.

2.6.1.2 Roadway segmentation

The roadway segmentation depends on the evaluation scenario and it is vital to provide accurate information on the link and route endpoints. The segmentation utilised to report travel times or speeds should be utilised in collecting reference data and accuracy measures calculation. Appropriate segment lengths are crucial in traffic studies. Ahsani, et al. (2019) recommended segment lengths ranging from about 300 m to 2400 m dependent on the FCD data application.

2.6.1.3 Time Intervals

The evaluation scenario defines the time interval required to collect and compare reference data. Depending on the evaluators, the time intervals can be either be rolling or fixed but the reference values should be assessed to match the exact same interval as the traveller information service data.

Moreover, sample size consideration is paramount when selecting a time interval. Collected samples are used in calculation of reference travel times or speeds. A time interval should be long enough to accommodate an adequate number of samples. Time intervals for reporting real-time data are 1, 2 and 5 minutes (Chandra, et al., 2014) while 15, 30- and 60-minutes intervals are normally used for historical traffic data and the average daily trend (Adu-Gyamfi, et al., 2015)

Selection of duration for FCD data collection is also dependent on the length of segments. Altintasi, et al. (2017) in their study on detection of urban traffic patterns from floating car data, used FCD which had been collected for a 1-minute intervals on small road segments of about 50 m segments from a Belgium-based traffic information provider.

2.6.1.4 Evaluation links or routes

Routes or links to evaluate should be classified by the level of variance (high or low) in travel times or speeds. Links and routes with high variance are considered challenging to monitor effectively and estimate conditions accurately. Accurate traffic information is provided through traffic parameters such as average speed and link occupancy (Altintasi, et al., 2019).

2.6.1.5 Network stratification

Stratification is employed according to facility type. The two main types of facility are freeways and arterials. Arterials are further divided into urban and non-urban arterials. Rural land-uses or suburban areas with low signal and access density distinguish non-urban arterial. High access point density, high signal density and greater land use intensity characterize urban arterials.

2.6.1.6 Time periods to evaluate

General times to consider include morning peak, midday, evening peak and overnight. The specific time values and duration are usually user defined, as they are dependent on travel conditions on each specific link. Off peak evaluations serve to determine whether traveller information service provide valid estimates in low volume scenarios. Therefore, in case of limited resources for data collection, focus should be put on peak periods (Altintasi, et al., 2019).

The most challenging times for accurate information are characterised by non-recurring events such as special events or incidents on the link which tend to disrupt normal traffic. Despite it being difficult to predict such times, efforts should be employed to select times that include non-recurring events unlike in historical data when the non-recurrent events have already happened.

2.6.1.7 Benchmark data collection method

The most preferred methods of reference data collection are re-identification and test vehicle methods. The process of considering which method to utilise depends on several aspects such as required sample size, available funds, staff knowledge and expertise, equipment and many evaluation parameters. Traffic data collected from stationary devices can record the frequency and the number of vehicles traversing which is advantageous but such methods have high installation and maintenance costs (Kessler, et al., 2018). Due to the various considerations required, the process of benchmark data collection is discussed in section 2.6.2.

2.6.1.8 Accuracy measures

Accuracy measures the proportion to which estimated values and benchmark values are relative to each other. According to Kessler, et.al (2018), the level of similarity should be adequate enough to discuss the quality of traffic information between the referred methods of data collection. The selection of accuracy measures is based on the evaluation scenario. Evaluation scenarios include link speeds, route travel time and congestion category. Table 7 shows some of the commonly used accuracy measures for each assessment scenario. Link speed accuracy measures are discussed in more detail in section 2.6.3.1.

Table 7: Data evaluation accuracy measures (Turner, et al., 2011)

Evaluation scenario	Accuracy measures
Route travel time	• Average absolute error percent error (%) • Average absolute error per unit length (seconds per kilometre) • Average absolute error (minutes)
Congestion category	• Percent of correct category classifications within a specified boundary tolerance.
Link speed	• Average absolute error (km/hr) • Signed error bias (km/hr) • Root mean square error (RMSE) (km/hr) • Percent of TIS values within a specified error tolerance of the benchmark value.

2.6.1.9 Frequency and duration of evaluations

A data quality evaluation is indicative of evaluated data at a certain time. However, the results may differ if the assessment is repeated. According to Turner et.al (2011), consideration should be made on the repetition and time period of an evaluation.

Data quality evaluations are conducted in short, medium or long-term assessments. A short-term assessment (a day or a peak period) can have inadequate samples hence providing unreliable data. A longer duration such as a medium-term (a week or a month) provide a larger sample hence more accurate. These durations are functions of average sample size and cost. The longer the evaluation period, the higher the cost incurred and the more the sample size collected. Re-evaluation of a data quality assessment is a complex exercise and mostly considered due to changes in network, changes in local travelling behaviour or due to business requirements.

It is recommended that for specific studies that focus on evaluation of the accuracy of traffic information, short term trends to be used. Adu-Gyamfi, et. al (2015) suggested that long term trends (bi-diurnal, bi-weekly, monthly among other long-term variations) are relevant for general performance assessment while medium-term (1 - 3 hours) and short-term trends (15 – 30 minutes) are used for evaluating accuracy with probe-based data. In their study, they looked at accuracy of congestion detection with probe-sourced traffic data.

2.6.1.10 Evaluation objectivity and transparency

It is crucial to maintain a transparent and objective assessment process, more so for performance-based data service agreements. Even the perception of misconduct could question the assessment of the results. Therefore, the most effective approach promoting objectivity and transparency has been the purchaser hiring and using third party evaluators who publicly document the entire evaluation method and results. The most commonly used third party evaluators are university research groups and private consultants.

2.6.2 Collecting and reducing reference data

Reference data is defined as a set of data values considered a true representation, against which estimated parameters are compared (Turner, et al., 2011). Reference data is also referred to as benchmark data. The phrase ground-truth is sometimes used but not quite preferred. The true values such as the true average travel speed or time are rarely measured but are provided as estimates with some level of confidence.

Considering benchmark data as the standard values for reference, caution must be adhered to during benchmark data collection, reduction and reporting. The most preferred methods of reference data collection are test vehicle and re-identification methods as traffic conditions are effectively managed with accurate travel time information (Haghani, et al., 2010).

2.6.2.1 Test Vehicle method

For decades, the test vehicle technique has been used. A vehicle is deployed for data collection after being instrumented with traffic data equipment such as GPS and driving instructions provided to the driver. Cumulative travel time is recorded at predefined points along the travel route by an observer and the information is converted to travel time, speed and delay for every section along the route. These vehicles are referenced as "active" test vehicles (Turner, et al., 1998).

Levels of instrumentation used to measure travel time with such vehicles are:

- Manual - elapsed time is recorded manually at predefined data collection points using an observer in the vehicle
- Distance Measuring Instrument (DMI) – an electronic DMI connected to the transmission of the test vehicle provides speed and distance information.
- Global Positioning System (GPS) – a system of satellites orbiting the earth relay signals that determine the test vehicle's position and speed.

The driver is considered as part of the traffic data collecting team and therefore, the behaviour and driving styles are guided to match the desired driving behaviour (MnDOT, 2007). The common test vehicle driving styles are as follows:

- Average car – the test vehicle travels according to how the driver judges the average speed of other vehicles in the traffic stream.
- Floating car – the driver strives to estimate the median speed in the traffic by attempting to safely pass and be passed by an equal number of vehicles.
- Maximum car – the driver navigates according to the posted speed limit unless hindered by actual traffic conditions or safety considerations.

The floating car approach is the most common but it is important to note that most studies integrate the floating car and average car styles especially in high traffic condition due to difficulties in keeping track of high passing and passed vehicles (Turner, et al., 1998). This method also results in a very small sample size (typically only one or two trajectories are recorded by one or a limited number of drivers) and is prone to impact the driver behaviour of the test vehicle driver.

2.6.2.2 Re-identification method

Re-identification methods comprise of identifying a unique vehicle attribute at the initial point of a roadway segment under investigation and re-identifying the same vehicle at the end of the segment (Gwara & Andersen, 2018). Bluetooth enabled devices, license plates and electronic toll tag identifiers are some of the common re-identification methods. Bluetooth data contains a media access control (MAC) address that is unique to each vehicle. Electronic toll tags and license plates are common attributes in the Open Road Tolling (ORT) system. Automatic Number Plate Recognition (ANPR) is used for identification in the ORT system. The licence plate number is unique to each vehicle.

Re-identification method of traffic data collection relies on "passive" probe vehicles since the probes are in the traffic stream primarily for a different purpose other than evaluation. Therefore, speed data can be evaluated from the travel time information collected between two points (initial and end) of the road segment. It is recommended that freeways, urban arterials and non-urban arterials with high variance levels be sampled via re-identification techniques of data collection. Freeways and non-urban arterials, which are inclusive of suburban and rural links, with low variance levels, may be sampled using a test vehicle or re-identification methods (Turner, et al., 2011).

Table 8 compares the two common methods of reference data collection based on characteristics such as cost, extent of coverage and overall experimental control.

Table 8: Characteristics of re-identification and test vehicle techniques (Turner, et al., 2011)

Characteristics	**Re-identification**	**Test Vehicle**
Coverage depth	On high-volume roads and times, it provides more samples per time interval	Very limited samples are provided compared to re-identification
Breadth of coverage	Coverage is limited to the positioning of the equipment	Coverage is greater than re-identification. However, this is at the cost of coverage depth
Technology adoption	Dependent on technological technique used for re-identification. Bluetooth is relatively new for applications in traffic monitoring but a mature communication standard	DMI and GPS technologies are mostly used in transportation applications.
General experimental control	Should filter vehicles with multiple passengers such as transit vehicles and slower modes such as bikes and buses Should filter outliers to remove vehicles that veer of the designated route Should minimize error by ensuring adequate sensor spacing	Enforce its standard application by establishing a driving protocol for test vehicles
Range of costs	Cost depends on the re-identification method. Bluetooth devices can provide a higher number of data points per data collection dollar than test vehicle.	Cost is normally the curbing factor, and mostly limits test vehicle runs to a statistically insufficient sample size.

Research conducted by Turner *et.al* (2011) noted that Bluetooth re-identification has been considered the accepted practice for statistically valid speed accuracy evaluations. However, this study will explore the use of traffic monitoring devices that collect spot speeds as the primary method of re-identification. These devices are inclusive of inductance loop detectors, piezoelectric sensors and magnetic sensors.

2.6.2.3 Inductance loop detector method

Inductance loop detector is the most widely administered point detection device. Placing two detectors in series provides a more accurate vehicle speed measurement. The arrangement is known as a loop or speed trap. However, a study conducted by the Texas Transportation Institute, found out that the accuracy of the inductance loop trap is dependent on factors such as inductance loop wire type, consistency of design and trap length (Woods, et al., 1994).

In order to capture unique vehicle features, the vehicle signature matching method is used. A probe vehicle is identified and matched between two consecutive locations providing link-based travel time and speed. Stationary counted traffic data are advantageous in that they register the number and frequency of every passing vehicle but with a low spatial availability. Kessler *et.al* (2018) did a study comparing speed data from stationary detectors against FCD and suggested that stationary detector data can be considered as benchmark with an average distance 1.3 km between detector cross-sections.

2.6.2.4 Probe vehicle sample size

Probe-based sources of traffic data are becoming more cost-effective alternatives in gathering relatively accurate traffic information. However, the accuracy of the probe-based data remains a concern (Adu-Gyamfi, et al., 2017). The number of probe vehicles is a critical attribute for evaluating accuracy of traffic data (Chandra, et al., 2014).

The minimum sample size is determined by the t-statistic, the coefficient of variation, standard deviation, relative allowable error and sample mean parameters (Schneider IV et al., (2010), Gwara and Andersen, (2018)).

2.6.3 Computing and reporting accuracy measures

Accuracy is a qualitative evaluation of freedom from error. It is the degree of agreement between an array of values and a source of data assumed to be correct, referred to as benchmark data (Turner, et al., 2011).

Computing and reporting accuracy measures requires reference values and traveller information service (TIS) values to be collected under the same conditions. The date, time and location should be matched as close as possible to minimize the differences in the observed and benchmark data. Any alteration done in post processing on TIS estimates should consequently be applied on the benchmark data (Turner, et al., 2011).

Research conducted by the Texas Transportation Institute in collaboration with the University of Virginia proposed the use of confidence intervals to indicate uncertainties in benchmark values and provide a fair assessment of observed data (TTI, 2012)). The authors also advised that assessment results should be calculated and reported separately in different traffic conditions ranging from light flow to heavy congestion (Table 9) to characterise observed data accuracy adequately (TTI, 2012).

Table 9: Recommended speed ranges for accuracy statistics computation and reporting (Turner, et al., 2011)**.**

Range	**Benchmark speed range**	
	Access controlled roads (e.g. freeways)	**Arterial Streets**
1. Light flow	Greater than 100km/h (60mph)	Greater than 65 km/h (40 mph)
2. Transition	70 to 100 km/h (45 to 60 mph)	50 to 65 km/h (30 to 40 mph)
3. Slowing	50 to 70 km/h (30 to 45 mph)	30 to 50 km/h (20 to 30 mph)
4. Stop and Go	Less than 50 km/h (30 mph)	Less than 30km/h (20 mph)

NB: A conversion of 1 mph = 1.609 km/h was used and rounded off to the nearest 5 or 10 km/h

As indicated in Section 2.6.1.8, accuracy measures are dependent on the evaluation scenario. It is important to note that there is no single best accuracy measure for all possible evaluation scenarios. The user or the purchaser defines acceptable accuracy values based on their intended purpose of the data.

This research focuses on speed data on particular segments. Therefore, the scholar shall present the accuracy measures linked to the link speed evaluation scenario as discussed hereafter.

2.6.3.1 Link speed accuracy measures

The accuracy measures employed in the link speed evaluation scenario are:

a) **Root mean square error (RMSE)**

This measure gives greater weight to higher magnitude errors (by squaring the error term). The RMSE therefore measures the difference between raw data and benchmark data (Siripollawat, 2012).

The error is the difference between the observed value and the benchmark value. To compute the error as a percent, the error is divided by the benchmark value (this is also applicable to the other error formulations presented for this evaluation scenario). RMSE units are given as distance per unit time e.g. mph (miles per hour) while %RMSE is the RMSE divided by the mean of all benchmark data values (**Equation 2**). **Equation 1** indicates the RMSE formulae. It is commonly available in various statistical software applications.

Equation 1

$$\boldsymbol{Root\ mean\ square\ error, RMSE\ (km/hr)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - x_{ref}\right)^2}$$

Equation 2

$$\boldsymbol{Root\ mean\ square\ percent\ error} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - x_{ref}}{x_{ref}}\right)^2}$$

where x_i = the observed value

x_{ref} = the benchmark value

n = the total number of observed data values

b) **Signed error bias**

The average error is also referenced as the bias or the signed error bias (SEB). The resulting positive or negative sign after calculation indicates whether the observed values are consistently higher or lower than the reference values. The signed error can be solved to a percentage value through the explanation presented in a) above

It is beneficial to note that it should only be used as a measure to reveal if there is a consistent bias in the observed or TIS values because large positive and negative errors can nullify each other and lead to a deceptive estimate of the magnitude of the error. **Equation 3** shows the calculation method.

Equation 3 $$\boldsymbol{Signed\ error\ bias\ (km/hr) = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{ref})}$$

Equation 4 $$\%\ \boldsymbol{Signed\ error = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - x_{ref}}{x_{ref}}\right)}$$

where x_i = the observed value

x_{ref} = the benchmark value

n = the total number of observed data values

c) Average Absolute Signed Error (AASE)

The average absolute error (**Equation 5**) assesses the average magnitude error notwithstanding the sign. Expressing this error as a percentage results in the average absolute percent error (AAPE). AAPE (**Equation 6**) and percentage Signed error may be used to compare the relative accuracy of attributes such as traffic volume count and speed measurement accuracy. However, it is not advisable to incorporate them for route travel time scenario as they project misleading results when comparing different route lengths (Turner, et al., 2011).

Equation 5 $$\boldsymbol{Average\ absolute\ signed\ error\ (km/hr) = \frac{1}{n}\sum_{i=1}^{n} abs(x_i - x_{ref})}$$

Equation 6 $$\boldsymbol{AAPE\ (km/hr) = \frac{1}{n}\sum_{i=1}^{n} abs\left(\frac{x_i - x_{ref}}{x_{ref}}\right)}$$

where x_i = the observed value

x_{ref} = the benchmark value

n = the total number of observed data values

d) Percent of TIS values within X of the benchmark value

This accuracy measure is a straightforward measure and provides the ability to report a consistent measure by varying the error tolerance on various facilities such as arterials and freeways or locations such as rural or urban.

X is a specified error tolerance in mph, percentage or statistical confidence interval.

Equation 7 $$\textit{Percent of TIS values within } X = \frac{\textit{Number of TIS values within } X}{\textit{Total number of comparisons}}$$

The signed bias error, average absolute error and root mean square error can be quantified for both single link-date-time combination and several links taking into consideration the speed ranges tabulated in Table 9. The percent of TIS values within X is unfitting when few reference comparisons are present. However, when computed in the presence of enough comparisons, the speed ranges are applicable.

The speed ranges in Table 9 are recommended for evaluating accuracy measures for route travel times and link speeds in different traffic conditions (Turner, et al., 2011). They should be used cautiously as comparison of evaluation results can differ when the different speed ranges are adjusted.

2.7 SELECT CASE STUDIES RELEVANT TO THE RESEARCH

This section presents some of research studies conducted to assess the accuracy and reliability of FCD. Floating car data used in such studies is gathered from on-board and potable GPS-enabled devices and smartphones. Data service providers such as TomTom, INRIX, HERE and Google play a role in providing FCD commercially. The projects focused herein were considered relevant to the overall aim of this research.

2.7.1 Evaluation of alternative technologies to estimate travel time on rural interstates, USA

2.7.1.1 Project summary

The purpose of the study was to present the findings of an assessment of alternative technologies to estimate travel time along a segment of Interstate 91 in Western Massachusetts, USA. The research area was reported to have low traffic volumes hence resulting in relatively low sample sizes. The authors also aimed to contribute towards the achievement of the Regional Traveller Information Centre (RTIC) and MassDOT I-95 ITS project (Jia, et al., 2013).

2.7.1.2 Data source

The study utilised three sources of data. Data collected from INRIX data service providers and data collected at the I-91 study site via Bluetooth technology (Bluetooth readers) was compared and evaluated against a third set of data collected through a video-based license plate matching method. The license plate-based data was incorporated and considered the benchmark data.

2.7.1.3 Methodology

Outliers and significant errors developing from the data collection process were first identified and deleted. This was done by applying normal distribution to get mean and standard deviation. Any data that fell out of the defined interval was considered erroneous. Mean squared error (MSE), Root mean square error (RMSE) and Mean absolute error (MAE) were considered as statistical measures to quantify the difference between the benchmark data from license plate-based method and the GPS data collected by INRIX and Bluetooth data. However, to interpret the data more reliably, the researcher applied the mean absolute percentage error (MAPE) to compare errors between data sources. The absolute error compared to the relative error was considered to be more informative.

The data collection was done along three segments on the Interstate 91 in Western Massachusetts with presence of low traffic sample sizes at a 5-minute time interval. The samples were divided into batches per segment. Segment 1 had 35 batches of data while segments 2 and 3 had 34 and 32 batches respectively.

2.7.1.4 Findings

The accuracy of travel times estimated through the MAPE method of statistical analysis comparing the benchmark data with GPS-based data and Bluetooth-based data (BT) provided acceptable estimates of accuracy. In Segment 1, the MAPE between BT and benchmark was 5.48% and for GPS it was 3.97%. For segments 2 and 3, the MAPE for BT and GPS were 4.24%, 3.33% and 5.91%, 4.88% respectively.

A further t-test analysis proved that data from GPS devices were statistically not different from benchmark data while Bluetooth data were statistically different from the benchmark data collected. However, the difference in travel times from Bluetooth devices and on-board GPS devices compared to travel times from license plate recognition technology were less than 1.2%.

The data analysis showed that accuracy of travel times estimated from Bluetooth-based and GPS-based are acceptable since their MAPEs were less than 6% along the segments investigated. In addition to that, results from a comparison done on the three sources of data to a floating car suggested that technological error is negligible when other sources of error e.g. aggregation error, sampling error and measurement error are controlled and eliminated.

2.7.2 Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel.

2.7.2.1 Project summary

The research paper presents an operational system of measuring traffic speeds and travel times based on information gathered from cellular phones and comparing the information with data obtained by dual magnetic loop detectors. The comparison utilised data from a 14 km freeway (Ayalon freeway) with 10 interchanges in both directions, collected between January and March of 2005. The dataset contained 1284587 valid loop detector speed measurements and 440331 valid measurements from the cellular system with each measurement covering a 5-min interval (Bar-Gera, 2007).

2.7.2.2 Data source

Three data sets were applied in this study. Data from dual magnetic loop detectors and data obtained from cellular phone-based systems. The loop detectors collected traffic data for all lanes after approximately every 500 m. The cell phone-based system received observations for about 1-3% of the total traffic during daytime. Travel times were converted to mean section speed by getting the ratio of the road section length to estimated travel time. 25 floating car measurements were conducted as additional observation data for a week in the stipulated period.

2.7.2.3 Methodology

Presence of large datasets prompted the researcher to consider graphical representation as a more useful tool for data comparison. The analysis was done in two sections: comparison of speeds by location and time and comparison of travel times along the entire roadway. The researcher conducted comparison statistics for travel time computations between loop detector data and cellular phone data to get the average absolute relative difference. It is worthwhile to note that none of the datasets was used as benchmark data.

2.7.2.4 Findings

The writer concluded that there was a good correspondence between the loop detector system and the cellular phone system while the floating car travel time measurements provided additional assurance for data accuracy. The researcher predicted increased usage of cellular phone-based data for various practical applications such as advanced traveller information system and evaluating system performance for modelling and planning.

2.7.3 Validation of TomTom historical average speeds on freeway segments in Gauteng, South Africa

2.7.3.1 Project summary

The project aimed at evaluating the quality of TomTom data. Gwara and Andersen (2018) compared benchmark speeds on six directional segments on the N1 and R21 freeways to TomTom historical speed estimates in Gauteng province. The N1 and R21 freeways are classified as urban roads. The Automatic Number Plate Recognition (ANPR) data, a component of the Open Road Tolling was utilised as a source of the reference data. This was part of the investigation due to the advancement in technology to provide innovative and alternative sources of traffic data.

2.7.3.2 Data source

The data sources included TomTom historical data and data from Open Road Tolling (ORT) system present on some sections of Gauteng province freeway network in South Africa. The ORT uses ANPR to capture traffic data when the vehicles traverse through a gantry. The TomTom historical data was downloaded from the Traffic Stats on the TomTom website while the ORT was requested from the South African Road Agency Limited (SANRAL).

2.7.3.3 Methodology

The study focused on six freeway segments along the N1 Ben Schoeman, N1 Western bypass and R21 Albertina Sisulu in Gauteng province. The segments were defined as the distance between two consecutive gantries on the freeway and chosen as per the direction of traffic flow, that is, northbound or southbound. Trips that were made from the first gantry and exited before the second gantry or exited the freeway and re-entered before the second gantry, were not considered due to longer travel times.

TomTom historical data from the TrafficStats database was already processed hence no further manipulation was required. Quality assurance checks were conducted by filtering out undesirable travel speeds and times on the ORT raw data requested from SANRAL and necessary data processing performed to obtain significant data for comparison. The ORT was employed as the benchmark data for evaluating the quality of TomTom historical data. The trips were allocated according to 15-minutes and 1-hour intervals from which data derived from the two sets of intervals was compared. The research adapted the procedure provided by Turner *et al.* (2011) and Battelle (2004) where accuracy measures were conducted by incorporating the following error measurements: The Average Absolute Speed Error (AASE), the Signed Error Bias (SEB) and the Signed Error.

2.7.3.4 Findings

The error measurements were considered within a certain range to provide a good accuracy estimate of the TomTom historical data. The AASE was to be evaluated to 10 km/h , the SEB within ±7.5 km/h and the allowable signed error was considered to be accurate if it was within ±10%. From the 15-minutes interval analysis, the AASE, SEB and Signed Error were evaluated to 6.4 km/h, -6.3 km/h and -5.8% respectively while the 1-hour interval analysis the error quantities were evaluated to 6.5 km/h, -6.3 km/h and -6.2% respectively.

In terms of time interval selection for evaluating accuracy, a two sample assuming equal variances t-test was conducted for the error measurements derived from the 15-minutes and 1-hour time intervals. The results concluded that there were no significant differences between the error measurements. The researcher concluded that the quality of TomTom historical data was in line with the reference speeds used for comparison despite some data consumers valuing certain data quality measures over others. The researcher recommended application of composite data quality measures bearing in mind its meaning and implications. It was also noted that TomTom generally underestimates traffic speeds.

Moreover, the writer highlighted the potential usage of probe data as a key source of traffic data in prevailing applications of transportation engineering and operations.

2.8 CONCLUSION

In this chapter, the sequential process of evaluating the accuracy of traffic data based on speed data was detailed. To portray a clear understanding of the type of facility to be considered in the research, a discussion was presented in section 2.2 focusing on the road classification system in South Africa. It is worthwhile to note that the guidelines presented shall be adapted for assessing historical probe data.

CHAPTER 3 : METHODOLOGY

3.1 INTRODUCTION

This chapter looks to provide an overview of the research methodology that will be utilised to assess the accuracy of commercial FCD along rural routes in the Western Cape of South Africa. Research design, description of study area and data sources are discussed.

3.2 RESEARCH DESIGN

The only published commercial FCD accuracy study in South Africa to date was conducted by Gwara and Andersen (2018) and considered the accuracy of FCD on urban freeways in Gauteng. The design approach considered in this thesis was based on the quantitative research adopted from the study of Gwara and Andersen (2017). In their study, FCD and benchmark speed data was obtained from six directional freeway segments aggregated over 15-min and 1-hour time intervals. The study considered only weekdays (Monday to Friday) over the month of February 2015. Gwara and Andersen (2018) evaluated the average speed error per road segment that could be observed between FCD-reported speeds and average speeds reported between the gantries of the Open Road Tolling system implemented in Gauteng. They evaluated Signed error bias (SEB), Average absolute signed error (AASE) and Signed Error (SE) (%) per road segment. Additionally, they averaged the speed bias results derived from each individual road segment to describe speed bias in the overall study area.

Studies conducted by Haghani et al. (2013), de Fabritiis et al. (2008) and Adu-Gyamfi et al. (2017) used short term intervals of 5 minutes, 15 minutes and 30 minutes for evaluating the reliability and accurately quantifying the similarities and dissimilarities of probe-sourced traffic speed data and benchmarked local sensor data. Gwara and Andersen (2017) compared the aggregation level between 15 minutes and 1-hour intervals and concluded that there was no significant difference when evaluating traffic speed accuracy between benchmark data and historical FCD for the two different aggregation periods. Accordingly, 1- hour aggregation was used for this study.

Static sensor data was used as a source of benchmark data in this research. Several studies have been conducted globally to compare the reliability and accuracy of FCD against static traffic sensor data such as inductive loop detector data and radar data (Ahsani, et al., 2019) (Adu-Gyamfi, et al., 2017). In this thesis, data was aggregated per source in hourly periods for all weekdays (Monday to Friday) for the month of February 2019 to calculate the SEB and AASE and SE (%) per directional segment. The results were then averaged per location to calculate the speed error for all rural and urban locations to evaluate the difference between rural and urban speed error in FCD.

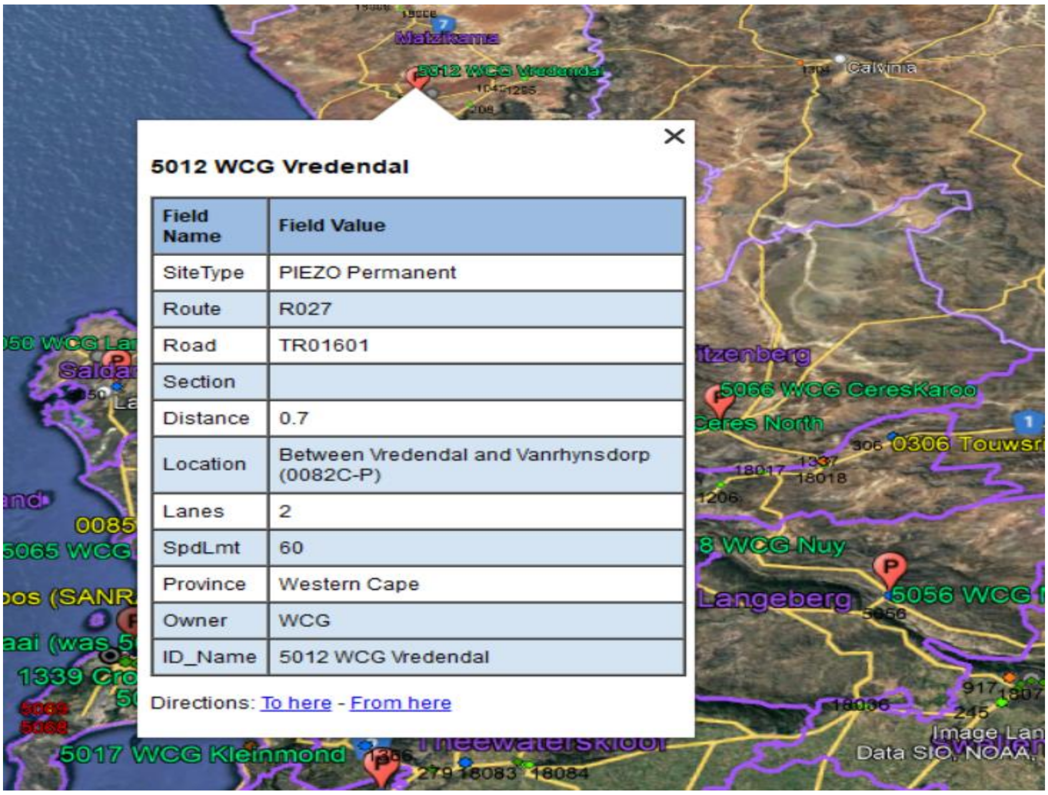
As part of their analysis, Gwara and Andersen (2018) used two traffic datasets and employed a statistical hypothesis test to determine if the TomTom and Open-Road Tolling (ORT) average speeds were significantly different at a 95% confidence interval. Wang (2012) conducted a statistical paired t-test in his estimated freeway travel time delay study to estimate the variation in FCD and loop detector mean.

In this study, a two-sample t-test assuming equal variances was conducted to determine the difference in sample mean for SEB and AASE in rural and urban areas. The stations are captured in Table 10 and Table 11 respectively.

3.3 DESCRIPTION OF THE STUDY AREA

3.3.1. General information

The Western Cape Department of Transport has a network of 7-day and permanent traffic observation stations located around the Western Cape. Data is collected by static traffic sensors, particularly inductive loops. Because monthly traffic data was used in this study, data from the permanent traffic count stations were preferable. An example of location information for a permanently placed loop detector is captured in Figure 5 for site 5012, close to Vredendal along the R27.



5012 WCG Vredendal

Field Name	Field Value
SiteType	PIEZO Permanent
Route	R027
Road	TR01601
Section	
Distance	0.7
Location	Between Vredendal and Vanrhynsdorp (0082C-P)
Lanes	2
SpdLmt	60
Province	Western Cape
Owner	WCG
ID_Name	5012 WCG Vredendal

Directions: To here - From here

Figure 5: **Example of Western Cape Government traffic count station information (Google Earth)**

3.3.2. Rural station selection

Rural areas are characterised by sporadic developments which require roads with longer reaches of connectivity. Therefore, rural roads require higher levels of mobility roads than urban roads. However, both urban and rural roads have the same classes but at different scales and standards (COTO, 2012). A detailed discussion on road classification is covered in Section 2.2.

Considering the Road Access Guidelines for South Africa (EDAT, 2002), the classification is divided into higher order arterials and lower order roads. The higher order arterials are inclusive of freeways, expressways and primary arterials while the Lower Order roads encompass the district distributors, local distributors and access roads. Understanding of the classification system were used to differentiate stations from either a rural or urban setting. In both cases, arterial roads were selected. Traffic data is not calculated on lower-order roads.

Comparison of data collected from static road side sensors. FCD was made on selected high order roads in rural and urban regions in the Western Cape. A total of 15 rural and 13 urban roads were selected for analysis. Any station with missing data or traffic speeds recorded as zero at any hour of the day were filtered out from analysis. This resulted in a final study area comparing 12 rural traffic count stations and 7 urban counting stations which had complete information and were considered for further analysis and compared with FCD. Figure 6 shows the study area and relative locations of the selected rural counting stations throughout the Western Cape.

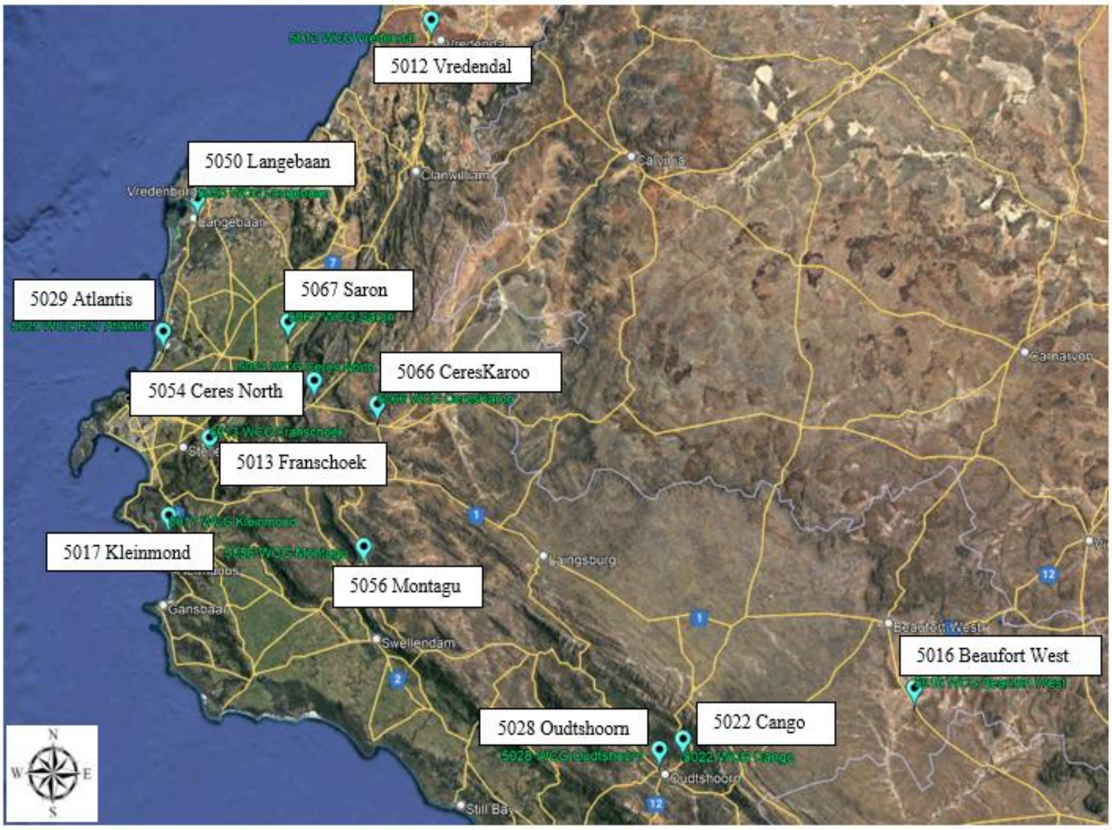


Figure 6: Location of rural permanent counting stations (Google Earth)

Details of the rural counting stations are tabulated in Table 10, including information on the name of the counting station, the route on which it is situated, the speed limit and information on the location of the counting station.

Table 10: Descriptions for selected rural counting stations

Rural (Data Station)	Route	Speed limit (km/h)	Location	Municipality
5012 WCG Vredendal	R27	60	Between Vredendal and Vanrhynsdorp	Matzikama Local Municipality
5013 WCG Franschoek	R45	100	Between R301 T/O and Akkerdal Wine Estate (Franschoek)	Franschoek Municipality
5016 WCG Beaufort West	R61	120	Between Beaufort west and Aberdeen	Close to Beaufort West Municipality
5017 WCG Kleinmond	R43	120	Between Botrivier and Hermanus	Overstrand Local Municipality

5022 WCG Cango	R328	80	Between Oudtshoorn and Cango Caves	Oudtshoorn Local Municipality
5028 WCG Oudtshoorn	R62	80	Between R328 T/O and Oudtshoorn	Oudtshoorn Local Municipality
5029 WCG Atlantis	R27	120	Between R307 T/O and Brakkefontein Rd T/O	City of Cape Town Metropolitan Municipality
5050 WCG Langebaan	R27	120	South of Langebaan turnoff	Saldanha Bay Municipality
5054 WCG Ceres North	R303	100	Between Ceres and Prince Alfred Hamlet	Witzenberg Local Municipality
5056 WCG Montagu	Van Riebeeck Street	100	Off R62	Langeberg Municipality
5066 WCG Ceres Karoo	R355	80	Between R46 Touwsriver T/O and Calvinia	Witzenberg Local Municipality
5067 WCG Saron	R44	120	Between R44/R46 intersection and Saron	Drakenstein Municipality

3.3.3. Urban station selection

Figure 7 shows the study area and relative locations of the selected urban counting stations. It should be noted that station 5057 (Eikendal) is in fact located in a rural zone. However, the traffic that it carries is urban in nature, because this is the main route between Somerset West and Stellenbosch, which is heavily trafficked in the peak periods, similar to urban traffic patterns. This is due to the high number of work opportunities in Stellenbosch that attract traffic from surrounding towns.

Figure 7: Location of urban permanent counting stations (Google Earth)

Details of the urban counting stations are tabulated in Table 11. They include the name and location of the counting station, the route on which the station is located and the speed limit of the road.

Table 11: Descriptions for selected urban counting stations

Urban (Data Station)	Route	Speed limit (km/h)	Location	Municipality
1339 Croydon	R102	100	Between Firgrove and Croydon Olive Estate	City of Cape Town Municipality
5001 WCG Macassar	M9	60	Between Macassar Rd and 3rd St intersection and Macassar Rd and Old Main Rd intersection	City of Cape Town Municipality
5057 WCG Eikendal	R44	100	Between Somerset West and Stellenbosch	Cape Winelands District Municipality
5058 Koeberg NB	N1	100	Between Koeberg and Sable Rd I/C (NB only)	City of Cape Town Municipality

5059 Koeberg SB	N1	100	Between Koeberg and Sable Rd I/C (SB only)	City of Cape Town Municipality
5064 WCG Airport	Airport Approach Rd.	80	On the Airport approach Rd between N2 and Cape Town international airport	City of Cape Town Municipality
5075 WCG Polkadraai	M12	100	Between Zevenwacht Link Rd T/O and Skoonheid Wingerde	City of Cape Town Municipality

3.3.4. Discussion of sample sites

The final study area comprised of 12 rural traffic count stations and 7 urban counting stations which had complete information and were considered for further analysis and were compared with FCD derived from TomTom. The few numbers of sample sites were as a result of limited permanent counting stations with complete data in the Western cape. Some selected sites had incompletely captured data whereby data were missing for certain hours of the day or missing data in a couple of days of the month. The counting stations especially the rural stations are also located in areas where vehicle transportation is not the main means of transport. Examples of counting stations that were retracted from the analysis are 5074 WCG Spier (on R310 Barden Powell Drive) which was an urban counting station and 5021 WCG Zebra (on R62) which was a rural counting station.

A study conducted by Gwara and Andersen (2018) used six freeway segments in Gauteng province, South Africa, for collecting data in order to validate the accuracy of TomTom historical average speeds. The Texas Transport Institute (TTI) (2012) considered 11 freeway segments in the evaluation of TomTom historical speed data. Moreover, a study conducted by Bruwer, et al. (2022) on the impact of probe sample bias on the accuracy of commercial FCD in urban areas, considered six cities in South Africa of which the researcher looked at a number of directional segments per city. The segments ranged from 15 directional segments in Pretoria (Gauteng province) to 42 directional segments in Cape Town (Western cape province).

Therefore, with evidence provided from the studies referenced herein, the 12 rural traffic count stations and 7 urban traffic count stations were sufficient to provide conclusive results in this study.

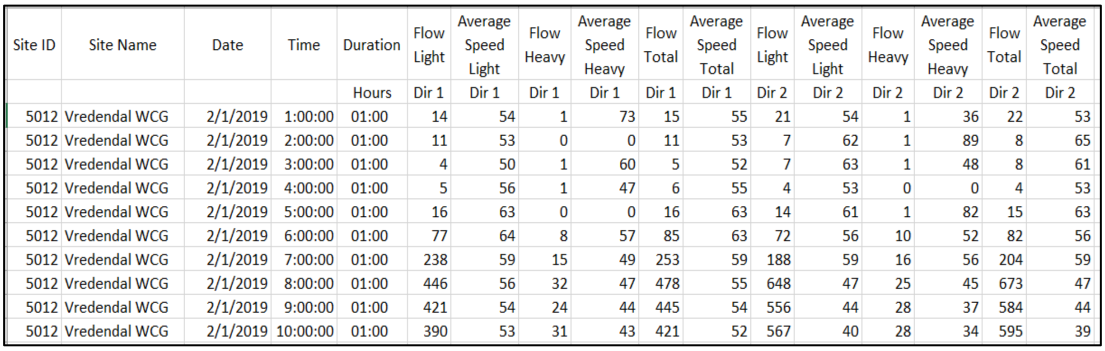
3.4 DATA SOURCES

3.4.1 Loop detector data

The Eulerian data collected via inductive loop detectors was provided by the Western Cape Department of Transport derived from permanent and 7-day traffic counting stations. Rural counting stations are tabulated in Table 10 while urban stations are present in Table 11. Permanent stations were used while 7-day counting stations were avoided to avoid bias. Permanent stations collect data 24-hours a day and throughout the year (365 days). Short duration counts are prone to non-recurring events which can disrupt normal traffic patterns and data recorded over the short investigation period might not be a true representation of the traffic pattern in an area (Turner, et al., 2011).

Data quality evaluations are conducted in short, medium or long-term assessments. A short-term assessment (a day or a peak period) can result in an inadequate sample hence providing unreliable data. A longer duration such as a medium-term (a week or a month) provide a larger sample and are hence more accurate. These durations are functions of cost and average sample size. The longer the evaluation period, the higher the cost incurred and the more the sample size collected (Turner, 2002).

Data was requested in hourly intervals for each day in the month of February 2019. The data package provided by the Western Cape Department of Transport contained Ms Excel workbooks, capturing each selected rural and urban counting station. A screenshot of data provided by the Western Cape Department of Transport is shown in Figure 8. Traffic data provided included the directional flow of the traffic according to light and heavy vehicles and the average speed per direction of light and heavy vehicles and all vehicles simultaneously. Data was provided per hour of each day of the month of February 2019.



Site ID	Site Name	Date	Time	Duration	Flow Light	Average Speed Light	Flow Heavy	Average Speed Heavy	Flow Total	Average Speed Total	Flow Light	Average Speed Light	Flow Heavy	Average Speed Heavy	Flow Total	Average Speed Total
				Hours	Dir 1	Dir 1	Dir 1	Dir 1	Dir 1	Dir 1	Dir 2	Dir 2	Dir 2	Dir 2	Dir 2	Dir 2
5012	Vredendal WCG	2/1/2019	1:00:00	01:00	14	54	1	73	15	55	21	54	1	36	22	53
5012	Vredendal WCG	2/1/2019	2:00:00	01:00	11	53	0	0	11	53	7	62	1	89	8	65
5012	Vredendal WCG	2/1/2019	3:00:00	01:00	4	50	1	60	5	52	7	63	1	48	8	61
5012	Vredendal WCG	2/1/2019	4:00:00	01:00	5	56	1	47	6	55	4	53	0	0	4	53
5012	Vredendal WCG	2/1/2019	5:00:00	01:00	16	63	0	0	16	63	14	61	1	82	15	63
5012	Vredendal WCG	2/1/2019	6:00:00	01:00	77	64	8	57	85	63	72	56	10	52	82	56
5012	Vredendal WCG	2/1/2019	7:00:00	01:00	238	59	15	49	253	59	188	59	16	56	204	59
5012	Vredendal WCG	2/1/2019	8:00:00	01:00	446	56	32	47	478	55	648	47	25	45	673	47
5012	Vredendal WCG	2/1/2019	9:00:00	01:00	421	54	24	44	445	54	556	44	28	37	584	44
5012	Vredendal WCG	2/1/2019	10:00:00	01:00	390	53	31	43	421	52	567	40	28	34	595	39

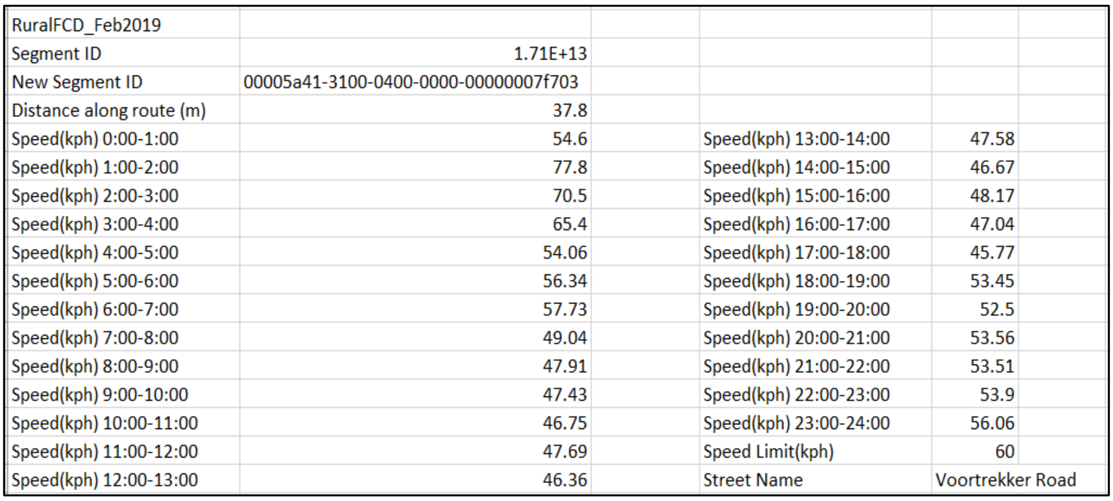
Figure 8: Benchmark data from Western Cape DoT

3.4.2 FCD

Historical FCD was obtained from the TomTom Move database (https://move.tomtom.com). Access to the database was granted through TomTom's partnership with the Stellenbosch University's Smart Mobility Lab. TomTom collects anonymised data from GPS measurements and builds a traffic database from the collected data. This data is widely used for transportation planning and operations (Dannehy & Krootjes, 2013) (Gwara & Andersen, 2018).

The Traffic Stats platform in the TomTomMove portal provides a platform to conduct area and route analysis and traffic density analysis of historical FCD. A typical area and route analysis define a specific route or area and generate average speed, average travel time and sample size data (TomTom, 2018). In this thesis, a route analysis was conducted on preselected rural and urban roads presented in Table 10 and Table 11. Results from the analysis were downloaded in Ms Excel format and each workbook contained the following data: route speed (harmonic average), route speed (arithmetic average), speed median, sample size and percentile speeds per directional segment.

A sample of the TomTom FCD information provided in the route speed (arithmetic average) worksheet is shown in Figure 9.



RuralFCD_Feb2019			
Segment ID	1.71E+13		
New Segment ID	00005a41-3100-0400-0000-00000007f703		
Distance along route (m)	37.8		
Speed(kph) 0:00-1:00	54.6	Speed(kph) 13:00-14:00	47.58
Speed(kph) 1:00-2:00	77.8	Speed(kph) 14:00-15:00	46.67
Speed(kph) 2:00-3:00	70.5	Speed(kph) 15:00-16:00	48.17
Speed(kph) 3:00-4:00	65.4	Speed(kph) 16:00-17:00	47.04
Speed(kph) 4:00-5:00	54.06	Speed(kph) 17:00-18:00	45.77
Speed(kph) 5:00-6:00	56.34	Speed(kph) 18:00-19:00	53.45
Speed(kph) 6:00-7:00	57.73	Speed(kph) 19:00-20:00	52.5
Speed(kph) 7:00-8:00	49.04	Speed(kph) 20:00-21:00	53.56
Speed(kph) 8:00-9:00	47.91	Speed(kph) 21:00-22:00	53.51
Speed(kph) 9:00-10:00	47.43	Speed(kph) 22:00-23:00	53.9
Speed(kph) 10:00-11:00	46.75	Speed(kph) 23:00-24:00	56.06
Speed(kph) 11:00-12:00	47.69	Speed Limit(kph)	60
Speed(kph) 12:00-13:00	46.36	Street Name	Voortrekker Road

Figure 9: FCD from Traffic Stats analysis

3.5 CONCLUSION

The research methodological chapter introduced the design approach that was followed for the analysis. It covered the study design, study coverage area, data sources and time interval to be utilised for speed accuracy error measurements and further statistical analysis.

CHAPTER 4 : DATA COLLECTION

4.1 INTRODUCTION

This chapter looks to provide an overview of the research data collection process for both the benchmark data and the FCD used to assess the accuracy of commercial FCD along rural routes in the Western Cape of South Africa. Collection of benchmark data and FCD is detailed, and the methods to analyse and statistically compare these sets of data was discussed in Literature review Chapter 2.

To adequately process traffic patterns from traffic data collected, spatial sampling and temporal sampling should be taken into consideration. Spatial sampling is the process of selecting a subset of spatial elements in a geographical area which can reflect the type of roads under investigation. Temporal sampling relates to selection of time elements of the study. This includes the month of sampling, day of the week, length of data collection, peak hours, off-peak hours and analysis period (Gwara & Andersen, 2018) (Schneider IV, et al., 2010).

Data for this study was collected for a period of three years from 2019 to 2021. However, due to the COVID 19 pandemic, an executive decision was made to sift out data collected for 2020 and 2021. This was due to probable effects on the data accuracy as a result of lockdown and restricted travels on traffic patterns. Analysis was therefore done for the year 2019. Data collection for this study was performed over the weekdays (Monday to Friday) in the month of February. February was chosen as the representative month as traffic volumes are higher in summer and most people are back to work and learners back to school. There are also no public holidays in February hence it would be a true representation of the traffic patterns.

Computing and reporting accuracy measures requires reference data and FCD data to be collected under the same conditions. The date, time and location should be matched as close as possible to minimize the differences in the observed and benchmark data. Any alteration done in post processing on TIS estimates should consequently be applied on the benchmark data (Turner, et al., 2011).

Statistical analysis was conducted to compare the data sets from FCD and reference speeds. Accuracy measures mentioned in section 2.6.3 were applied for the evaluation and are discussed in this chapter.

4.2 DATA

4.2.1 Benchmark data

Traffic data requested from the Western Cape Department of Transport was received in Microsoft Excel worksheets for preselected rural and urban locations. Spot speeds from the traffic observation stations were collected through single inductive loop technology. Single inductive loop technology can differentiate between light and heavy vehicles (TES, 2021). The data is logged in either in 15-minutes or 60-minutes intervals. The 60-minutes intervals were chosen for this study.

The excel workbook contained tabulated light and heavy vehicle speeds for each hour of the day for the month of February 2019, as well as the average speed of the road segment of all vehicles combined. To confirm the average speeds presented, an additional check was done by recalculating the average speeds for each counting station per directional segment. The recalculated average speeds were similar to the average speeds received from Western Cape government. Figure 8 shows the type of data that was received.

In order to have desirable data for analysis, data filtering was applied to address any undesirable observations. Data filtering was conducted in accordance with Haghani et.al (2010) process of separating unacceptable data points from observed data. The following characteristics for filtering out data were applicable to this study;

- Observations with zero speed and blank entries. This was regarded as missing data and therefore ignored;
- Observations that were significantly different from the average speed. This were outliers as they were abnormally off from the rest of the data.

4.2.2 FCD data

Floating car data was obtained using the TomTom Move portal by conducting a route analysis on the various road segments corresponding to the location of individual WCG count stations using the Traffic Stats module. Data from Traffic Stats evaluations contain information on average and median speed, speed limits, street names, speed percentiles and sample sizes of number of probes per directional segment (TomTom, 2018).

4.2.2.1 Traffic Stats data collection procedure

The following procedure describes how to collect data from the Traffic Stats module of the TomTomMove portal (Dannehy & Krootjes, 2013) (TomTom, 2018).

i) Log in

The sign in page is first accessed by opening the TomTom Move link https://move.tomtom.com/

Log in credentials that is, email address and password, were provided through the Smart Mobility Lab at Stellenbosch University. The license to access the database was granted through the university's partnership with TomTom.

ii) Choose a report type

To start on a report, a report name was created by including the researcher's surname, location for analysis, station and month of data collection. An example of a report name that was used to save data that was collected from a rural location was Kiautha_RuralFCD_5012.Vredendal_Feb2019. This format was convenient since several researchers had access to the portal and this ensured you download the correct data for analysis.

Route analysis report was therefore selected as the report type and the initial step of report creation was completed.

iii) Choose date(s)

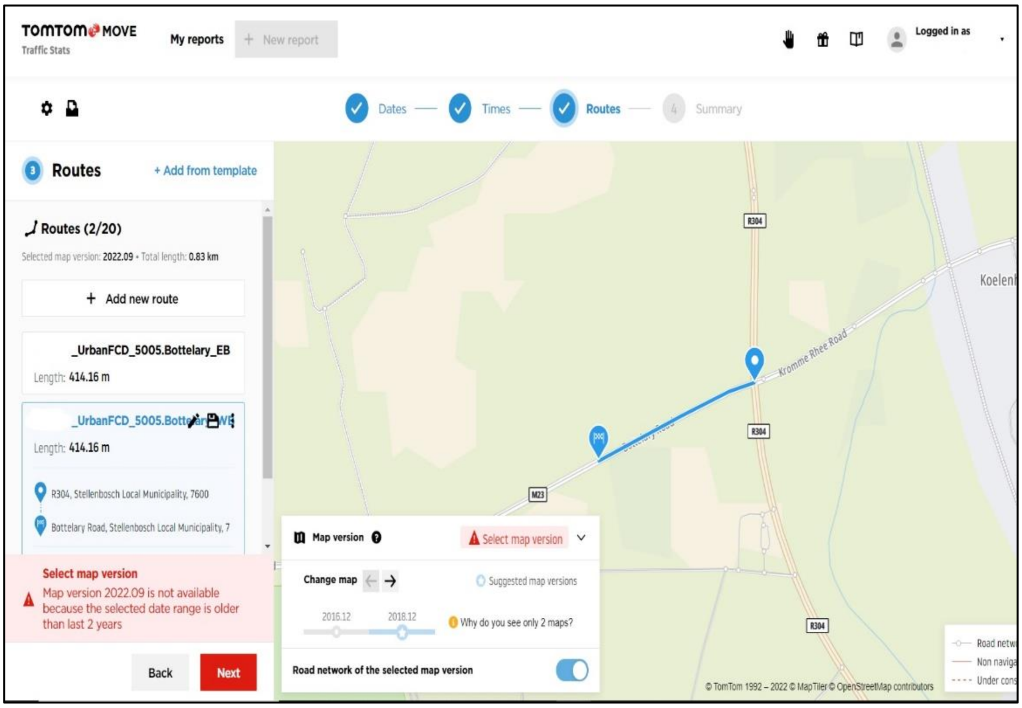
The date for this study was selected as February 2019. Since only weekdays (Monday to Friday) were to be analysed, weekends (Saturday and Sunday) were retracted from the days of the month and the new dates were saved as a template. Saving the date as a template was a time saving technique as it would be applicable across all the data that was to be collected from preselected rural and urban roads.

iv) Choose time period(s)

Observations were to be done on hourly intervals. Therefore, 1-hour periods were added 24 times to aggregate the data from February to each hour of the day. The time period was also saved as a template since it would be applicable across all the preselected rural and urban roads under investigation.

v) Select route or area

TomTom data is aggregated in links. There was need to locate the shortest possible link that correlated with the location of the comprehensive traffic observation station from google earth where reference data was collected. This was matched as close as possible to a link on Traffic Stats portal. A TomTom link is the distance between two nodes as shown in Figure 10.



Routes (2/20)

Selected map version: 2022.09 • Total length: 0.83 km

_UrbanFCD_5005.Bottelary_EB

Length: 414.16 m

_UrbanFCD_5005.Bott ar WB

Length: 414.16 m

R304, Stellenbosch Local Municipality, 7600

Bottelary Road, Stellenbosch Local Municipality, 7

Select map version

Map version 2022.09 is not available because the selected date range is older than last 2 years

Figure 10: Traffic Stats route selection

The TomTom link was selected to ensure that the inductance loop detector was located along the TomTom link. To correlate with the benchmark data, the links were selected in both directions of flow to obtain directional traffic information in both directions of travel for each road segment.

vi) Order report

The final stage of the data collection process comprises ordering the summarised information of the route analysis. Order report tab is selected and only ordered once a confirmation is done by the supervisor.

Once the reported is confirmed and ordered, the data is available for download. TomTom historical data was downloaded as Microsoft Excel workbooks for analysis. The data was ready for comparison and no additional process was required. The average speeds received from the Western Cape Department of Transport were arithmetic speeds and therefore, TomTom arithmetic speeds were selected for comparison.

4.3 CONCLUSION

The data collection chapter explained how reference data from the Western Cape Department of Transport and FCD from TomTom were obtained. The procedure for ordering traffic data from TomTom was also highlighted.

CHAPTER 5 : DATA ANALYSIS AND RESULTS

5.1 INTRODUCTION

This chapter discusses the statistical analysis for measuring the accuracy of commercial FCD along the rural routes under investigation. Speed profiles are presented to show the variations in FCD and benchmark data. Statistical hypothesis testing was then conducted to verify if the benchmark and FCD average speeds were significantly different for rural and urban regions.

5.2 SPEED PROFILES

This section presents the speed profiles for both the rural and urban counting stations selected for this research study. TomTom average speeds are graphically compared to benchmark speeds aggregated over a 1-hour time interval.

5.2.1 Rural count stations

5.2.1.1 5012 Vredendal count station

The northbound and southbound speed profiles of the 5012 Vredendal count station are shown in Figure 11.

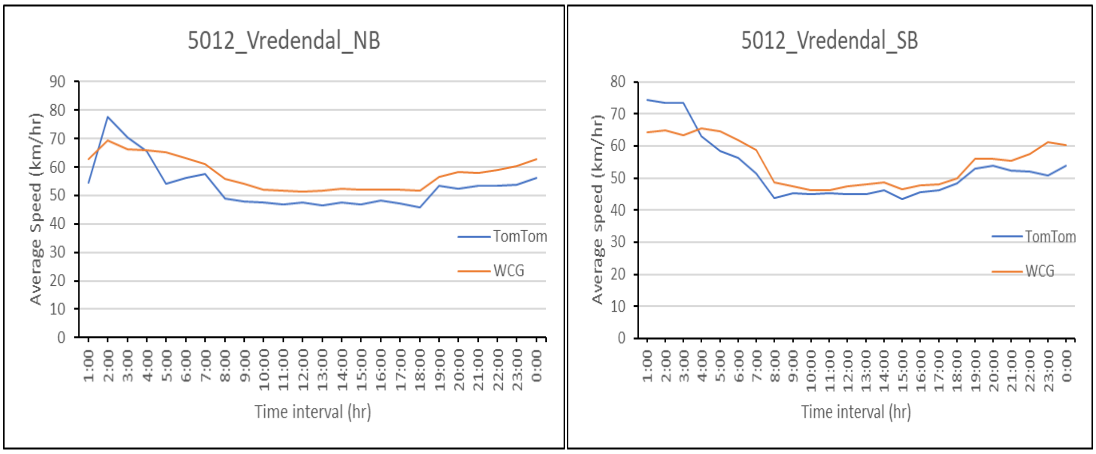


Figure 11: Speed profile for 5012_Vredendal count station

TomTom recorded higher speeds than WCG benchmark speeds both in the northbound and southbound between 01:00 and 04:00. From 04:00, the FCD reported speed are consistently lower than the benchmark speeds. There are also no pronounced peak times but there is noticeable speed reduction until 08:00 from which the speeds stagnate between 40 km/hr and 50 km/hr until 18:00. Thereafter, a slight speed increase was observed.

5.2.1.2 5013 Franschoek count station

The eastbound and westbound speed profiles of the 5013 Franschoek count station are shown in Figure 12.

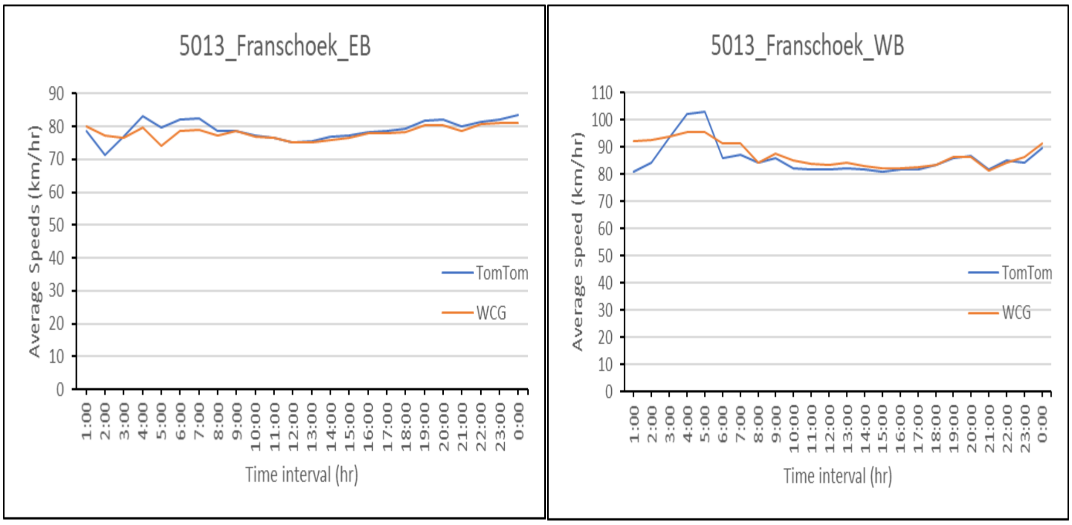


Figure 12: Speed profile for 5013_Franschoek count station

Relatively similar speeds were recorded from the 5013 Franschoek count station from 08:00.

It was observed that in the eastbound direction recorded speeds ranged between 70 km/hr and 85 km/hr while in the westbound, the traffic speeds were slightly higher ranging from 80 km/hr to 105 km/hr. However, the average speed was observed as 85 km/hr which was in line with the average of 100 km/hr within the area.

5.2.1.3 5016 Beaufort West count station

The eastbound and westbound speed profiles of the 5016 Beaufort West count station are shown in Figure 13.

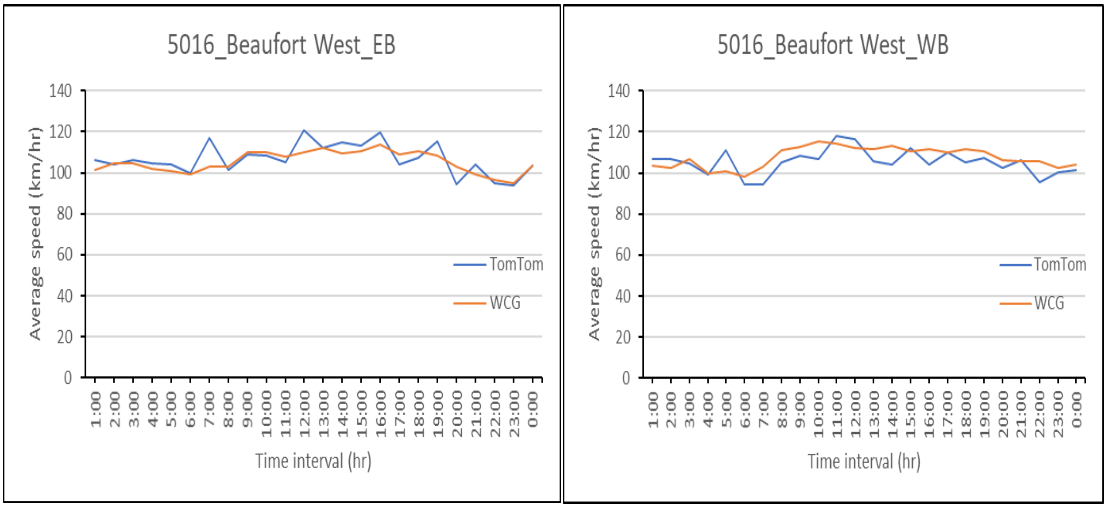


Figure 13: Speed profile for 5016_Beaufort West count station

The speeds in the eastbound and westbound had a matching profile with speeds recorded between 90 km/hr and 120 km/hr. Frequent 2-hour speed transitions were observed from FCD reported speeds with gradual increments from 06:00 to 16:00 and a slight decline thereafter. This was also the case with the benchmark speeds which recorded a parabolic speed increment to around 16:00 and reduced speeds were noticed until 24:00. This observation was different from the normal traffic patterns whereby free flow speeds are actualised before the morning peak and after the afternoon peak of which traffic speeds are usually relatively higher than the day speeds due to assumed traffic conditions.

5.2.1.4 5017 Kleinmond count station

The northbound and southbound speed profiles of the 5017 Kleinmond count station are shown in Figure 14.

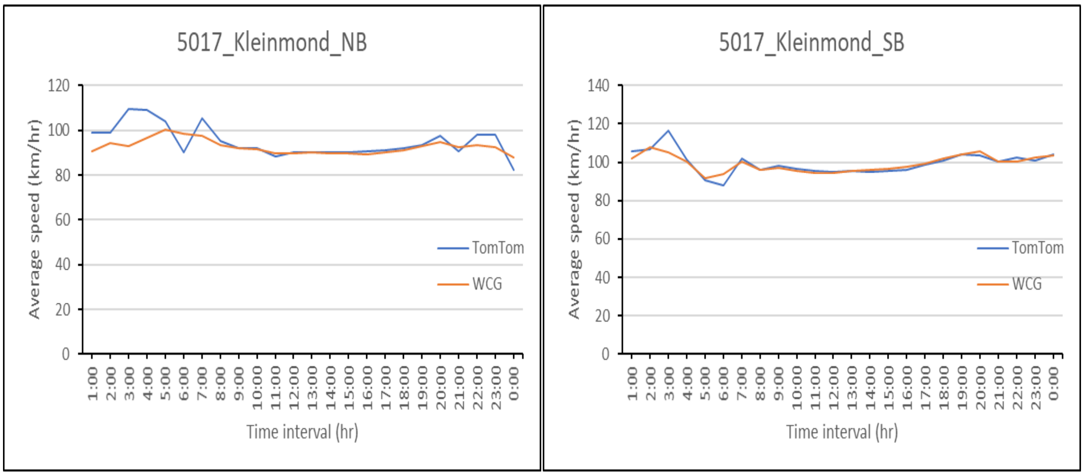


Figure 14: Speed profile for 5017_Kleinmond count station

5017 Kleinmond northbound realised a reduction in speeds from 05:00 to 11:00 after which an average speed of 90 km/hr was maintained until 19:00. It was also observed that higher FCD speeds were reported from 01:00 to 09:00 and from 19:00 to 24:00. The two methods of data collection used for this site recorded almost similar traffic speeds from 08:00 to 19:00.

In the southbound, there was a slight difference in speeds between 01:00 and 07:00. Thereafter, almost similar speeds were reported by FCD and benchmark data.

5.2.1.5 5022 Cango count station

The northbound and southbound speed profiles of the 5022 Cango count station are shown in Figure 15.

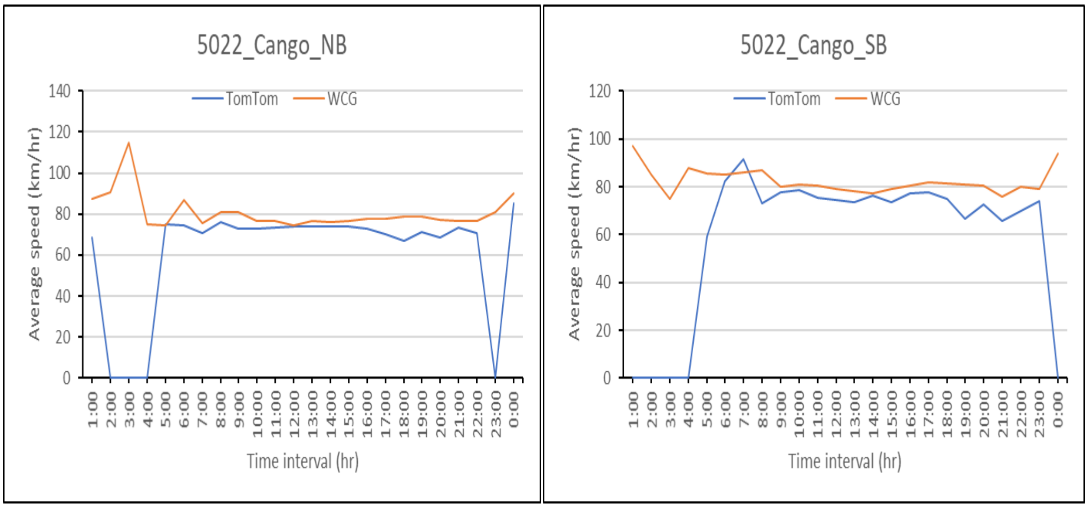


Figure 15: Speed profile for 5022_Cango count station

There were no FCD reporting devices present from 02:00 to 04:00 and at 23:00 in the northbound direction of the 5022 Cango count station. A similar observation was made in the southbound direction with no recorded speeds from 01:00 to 04:00 and between 23:00 and 24:00.

An average speed of 80 km/hr was observed for the benchmark recorded speeds both in the northbound and southbound while the FCD reported speeds ranged between 65 km/hr and 75 km/hr in both directions.

5.2.1.6 5028 Oudtshoorn count station

The eastbound and westbound speed profiles of the 5028 Oudtshoorn count station are shown in Figure 16.

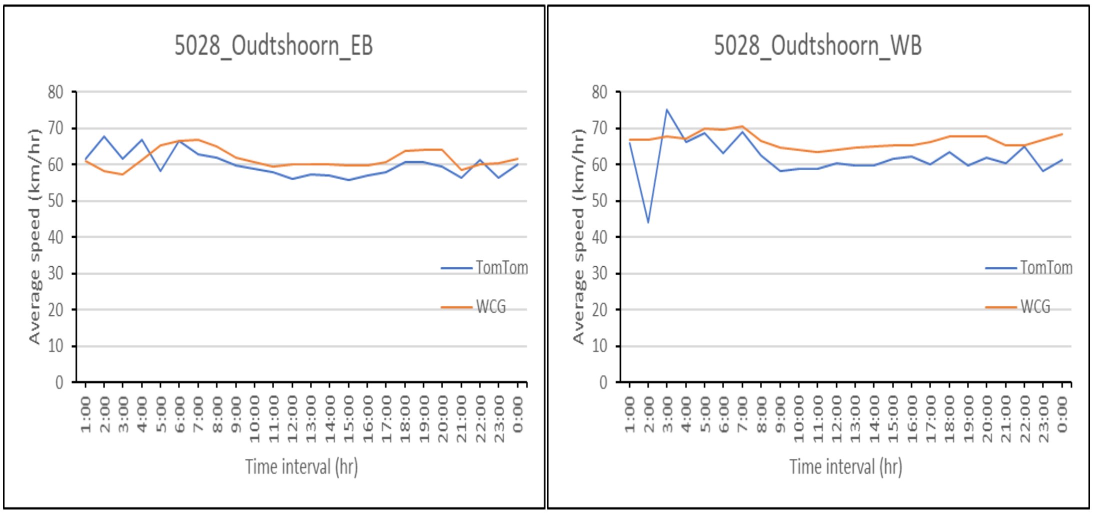


Figure 16: Speed profile for 5028_Oudtshoorn count station

In the eastbound, FCD reported speeds were higher than benchmark speeds from 01:00 to 04:00. It was observed that benchmark speeds slightly increased to 65 km/hr overlapping the FCD reported speeds. Higher benchmark speeds were then reported consistently throughout the rest of the day.

A steep reduction in speeds from 68 km/hr to 43 km/hr were observed between 01:00 and 02:00 and the speed increased again to 75 km/hr from 02:00 to 03:00. The speeds stabilized after 04:00 recording an average of 58 km/hr in the eastbound and 60 km/hr in the westbound.

5.2.1.7 5029 Atlantis count station

The northbound and southbound speed profiles of the 5029 Atlantis count station are shown in Figure 17.

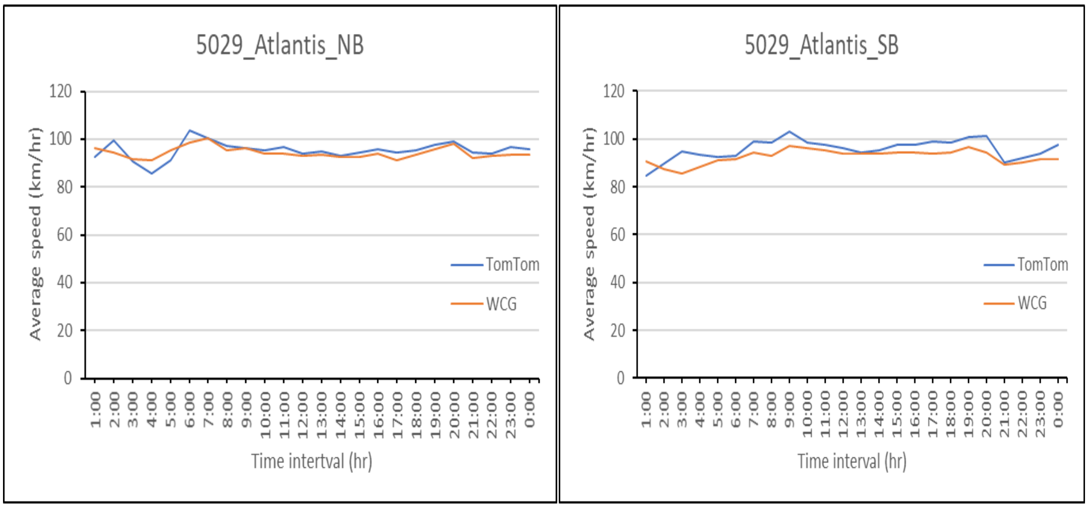


Figure 17: Speed profile for 5029_Atlantis count station

There was little variation in speeds noted throughout the analysis period with FCD speeds being consistently higher in the southbound. In the northbound, there was a short transition from 03:00 to 06:00 where benchmark recorded speeds were higher than FCD recorded speeds.

5.2.1.8 5050 Langebaan count station

The northbound and southbound speed profiles of the 5050 Langebaan count station are shown in Figure 18.

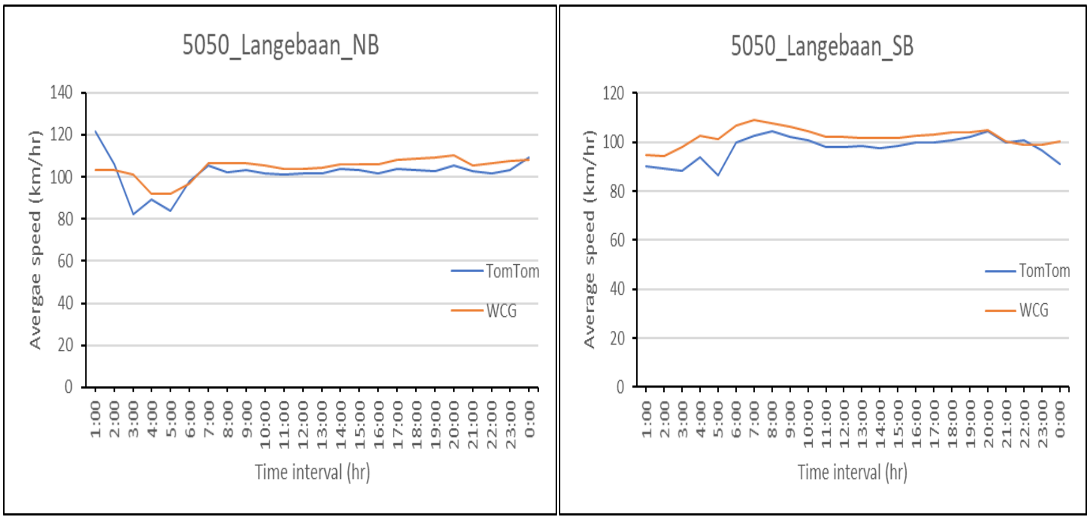


Figure 18: Speed profile for 5050_Langebaan count station

Lowest speeds of the 5050 Langebaan count station were recorded in the early hours of the day from 01:00 to 07:00. There were no pronounced peak durations and an average speed of 103 km/hr was observed from the Northbound speed profile while an average speed of 100 km/hr was observed in the Southbound. The FCD speeds were also consistently lower during the analysis period.

5.2.1.9 5054 Ceres North count station

The northbound and southbound speed profiles of the 5054 Ceres North count station are shown in Figure 19.

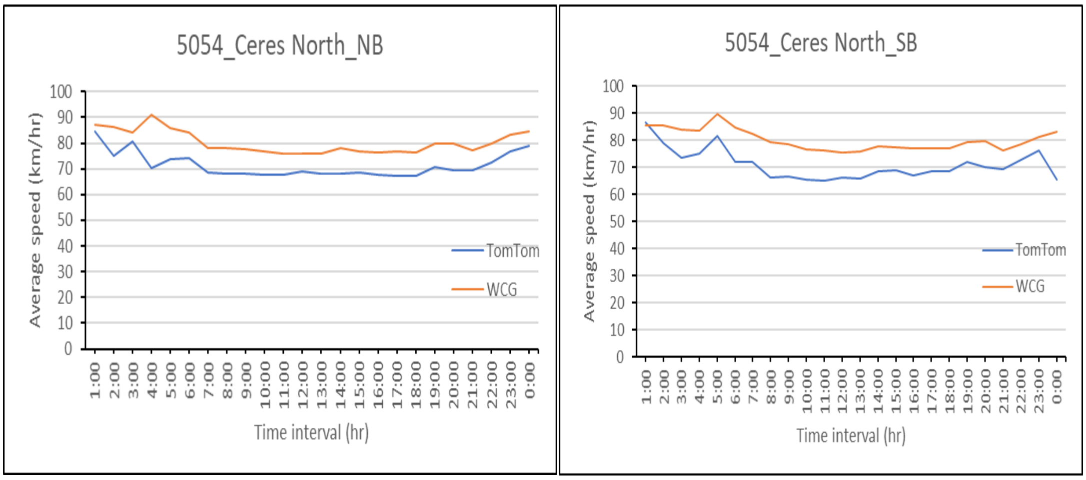


Figure 19: Speed profile for 5054_Ceres North count station

A difference of 10 km/hr was observed between FCD speeds and benchmark speeds for both the northbound and southbound speed profiles. This was consistent throughout the analysis period with no definitive peak periods. However, there was a notable speed decrease from 06:00

to an average speed of 69 km/hr recorded from the FCD reporting devices and 79 km/hr recorded by the inductance loops. The average speeds were maintained until 19:00.

5.2.1.10 5056 Montagu count station

The eastbound and westbound speed profiles of the 5056 Montagu count station are shown in Figure 20.

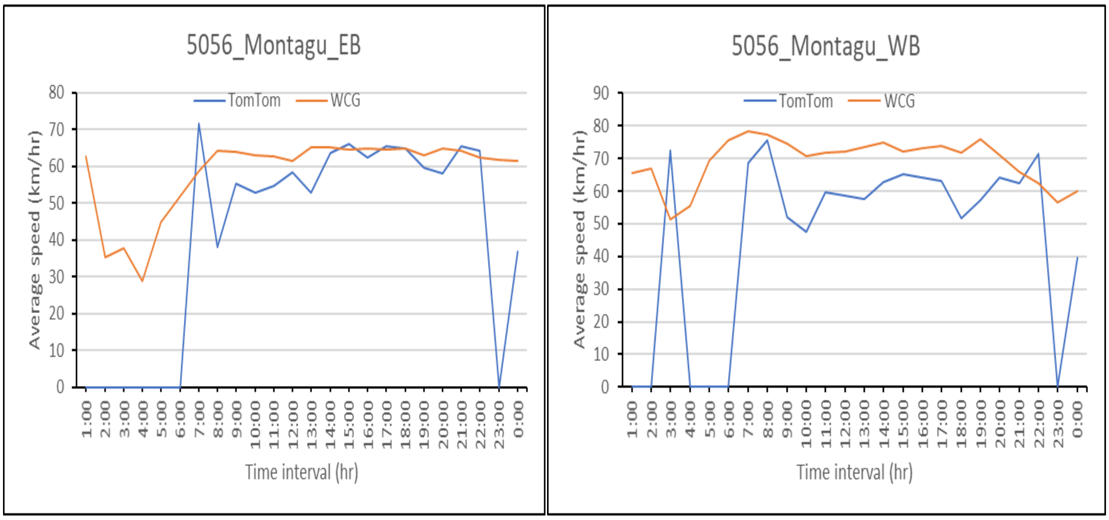


Figure 20: Speed profile for 5056_Montagu count station

There were no FCD reported speeds observed in the eastbound direction between 01:00 and 06:00 and at 23:00 along the 5056 Montagu count station. Speeds were also observed to be as low as 30 km/hr before 08:00 and increased to an average of 65 km/hr which was constant throughout the day. There was a little variation between the FCD speeds and the benchmark speeds with FCD recording lower speeds.

In the westbound, there were no FCD recorded speeds at 01:00 to 02:00, 04:00 to 06:00 and at 23:00. Also, the FCD reported speeds were consistently lower than the benchmark speeds in the larger duration of the analysis period. The speed difference, particularly in the westbound direction is substantial, with FCD indicated to be between 10 and 20 km/hr below the benchmark speed.

5.2.1.11 5066 Ceres Karoo count station

The eastbound and westbound speed profiles of the 5066 Ceres Karoo count station are shown in Figure 21.

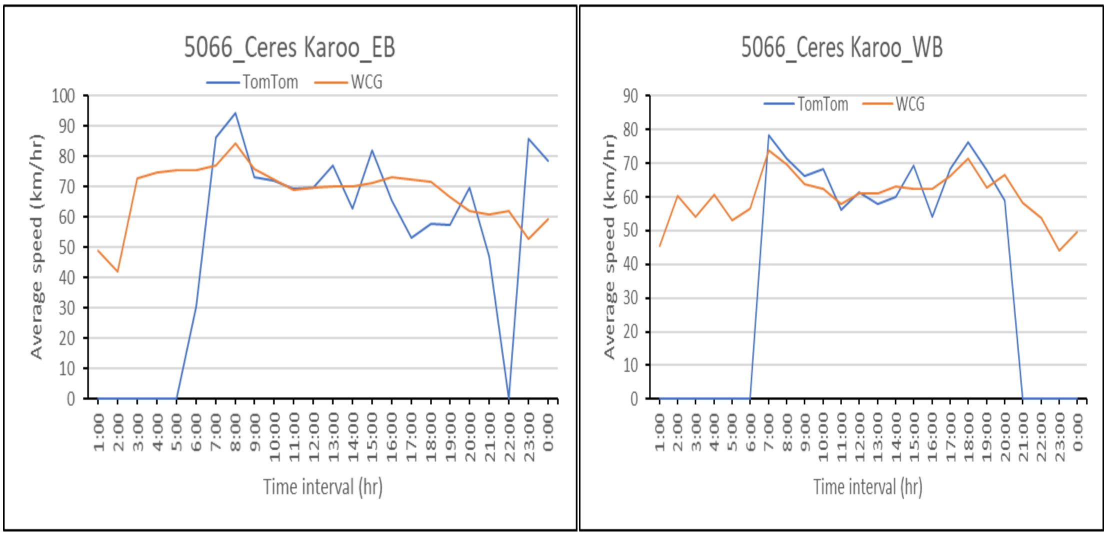


Figure 21: Speed profile for 5066_Ceres Karoo count station

A trend of frequent increase and decrease in average speeds was noted and especially from FCD speeds both in the eastbound and westbound. In the eastbound, there were no recorded FCD speeds from 01:00 to 05:00 and at 22:00 while in the westbound, there were missing FCD speeds from 01:00 to 06:00 and from 21:00 to 24:00. It was also observed that the speeds were interchangeably varying between the two datasets.

5.2.1.12 5067 Saron count station

The northbound and southbound speed profiles of the 5067 Saron count station are shown in Figure 22.

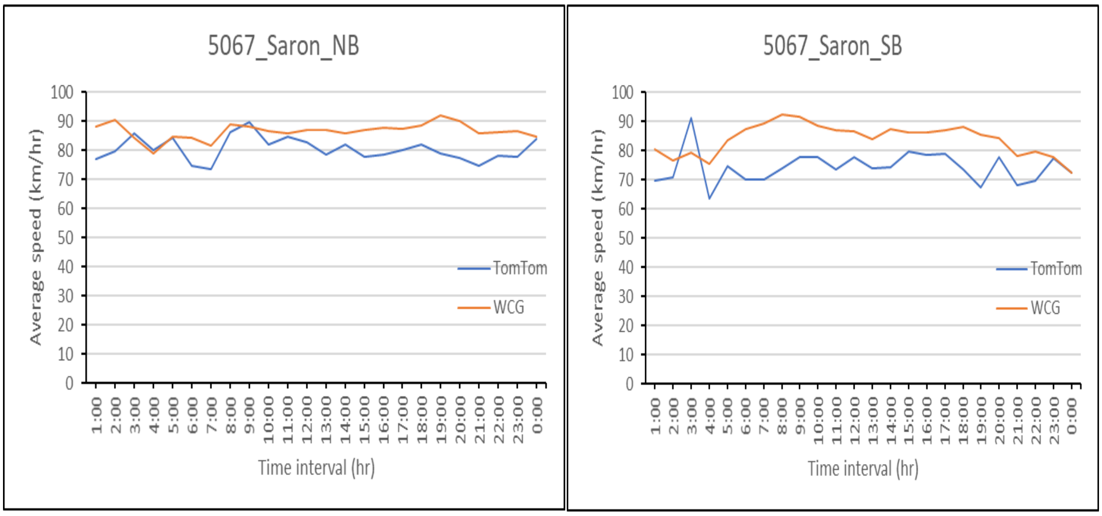


Figure 22: Speed profile for 5067_Saron count station

The average speeds were observed to be constant throughout the analysis period with the benchmark speeds recorded higher than the FCD reported speeds. A zigzag trend is notable from both the northbound and southbound speed profiles indicating frequent increase and decrease in traffic speeds along the 5067 Saron count station.

5.2.2 Urban count stations

5.2.2.1 1339 Croydon count station

The eastbound and westbound speed profiles of the 1339 Croydon count station are shown in Figure 23.

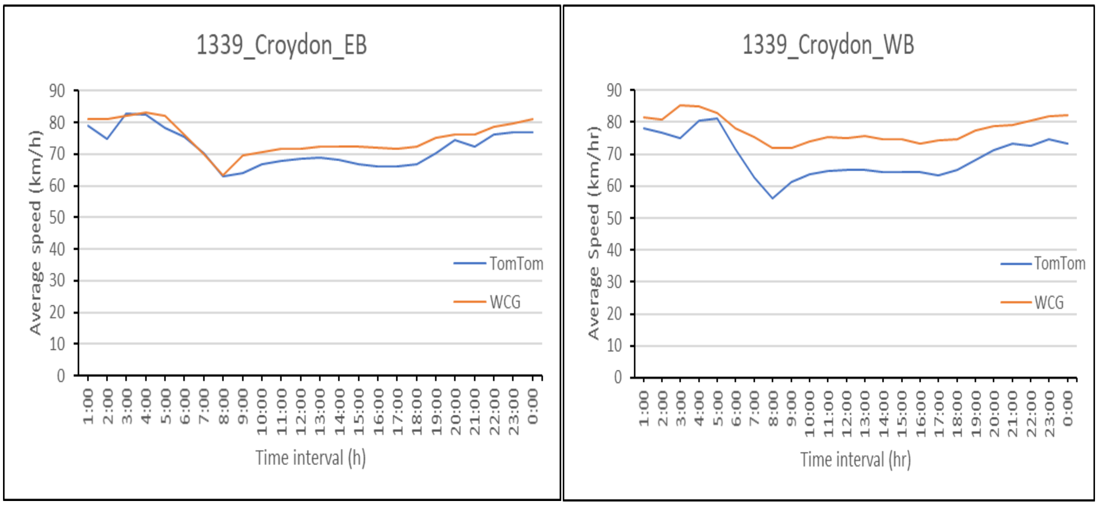


Figure 23: Speed profile for 1339_Croydon count station

Along the 1339 Croydon count station, there is a notable morning peak period between 04:00 and 09:00 characterised by a reduction in speeds from 82 km/hr to 64 km/hr in the eastbound direction and from 80 km/hr to 58 km/hr in the westbound direction. It was also noted that the datasets reported similar speeds during the morning peak in the eastbound direction from 06:00 to 08:00. It is interesting to note that the morning peak is visible in both directions of travel, however there is no clear afternoon peak, indicating that there is likely no congestion occurring in the afternoon peak which could influence speeds.

In the westbound, FCD speeds were recorded with a larger variation during the morning peak. The speeds changed from 80 km/hr at 05:00 to 58 km/hr at 08:00 while benchmark speeds changed from 85 km/hr at 03:00 to 72 km/hr at 08:00. FCD recorded lower speeds than benchmark speeds throughout the analysis period.

5.2.2.2 5001 Macassar count station

The northbound and southbound speed profiles of the 5001 Macassar count station are shown in Figure 24.

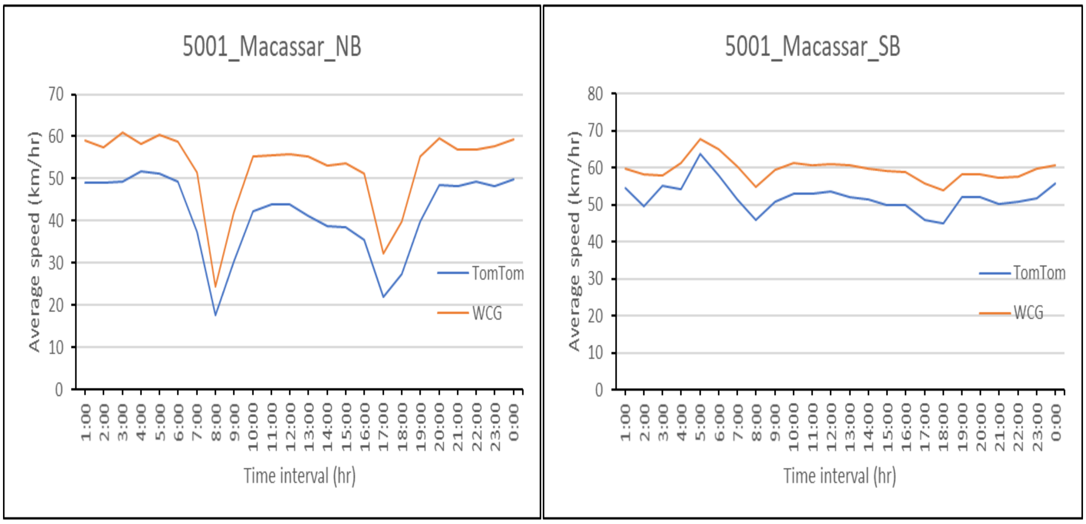


Figure 24: Speed profile for 5001_Macassar count station

Definitive morning and evening peak periods were observed along the 5001 Macassar urban count station with FCD recording lower speeds than benchmark speeds.

Looking at the FCD reported speeds in the northbound, there was a change in the recorded speeds from 50 km/hr to 18 km/hr and increased to 44 km/hr between 05:00 and 10:00. The speeds then plateaued until the evening peak between 15:00 to 20:00 where the dropped to a low of 22 km/hr 17:00. The traffic assumed free flow speeds after 20:00.

In the southbound, there were notable morning and evening peaks but not as pronounced as the northbound peaks. The morning peak period was noted between 05:00 and 10:00 while the evening peak was observed between 15:00 and 20:00.

5.2.2.3 5057 Eikendal count station

The northbound and southbound speed profiles of the 5057 Eikendal count station are shown in Figure 25.

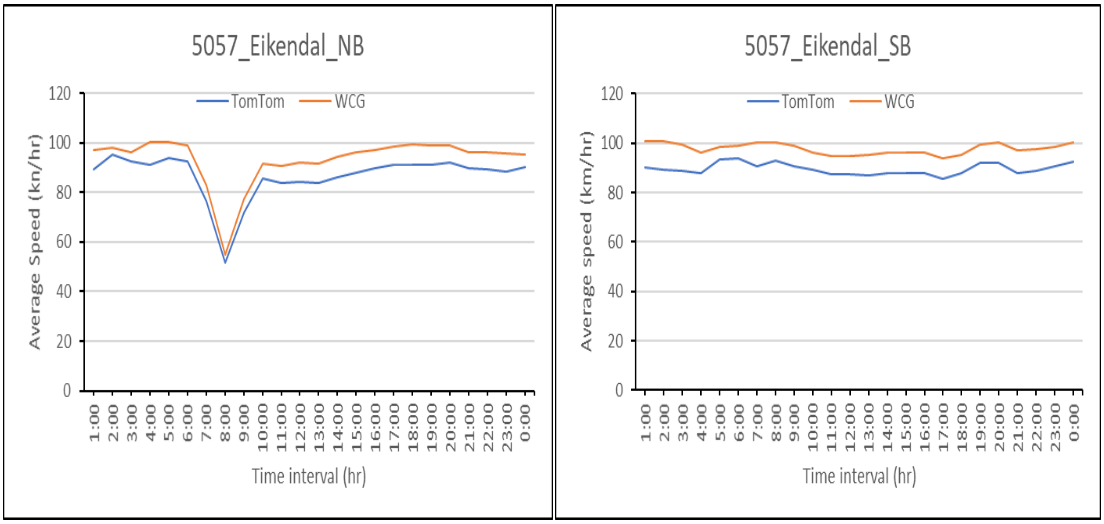


Figure 25: Speed profile for 5057_Eikendal count station

The difference in the northbound and southbound speed patterns indicate the different traffic patterns along the 5057 Eikendal count station. In the northbound, there was only one established peak in the morning from 06:00 to 10:00. The speeds were then averaged out through the rest of the analysis period.

In the southbound direction, there was a slight variation in the average speeds during peak times. However, the average speeds were consistent throughout the analysis period with a difference of 10km/hr consistently observed between the two datasets.

5.2.2.4 5058 Koeberg NB count station

The northbound speed profile of the 5058 Koeberg count station is shown in Figure 26.

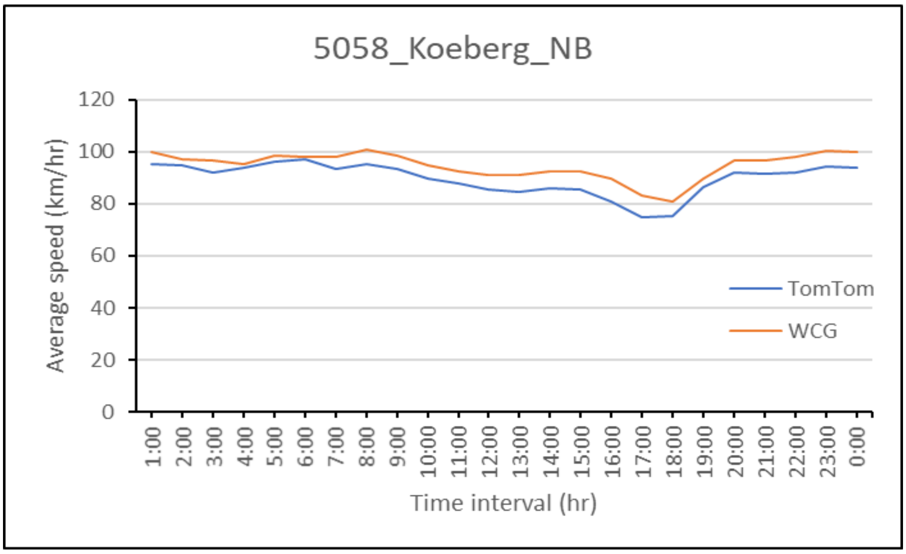


Figure 26: Speed profile for 5058_Koeberg count station

In the northbound direction, it was observed that traffic peaks in the evening hours between 15:00 and 20:00. The FCD reported speeds are also consistently lower than the benchmark average speeds. A maximum average speed 100 km/hr was also observed from the WCG data which is in line with the recommended travel speed along the N1.

5.2.2.5 5059 Koeberg SB count station

The southbound speed profile of the 5059 Koeberg count station is shown in Figure 27.

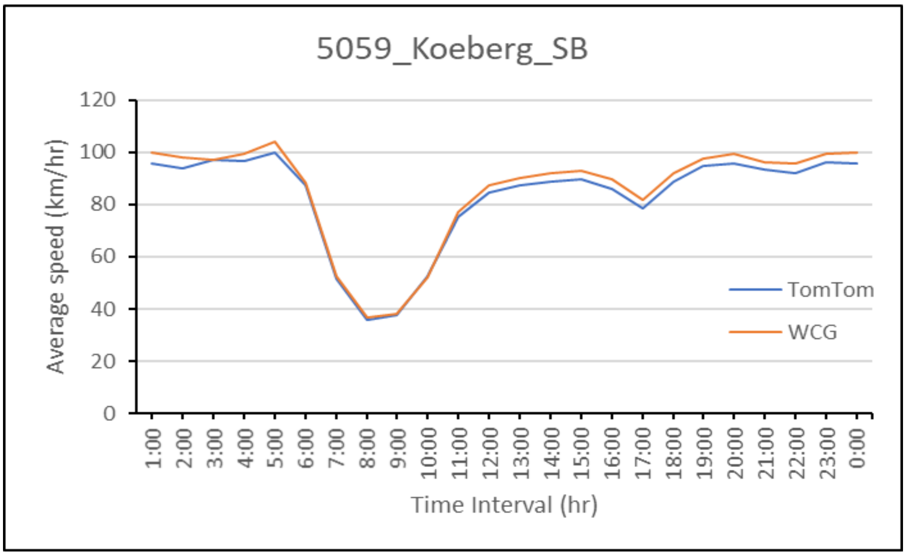


Figure 27: Speed profile for 5059_Koeberg count station

In the southbound, there is a defined morning peak period between 05:00 and 11:00. TomTom and WCG recorded similar average speeds during the morning peak period. There was a slight variation in speeds observed between FCD reported speeds and the benchmark speeds with FCD recording lower speeds throughout the analysis period. Moreover, an evening peak period was observed between 15:00 and 19:00 with FCD speeds decreasing to 82 km/hr between 16:00 and 17:00.

5.2.2.6 5064 Airport Approach count station

The eastbound and westbound speed profiles of the 5064 Airport Approach count station is shown in Figure 28.

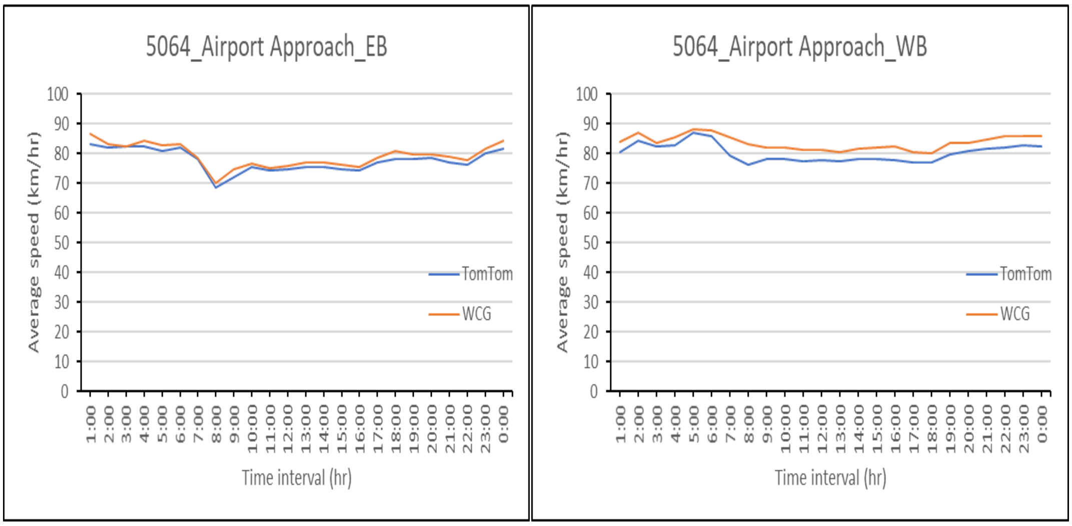


Figure 28: Speed profile for 5064_Airport Approach count station

In the eastbound of the 5064 Airport approach road counting station, three peak periods were observed. The peak periods were observed from 06:00 to 10:00, 15:00 to 18:00, and 20:00 to 23:00. The two datasets reported almost similar average speeds with the FCD reported speeds being lower than the benchmark speeds.

In the westbound, there was a slight variation in speeds of about 5 km/hr between the FCD speeds and benchmark speeds. Also, a morning peak period between 06:00 and 09:00 and an evening peak period between 16:00 and 19:00 were observed.

5.2.2.7 5075 Polkadraai count station

The northbound and southbound speed profiles of the 5075 Polkadraai count station is shown in Figure 29.

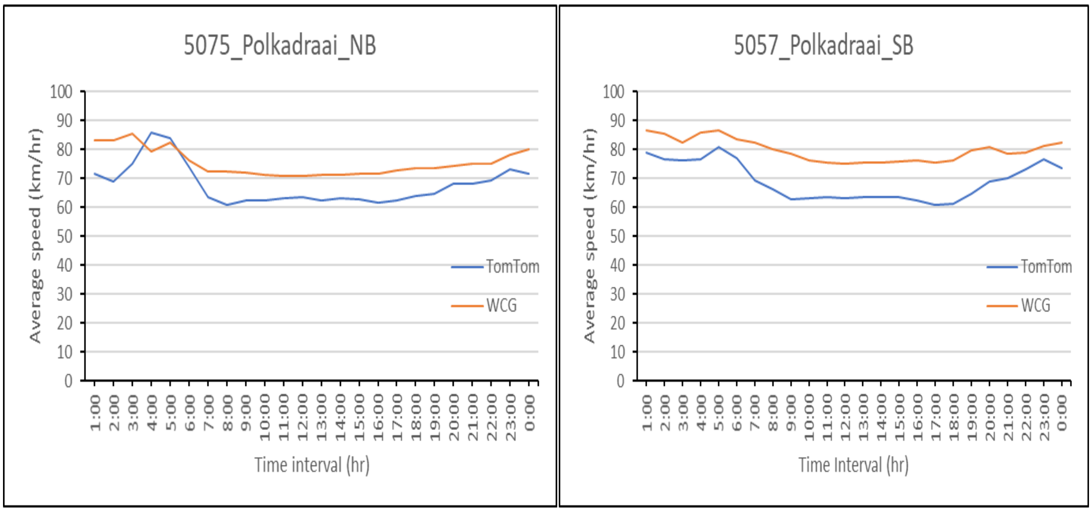


Figure 29: Speed profile for 5057_Polkadraai count station

At 5075 Polkadraai count station, a morning peak period was observed from 05:00 to 09:00 in both the northbound and southbound. Thereafter, the speeds stagnated at 70 km/hr in the northbound and 75 km/hr in the southbound until the evening peak period. The evening peak period was observed from 15:00 to 19:00.

FCD reported speeds were observed to be lower than the benchmark speeds. A difference of 10 km/hr was observed in the southbound direction.

5.3 ACCURACY MEASURES

Accuracy measures the extent of closeness between the certain parameter(s) and benchmark value(s). The selection of accuracy measures is based on the evaluation scenario. Evaluation scenarios include link speeds, route travel time and congestion category (Turner, et al., 2011). In this study, link speeds were selected as the evaluation criteria.

The accuracy measures linked to FCD reported speeds are;

1. Penetration rate;
2. Average absolute error (AASE) (units km/h);
3. Average error, also known as signed error bias (SEB) (units km/h);
4. Signed Error (%).

The level of quality to be achieved is not defined in the measurements of data accuracy guidelines and therefore it lies at the discretion of the user. The type of data and its application helps in determining the methods required to calculate the quality measures and the thresholds for evaluating the data quality (Turner, 2002) (Turner, et al., 2011). The evaluation criteria were similar to the analysis done by Gwara and Andersen (2017) and the maximum allowable error for the accuracy measures which was used in their study, listed in Table 12, are applied to this research as well.

Table 12: Maximum allowable error

Accuracy measures	Maximum allowable error
AASE	10 km/h
SEB	±7.5 km/h
Signed Error	±10 %

In this study, the average absolute speed error and signed error were adopted to show the differences between TomTom speed estimates and benchmark average speeds. Computation of the error measurements was discussed in Section 2.6.1.8. Turner et. Al (2011) recommended the use of arithmetic averages for computation when using root mean square error, average absolute error and average error for speed accuracy measurements. Therefore, arithmetic speeds from FCD were chosen for consistency in data analysis.

5.3.1 Penetration rate

Probe penetration rates are used to inform on the number of vehicles in a traffic population that are reporting FCD. Recent studies have shown that urban areas have a higher probe penetration hence leading to collection of more accurate FCD (Ahsani, et al., 2019) (Valadkhani, et al., 2017). More data sample points were collected for the rural area to have a larger rural study area and consequently improve the penetration rate.

Probe penetration rates were calculated for each directional segment and averaged per count stations. The results were then averaged per study region, that is, urban region and rural region to evaluate the respective penetration rates. Table 13 indicates the calculated rural penetration rates.

Table 13: Rural area penetration rates

RURAL (FEB 2019)			
	NB/EB	**SB/WB**	
Location	**Penetration Rate (PR)**		**Average Penetration Rate (%)**
5012_Vredendal	5.54	5.50	5.52
5013_Franschoek	13.14	13.44	13.29
5016_Beaufort West	7.08	8.94	8.01
5017_Kleinmond	12.63	12.74	12.69
5022_Cango	14.11	10.33	12.22
5028_Oudtshoorn	8.10	7.17	7.64
5029_Atlantis	9.00	10.89	9.95
5050_Langebaan	11.26	13.16	12.21
5054_CeresNorth	6.75	6.38	6.57
5056_Montagu	2.82	3.72	3.27
5066_CeresKaroo	4.64	9.24	6.94
5067_Saron	7.09	6.98	7.04
		Rural PR	**8.78**

It is interesting to note that the penetration rate at location 5056 Montagu is very low compared to other locations at only 3.27%. This location also indicated big differences in speeds measured by FCD and the counting station, which may indicate that penetration rate impacts FCD speed accuracy.

Table 14 indicates the calculated urban area penetration rates. The highest observed urban penetration rate was along the Airport approach road which was probably due to an expected high population of people using smartphones to estimate their estimated arrival time to the airport.

Table 14: Urban area penetration rates

URBAN (FEB 2019)			
	NB/EB	SB/WB	
Location	Penetration Rate (PR)		Average Penetration Rate (%)
1339_Croydon	13.34	10.96	12.15
5001_Macassar	13.87	13.29	13.58
5057_Eikendal	13.78	14.43	14.11
5058_KoebergNB	23.90		23.90
5059_KoebergSB		14.13	14.13
5064_Airport	27.55	29.16	28.36
5075_Polkadraai	12.08	11.89	11.99
		Urban PR	**16.89**

From the presented results in Table 13 and

Table **14**, it was observed that the rural area had a penetration rate of 8.78% while the urban area had a penetration rate of 16.89%. As expected, the population reporting FCD speeds was higher in the urban region compared to the rural region.

Studies conducted by Sanwal and Walrand (2016) and Srinivasan and Jovanis (1996) indicated that a probe penetration rate of 4% and 5% of the total vehicle population was sufficient for accuracy and reliability studies. Therefore, the probe penetration rate evaluated in the rural area was adequate to quantify the accuracy of traffic speeds reported in the rural area.

A regression analysis was conducted to evaluate if the penetration rate differs significantly between SEB and AASE in rural and urban areas. An ANOVA test indicated that penetration rate differed significantly between the two accuracy measures in the rural area (F=0.0035) which was less than 0.05 and there was no significant difference in the urban area (F= 0.078) being greater than 0.05.

5.3.2 Accuracy measures per region

The SEB and AASE were calculated per region that is, rural and urban. Table 15 shows the quantified SEB, AASE and SE (%) values for the rural routes per directional segment while Table 16 shows similar measurements for the urban routes.

Additional filtering of the data was done during the evaluation of the accuracy measures. In the case where the TomTom average speed largely differed from the benchmark average speeds, that counting station was considered as an outlier and removed from the analysis. This high difference in the average speeds reported by TomTom and by the WCG count station devices led to SEB and AASE values above the maximum allowable error values. This was visible in the 5026_Lemoenfontein station which was a rural route and in 5005_Bottelary which was an urban route.

In both instances, the reason leading to the large difference in the average speeds was that the counting station, was close to an intersection and the TomTom link extended to the intersection. The average speeds reported by the TomTom data from such a link cannot give a true representation of the traffic pattern due to drivers slowing down on approach to the intersection. This resulted in the very high SEB and AASE values, highlighted in red in Table 15 and Table 16. The results for these two locations were therefore not included in the further analysis.

Table 15: Rural routes accuracy measures

RURAL (FEB 2019)						
	NB/EB			SB/WB		
Location	SEB	AASE	SE (%)	SEB	AASE	SE (%)
5012_Vredendal	-4.414	5.464	-7.917	-2.069	4.446	-4.029
5013_Franschoek	0.991	1.611	1.269	-1.329	2.660	-1.527
5016_Beaufort West	1.479	3.806	1.388	-2.335	4.673	-2.116
5017_Kleinmond	2.343	3.806	2.486	0.304	1.668	0.269
5022_Cango	-5.893	5.934	-7.309	-6.505	7.069	-7.997
5026_Lemoenfontein	-17.030	17.030	-23.071	-23.151	23.151	-28.816
5028_Oudtshoorn	-1.509	3.283	-2.349	-4.653	5.261	-7.009
5029_Atlantis	1.069	2.253	1.129	3.087	3.619	3.338
5050_Langebaan	-2.843	4.795	-2.737	-4.490	4.651	-4.406
5054_CeresNorth	-8.913	8.913	-11.091	-9.109	9.200	-11.424
5056_Montagu	-5.212	7.365	-8.136	-9.832	13.220	-12.969
5066_CeresKaroo	-1.255	12.024	-0.611	0.766	4.121	1.066
5067_Saron	-6.136	6.494	-7.010	-9.645	10.642	-11.226

Other rural and urban routes were evaluated to SEB and AASE results that were within the maximum allowable error limits provided in Table 12. This showed a good estimation on the traffic data used for traffic accuracy measurements. The SE (%) were within the 10% maximum allowable error indicating that the data was also accurate to be considered for traffic planning studies (Gwara & Andersen, 2018).

Table 16: Urban routes accuracy measures

URBAN (FEB 2019)						
	NB/EB			SB/WB		
Location	**SEB**	**AASE**	**SE (%)**	**SEB**	**AASE**	**SE (%)**
1339_Croydon	-3.262	3.352	-4.383	-8.656	8.656	-11.315
5001_Macassar	-11.151	11.151	-21.808	-7.407	7.407	-12.526
5005_Bottelary	-30.997	30.997	-43.504	-17.216	17.216	-23.755
5057_Eikendal	-6.650	6.650	-7.112	-8.202	8.202	-8.387
5058_KoebergNB	-5.031	5.031	-5.376			
5059_KoebergSB				-2.594	2.645	-2.886
5064_Airport	-1.633	1.648	-2.056	-3.474	3.474	-4.165
5075_Polkadraai	-7.478	8.173	-10.022	-10.802	10.802	-13.679

It is also observed that SEB results for urban area are consistently negative while in the rural area there are some cases where the SEB values are positive and others negative. Consistent negative SEB values indicate that TomTom average speeds collected in urban regions were lower than the benchmark average speeds. Graphical illustrations of the observations are presented in Section 5.2.

The SEB and AASE were averaged for rural and urban areas, and are tabulated in Table 17. The accuracy measure values are within the maximum allowable limits indicating that the TomTom speed data is adequately accurate in both rural and urban zones. To confirm if the urban and rural accuracy measures are significantly different, a hypothesis test was conducted.

Table 17: Rural and Urban accuracy measures

	SEB (km/h)	**AASE (km/h)**
Rural	-3.171	5.707
Urban	-6.362	6.433

Box plots captured in Figure 30 indicate the variation observed in the accuracy measures between urban and rural locations. The SEB compares how FCD reported speed estimates vary from benchmark speeds. In Figure 30, the SEB plots for both rural area and urban area indicate that most FCD speed estimates were lower than the benchmark speeds (SEB<0). However, some counting stations in the rural area recorded positive SEB values indicated by the difference observed in the rural and urban SEB plot. 4 out of 12 rural count stations recorded positive SEB while no positive SEB was observed in the urban area as shown in Table 16.

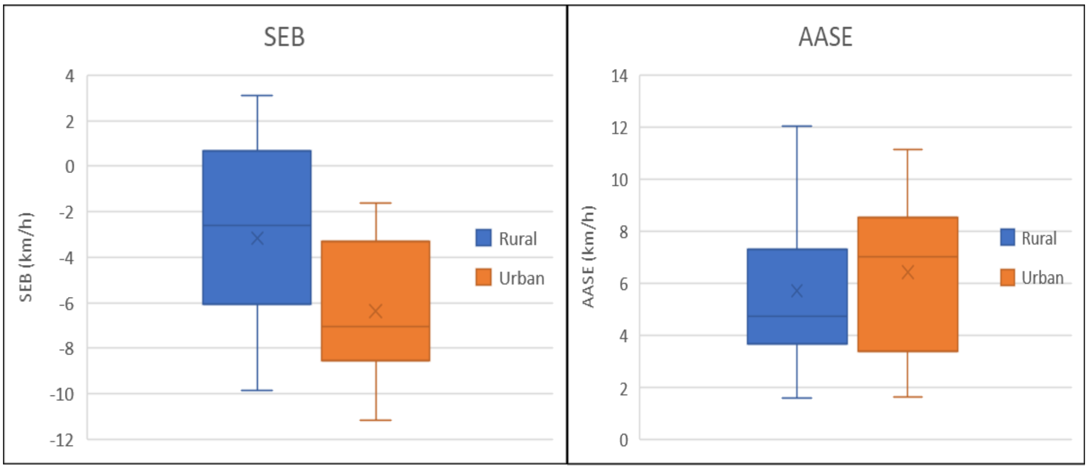


Figure 30: Variation in accuracy measures SEB and AASE per region

AASE indicate the extent to which FCD and benchmark speed data vary. The highest variation was measured in the urban region (AASE = 6.43 km/hr) showing that the FCD reported speeds are about 6 km/hr different to benchmark speeds. A lower AASE measurement was recorded for the rural area (AASE = 5.70 km/hr) indicating that the measurements from the rural area were more accurate than the urban region measurements. Significance of these variations are looked into in the following section.

5.3.3 Hypothesis testing

A null hypothesis test was conducted to evaluate if the SEB and AASE determined in the rural regions were significantly different from those measured in the urban regions. A 95% confidence interval was applied in the hypothesis testing. In all analyses, the null hypothesis was that the SEB and AASE in the rural area were not significantly different from SEB and AASE in the urban area.

5.3.3.1 F-Test to determine variance

An F-test was first conducted to determine if the variance between the urban and rural samples were equal or unequal. This helped in deciding which T-test to use for both SEB and AASE variables. A T-test for equal variance is applied if the F-statistic (ratio between the larger sample variance and the smaller sample variance) is less than 2 or if $F < F_{Critical}$.

The test was conducted using the data analysis tool pack in Microsoft Excel. The variances of SEB and AASE were calculated for rural and urban areas so as to differentiate which location had the higher and lower variance.

Table 18: SEB F-test results

F-Test Two-Sample for Variances		
	Rural	*Urban*
Mean	-3.17	-6.36
Variance	15.54	10.00
Observations	24	12
df	23	11
F	1.554	
P(F<=f) one-tail	0.226	
F Critical one-tail	2.617	

F-test results for SEB are captured in Table 18. The results showed that the SEB [F(1.554) < $F_{Critical}$ (2.617)] and consequently SEB [F(1.554) < 2] indicating that the variances are statistically similar.

F-test results for AASE are captured in Table 19. AASE [F(1.019) < $F_{Critical}$ (2.236)] and consequently SEB [F(1.019) < 2] . Both SEB and AASE have F-values less than $F_{Critical}$ values. These analyses confirm that a T-test with equal variance should be conducted to evaluate the difference or similarity between urban and rural AASE and SEB.

Table 19: AASE F-test results

F-Test Two-Sample for Variances		
	Urban	*Rural*
Mean	6.43	5.71
Variance	10.08	9.90
Observations	12	24
df	11	23
F	1.019	
P(F<=f) one-tail	0.461	
F Critical one-tail	2.236	

5.3.3.2 T-test with equal variance

The T-test with equal variance was conducted using the data analysis tool pack in Microsoft Excel. In checking and validating T-test results, the P-value is considered. P < 0.05 indicates that there is a significant difference between the parameters under comparison and the null hypothesis is rejected while a value of P > 0.05 indicates that the parameters are not significantly different and the null hypothesis is maintained.

Since the aim of the T-test is to evaluate if SEB and AASE values in the rural area were significantly different from the urban area, a two tailed t-test was used to verify the results. Table 20 indicates the results of the SEB t-test. The P-value was less than 0.05 confirming that SEB in rural areas is significantly different from SEB in urban areas. Therefore, the null hypothesis for SEB was rejected.

This is supported by information displayed in Figure 30. The mean value for rural region was -3.17 km/h while for urban region was -6.36 km/h. The box plot diagram illustrates the difference in mean values showing the significant difference. The SEB value in urban area is more on the negative side indicating that FCD reported in actually lesser than the benchmark speed in urban area than in rural area.

Table 20: SEB T-test solution

t-Test: Two-Sample Assuming Equal Variances		
	Rural	*Urban*
Mean	-3.17	-6.36
Variance	15.54	10.00
Observations	24	12
Pooled Variance	13.75	
Hypothesized Mean Difference	0	
df	34	
t Stat	2.434	
P(T<=t) one-tail	0.010	
t Critical one-tail	1.691	
P(T<=t) two-tail	0.020	
t Critical two-tail	2.032	

Table 21 shows the results of the AASE T-test. The P-value was greater than 0.05 indicating that there was no significant difference between AASE from rural areas and urban areas. In this case, the null hypothesis was accepted. This indicated that mean speeds derived from FCD and benchmark data from rural areas and urban areas are relatively similar. The box plot on the right side of Figure 30 shows that the mean is relatively the same with values of 6.43 km/h for urban region and 5.71 km/h for rural region.

Table 21: AASE T-test solution

t-Test: Two-Sample Assuming Equal Variances		
	Urban	*Rural*
Mean	6.43	5.71
Variance	10.08	9.90
Observations	12	24
Pooled Variance	9.96	

Hypothesized Mean Difference	0	
df	34	
t Stat	0.650	
P(T<=t) one-tail	0.260	
t Critical one-tail	1.691	
P(T<=t) two-tail	0.520	
t Critical two-tail	2.032	

5.4 CONCLUSION

The results of the study were discussed in this chapter. SEB values indicated that FCD consistently recorded lower estimates compared to the benchmark speeds for both rural and urban regions but with the urban region recording much lower estimates. A study conducted by Bruwer et al. (2022) on urban freeways in six cities in South Africa, showed that five out of the six cities had FCD underestimating the average speeds evaluating a range of SEB values from -2.74 km/h to 0.36 km/h.

Gwara and Andersen (2018) combined the SEB and AASE results for six freeway segments investigated in Gauteng and resulted in -6.3 km/h and 6.5 km/h averages for 1-hour time interval respectively. This indicated that FCD speeds are underestimated in different regions. The magnitude of variation provided by AASE was relatively similar to 6.43 km/h evaluated in this study. However, the AASE for rural area in this study resulted to 5.71 km/h which indicated minimal variation in magnitude between urban regions and rural regions concluded from this study.

Accuracy measurements were within the maximum allowable error showing that the travel speed analysis in urban and rural area derived conclusive and accurate results. A null hypothesis stating that AASE from rural areas was not significantly different from AASE in urban areas was accepted. This indicated that the difference in mean speeds derived from FCD and benchmark data is relatively similar in rural areas and urban areas. However, the null hypothesis on SEB accuracy was rejected showing that there were notable differences in TomTom and Benchmark speeds between rural and urban areas. Positive SEB speeds were recorded in the rural areas indicating that benchmark speeds were higher on some rural roads.

CHAPTER 6 : CONCLUSION AND RECOMMENDATIONS

The aim of this study was to investigate the accuracy of commercial FCD in the rural areas of the Western Cape in South Africa. The only FCD accuracy study that has been previously conducted in South Africa was conducted by Gwara and Andersen (2018) which considered the accuracy of FCD on urban freeways in Gauteng. No studies have determined the accuracy of FCD in rural areas in South Africa.

SEB and AASE accuracy measures were evaluated in the investigation. Data was analysed from 12 rural routes and 7 urban routes. The magnitude of speed error measured by AASE and SEB was within the allowable range, indicating that the FCD was sufficiently accurate at all locations in both the urban and rural areas. AASE measures were similar in both urban and rural zones. On the contrary, SEB measurements differed significantly between urban and rural areas. Factors leading to the speed bias may include the number of vehicles reporting FCD, service providers underestimating FCD speeds, socioeconomic characteristics between regions and driver behaviour in different regions being influenced by both internal and external factors such as age and traffic rules respectively.

Probe penetration rates for both urban and rural areas were calculated to investigate if levels of penetration affect accuracy of FCD reported speeds. It was observed that a high probe penetration rate is not directly proportional to a high level of speed data accuracy. Penetration rate in the rural areas was evaluated as 8.8% while in the urban area it was evaluated as 16.9% which shows that a higher population of probe vehicles was present in urban area than in the rural area. However, the average AASE measurement for the rural areas was measured as 5.71 km/hr while in the urban areas it was estimated as 6.43 km/hr which shows that speed estimates in the rural areas were more accurate with lesser variation between the FCD speeds and benchmark speeds.

This study shows that mean speeds derived from commercial FCD are relatively accurate in rural areas and can be used for transportation planning and evaluating speed information in rural areas of South Africa.

This study was conducted on higher order roads and there is a gap on verifying the accuracy of commercial FCD on lower order roads in both rural and urban areas. More studies should also be conducted in rural areas to improve on the accuracy of information provided to stakeholders in order to improve on the transportation sector in the rural areas and consequent standard of living.

REFERENCES

Adu-Gyamfi, Y. O. et al., 2017. Framework for Evaluating the Reliability of Wide-Area Probe Data. *Transportation Research Record: Journal of the Transportation Research Board*, pp. 93-104.

Adu-Gyamfi, Y. et al., 2015. *Reliability of Probe Speed Data for Detecting Congestion Trends.* Missouri, IEEE.

Ahsani, V., Amin-Naseri, M., Knickerbocker, S. & Sharma, A., 2019. Quantitative analysis of probe data characteristics: Coverage, speed bias and congestion detection precision. *Journal of Intelligent Transportation Systems*, 23(2), pp. 103-119.

Altintasi, O., Tuydes-Yaman, H. & Tuncay, K., 2017. Detection of urban traffic patterns from Floating Car Data. *Transportation Research Procedia*, Volume 22, pp. 382-391.

Altintasi, O., Tuydes-Yaman, H. & Tuncay, K., 2019. Quality of floating car data (FCD) as a surrogate measure for urban arterial speed. *Canadian Journal of Civil Engineering*, 46(12), pp. 1187-1198.

Bar-Gera, H., 2007. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Elsevier*, Issue 15, pp. 380-391.

Battelle, 2004. *Traffic Data Quality Measurement*, Washington D.C.: Cambridge Systematics Inc. & Texas Transportation Institute.

Bennett, C. R. et al., 2005. *Data Collection Technologies for Road Management*, Washington D.C.: World Bank East-Asia Transport Unit.

Bruwer, M. M., Walker, I. & Andersen, S. J., 2022. The impact of probe sample bias on the accuracy of commercial floating car data speeds. *Transportation Planning and Technology*, 45(8), pp. 611-628.

Chandra, F. J., Glann, G., Gunawan, F. E. & Gunawan, A. A., 2014. *Improving the Accuracy of Real-Time Traffic Data Gathered by the Floating Car Data Method.* Bandung, Indonesia, Institute of Electrical and Electronic Engineers.

Cohen, S. & Christoforou, Z., 2015. *Travel time estimation between loop detectors and FCD: A compatibility study on the Lille network, France.* Delft, The Netherlands, Elsevier B. V..

COTO, 2012. *TRH 26: South African Road Classification and Access Management Manual*, Pretoria: The South African National Roads Agency Limited.

Dannehy, M. & Krootjes, P., 2013. *How to use TomTom's Traffic Stats portal to access historic traffic data.*. [Online]

Available at: https://www.directionsmag.com/webinar/4559 [Accessed 18 May 2022].

de Fabritiis, C., Ragona, R. & Valenti, G., 2008. *Traffic Estimation And Prediction Based On Real Time Floating Car Data.* Beijing, China, International IEEE.

De Jongh, F. & Bruwer, M., n.d. *Quantification of the natural variation in Traffic flow on selected National Roads in South Africa,* Stellenbosch, South Africa: s.n.

DoT, 2007. *Rural Transport Strategy for South Africa,* Pretoria: s.n.

DoT, 2015. *National Land Transport Strategic Framework- Final Draft,* s.l.: DoT.

EDAT, 2002. *Road Access Guidelines,* Cape Town: Provincial Administration Western Cape.

FHWA, 2021. *U.S. Department of Transport.* [Online] Available at: https://www.fhwa.dot.gov/ohim/handbook/chap6.pdf [Accessed 09 06 2021].

Guillaume, L., 2008. *Road Traffic Data: Collection Methods and Applications,* Luxembourg: JRC European Communities.

Gwara, B. & Andersen, J., 2017. *Validation of Tom Tom historical average speeds on freeway segments in Gauteng, South Africa,* Stellenbosch: Stellenbosch University.

Gwara, B. & Andersen, J., 2018. *Validation of Tom Tom historical average speeds on freeway segments in Gauteng, South Africa.* Stellenbosch, Stellenbosch University.

Haghani, A. et al., 2013. *I-95 Corridor Coalition Vehicle Probe Project Validation of Inrix Data,* Georgia: I-95 Corridor Coalition.

Haghani, A. et al., 2010. Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board,* 2160(1), pp. 60-68.

Javid, M. A. et al., 2022. Analysis of Driver's Socioeconomic Characteristics Relating to Speeding Behavior and Crash Involvement: A Case Study in Lahore. *Infrastructures,* 7(18).

Jia, C. E. et al., 2013. *An Evaluation of Alternative Technologies to Estimate Travel Time on Rural Interstates,* Washington D.C.: Transportation Research Boards.

Jurewicz, C. et al., 2017. *Validation and applicability of floating car speed data for road safety.* Perth, Australia, Australian Road Safety Conference.

Kessler, L., Huber, G., Kesting, A. & Bogenberger, K., 2018. *Comparing Speed Data from Stationary Detectors Against Floating-Car Data.* Germany, IFAC (International Federation of Automatic Control).

Madihlaba, L. G., 2019. *Adoption of Intelligent Transport Systems for Sustainable Transportation in Secondary cities of South Africa: A case of Port Shepstone,* Durban: Durban University of Technology.

MnDOT, 2007. *Minnesota Department of Transport.* [Online] Available at: https://www.dot.state.mn.us/trafficeng/publ/tem/2009/Chapter-05.pdf [Accessed 27 May 2021].

NT, 2006. *National Treasury-Republic of South Africa.* [Online] Available at: www.treasury.gov.za › prov › 08. Chapter 7 - Roads and Transport.pdf

Oruc, A., Hediye, T.-Y. & Kagan, T., 2019. Quality of Floating Car Data (FCD) as a Surrogate Measure for Urban Arterial Speed. *Canadian Journal of Civil Engineering*, 19 April, pp. 1187-1198.

Samuelson, J. P., 2011. *Experience in Data Quality Assessment on Archived Historical Freeway Traffic Data,* Arizona: ProQuest LLC..

Sanwal, K. K. & Walrand, J., 2016. *Vehicle as probes.* California: s.n.

Schneider IV, W. H., Turner, S., Roth, J. & Wikander, J., 2010. *Statistical Validation of Speeds and Travel Times Provided by a Data Service Vendor*, Akron, Ohio: The University of Akron.

Sewell, S., Desai, S., Mutsaa, E. & Lottering, R., 2019. A comparative study of community perceptions regarding the role of roads as a poverty alleviation strategy in rural areas. *Journal of Rural Studies*, Issue 71, pp. 73-84.

Siripollawat, K., 2012. *Evaluation of Travel Time Data Quality and Impacts on the I-95 corridor and in the Hampton Roads Area of Virginia*, Virginia: University of Virginia.

Smith, A. & Visser, A., 2001. *A South African Road Network Classification based on Traffic Loading.* Johannesburg, South Africa, Document Transformation Technologies.

Srinivasan, K. & Jovanis, p., 1996. Determination of number of probe vehicles required for reliable travel time measurement in urban network.. *Transportation Research Record*, Volume 1537, pp. 15-22.

Stander, H. & Brink, J., 2016. *Perspectives on the future of personal transport in South Africa.* s.l., South Africa Transport Conference (SATC).

TES, 2021. *Electronic Traffic Surveying Services.* [Online] Available at: https://www.tes.co.za/trafficmonitoring.html [Accessed 12 10 2022].

TomTom, 2018. *Traffic Stats: Road network analysis and bottleneck intel.* Netherlands: TomTom N.V.

TTI, 2012. *Evaluation of TomTom Historical Average Speed Data,* Texas: Texas Transportation Institute.

Turner, S., 2002. *Defining and Measuring Traffic Data Quality,* Texas: Texas Transportation Institute & Cambridge Systematics Inc..

Turner, S. M., Eisele, W. L., Benz, R. J. & Holdener, D. J., 1998. *Travel Time Data Collection Handbook,* Texas: Texas Transportation Research Institute.

Turner, S., Richardson, J., Fontaine, M. & Smith, B., 2011. *Guidelines for Evaluating the Accuracy of Travel time and Speed data,* Virginia: Texas Transportation Institute, University of Virginia & Virginia Center for Transportation Innovation and Research.

Valadkhani, H. A., Hong, Y. & Ramezani, M., 2017. *Integration of Loop and Probe Data for Traffic State Estimation on Freeway and Signalized Arterial Links.* s.l., International Conference on Intelligent Transportation Systems.

van As, S. & van Niekerk, A., 2004. *THE OPERATIONAL ANALYSIS OF TWO-LANE RURAL HIGHWAYS.* Pretoria, Document Transformation Technologies cc.

Vanderschuren, M., 2006. *Intelligent Transport System for South Africa: Impact Assessment through Microscopic Simulation in the South African context.* Thesis Series no. T2006/4 ed. Delft, The Netherlands: TRAIL Research School.

Wang, Z., 2012. A Compariosn of Floating Car vs. Loop Detector Estimated Freeway Travel Time Delay. *International Journal of Transportation Science and Technology,* 1(2), pp. 147-169.

Woods, D. L., Cronin, B. P. & Hamm, R. A., 1994. *Speed Measurement with Inductance Loop Speed Traps,* Austin, Texas: Texas Department of Transportation.

Yatskiv, I., Grakovski, A. & Yurshevich, E., 2013. *An Overview of Different Methods Available to Observe Traffic Flows Using New Technologies.* Latvia, Transport and Telecommunication Institute.