UNDERSTANDING TRAVEL BEHAVIOUR IN CAPE TOWN THROUGH INVESTIGATING THE APPLICATION OF ALTERNATIVE DATA SOURCES

M COETSEE[1]*, L PRETORIUS[1], C KROGSCHEEPERS[1]*** and S DE BRUIN[1]******

[1]Innovative Transport Solutions (Pty) Ltd, Postnet Suite 127, Private Bag X7,
Tygervalley 7536; Tel: 021 9146211;
*Email: minette@itsglobal.co.za; **Email: lynne@itsglobal.co.za;
***Email: christoff@itsglobal.co.za; ****Email: stefan@itsglobal.co.za

ABSTRACT

The City of Cape Town previously developed and continuously updates a transport model to predict future demand on the transport system in Cape Town and to inform long-term decision-making about the road and public transport network. As part of maintaining and updating the transport model, the City of Cape Town must collect transport data on travel behaviour regularly.

In the past, undertaking household travel surveys was an important source of information on people's travel behaviour in Cape Town. In an environment where people are hesitant to participate in surveys due to health risks associated with the Covid 19 pandemic and security concerns, an alternative approach to collecting the required information was investigated.

With advancements in technology, data sources have become available that have traditionally not been used in understanding people's travel behaviour in Cape Town. This offered an alternative approach to collecting travel behaviour information, but the limitations of the available data source had to be investigated, understood, and tested before it could be used to provide information on travel behaviour.

This paper aims to report on the investigation into the use of the various data sources available to the City of Cape Town, its application in collecting travel behaviour information, and its limitations. The data sources investigated include on-board vehicle surveys (bus and minibus taxis), ticket information, personal transport interviews, probe data and cell phone data.

1. INTRODUCTION

Information on trip-making behaviour is pivotal in informing and shaping effective transport planning strategies. Personal interviews provide valuable insights into travel patterns, individual preferences and behaviour, by systematically gathering information directly from individuals and households. Through household travel surveys, comprehensive data is collected about the movement of individuals in the household and encompasses modes of transport, trip purposes, home and work addresses, times of departure and demographic information. This data is instrumental in understanding the demands and challenges within transportation systems, helping planners and policymakers make informed decisions, guiding infrastructure investments and formulating policies.

Conducting household travel surveys poses several challenges, notably for security, health concerns, and changing living arrangements. Security concerns can hinder the willingness of individuals to interact with interviewers and to participate in surveys. Traditionally, surveyors were mostly invited into homes where the lengthy surveys were explained and completed for each household member. Additionally, health considerations, as highlighted by events such as the Covid-19 pandemic, have added a layer of complexity to conducting in-person surveys. Concerns about virus transmission may deter respondents from allowing surveyors into their homes or participating in face-to-face interviews. Furthermore, the rise of gated communities and security complexes presents another challenge, as access restrictions limit the ability of interviewers to reach residents and collect representative data.

Conducting household travel surveys is also expensive, and includes team recruitment, training, transport, and materials. The costs are exacerbated when a deployed survey team achieves a low response rate and the survey period is extended to achieve a reasonable representative sample.

The use of alternative data sources into transport modelling for planning purposes represents a significant shift from traditional household surveys, offering transformative opportunities for authorities and transport planners to use existing, readily available data sources as a more efficient and cost-effective alternative to inform transport models.

These alternative data sources contribute to overcoming the limitations of traditional household travel surveys that may offer richer and more dynamic datasets. However, it is important to note that the adoption and effectiveness of these sources can vary based on sample sizes, biases, level of representation and accuracy.

In the City of Cape Town study (2023), the following alternative data sources were investigated as a means to provide the necessary data on people’s origins and destinations to update the transport model:

- Probe data.
- On-board surveys (on minibus taxis and buses).
- Ticket information (MyCiTi bus rapid transit (BRT)).
- The National Household Travel Study (NHTS).
- Cordon Surveys.

1.1 Aim of Paper

This paper aims to provide insight into the effectiveness of utilising alternative data sources for transport planning and to demonstrate how these sources have been applied to gather information on travel behaviour. It aims to highlight instances where the data sources have proven successful in enriching the understanding of travel patterns and behaviours among the city's residents. Additionally, the paper intends to address the limitations encountered during the utilisation of these data sources, acknowledging challenges or constraints that may impact the comprehensiveness or accuracy of the collected travel behaviour information. By offering an examination of both successes and limitations, the paper aims to contribute to an increased understanding of the practical implications and insights derived from employing varied data sources.

1.1.1 Problem Statement

The City of Cape Town is facing challenges in gathering travel behaviour data, which is required for developing and updating its transport models. Traditional methods like household interviews are becoming less effective because people are less accessible due to safety, security, health concerns, and physical limitations. To address these issues, there is a need to find alternative approaches to collect the required data for planning models. The emergence of technology and big data opens new possibilities by introducing alternative data sources not traditionally used in transport planning. The problem is that these alternative data sources need to be fully comprehended to determine their usability. Factors like biases, accuracy, and representativeness are introduced with alternative data sources that were not relevant to traditional travel surveys.

1.1.2 Scope of Paper

This paper explores the efficiency and limitations of alternative data sources in informing transport modelling and planning strategies within the context of the City of Cape Town. The scope encompasses an examination of diverse alternative data sources, including on-board surveys, probe data, ticket information, the National Household Travel Study, and cordon surveys.

The investigation aims to:

Evaluate Alternative Data Sources: The paper scrutinises various alternative data sources, assessing their ability to capture comprehensive travel behaviour information. This includes a detailed examination of probe data, on-board surveys for different modes of transport, ticket information, and cordon surveys.

Assess Reliability and Representativeness: Statistical analyses are employed to validate the reliability and representativeness of alternative data sources. The focus is on understanding the correlation between these alternative sources and personal interviews.

Examine Origin-Destination Patterns: The paper delves into the useful extraction of Origin-Destination (O-D) patterns from different alternative data sources. It scrutinises the suitability of these patterns for informing the City of Cape Town’s transport model, emphasising successful sources and identifying challenges.

Consider Impacts of External Factors: The paper acknowledges and evaluates the impact of external factors such as the Covid-19 pandemic and declining rail services on travel behaviour, emphasising their influence on O-D and modal split patterns.

Provide Recommendations for Future Planning: Based on the findings, the paper offers insights and recommendations for the City of Cape Town's future transport planning strategies. This includes considerations for adaptive planning in response to dynamic societal and technological trends.

By delving into the successes and challenges of alternative data sources, this paper aims to contribute valuable insights to the discourse on urban travel behaviour studies and advance the understanding of the practical implications of integrating diverse data sources into transport modelling for effective and resilient planning.

2. LITERATURE REVIEW OF THE APPLICATION OF ALTERNATIVE DATA SOURCES IN TRANSPORT MODELLING

The collection of alternative data sources used internationally by various cities and institutions is comprehensive and ever-changing due to the fast development of technology and related applications. The subsequent listing of data sources and technologies captures the prevalent international trends in the use of alternative data for transport planning.

Global Positioning System (GPS) Data: GPS technology, obtained from smartphones, in-vehicle systems, or dedicated GPS devices, provides real-time and accurate information about travel routes and speeds, enhancing the accuracy of travel data (Koutsopoulos, 2003).

Electronic Ticket/Fare Collection Card Data: Public transportation systems worldwide leverage electronic cards to collect data on users' travel patterns, helping to understand modal choices, frequency, and entry/exit points (Pelletier, Trepanier & Morency, 2011).

Mobile Applications: The rise of mobile apps dedicated to travel and navigation contributes to data collection, offering insights into user preferences, route choices, and real-time travel behaviour (Zheng, Capra, Wolfson & Yang, 2010).

Social Media Analytics: Mining geospatial information from social media platforms provides valuable data on travel preferences, sentiments, and emerging trends (Hawelka, Sitko, Beinat, Sobolevsky, Kazakopoulos & Ratti, 2014).

Big Data Analytics: Using large datasets from diverse sources, including mobile networks, sensors, and open data, enables a holistic understanding of travel patterns and behaviour (Calabrese, Colonna, Lovisolo, Parata & Ratti, 2011).

3. METHODOLOGY

In the project undertaken for the City of Cape Town, a set of available alternative data sources were identified. These sources included probe data, on-board surveys, ticket information and cordon surveys, each offering unique insights into various aspects of travel behaviour and patterns within the City.

Alongside these alternative data sources, the City of Cape Town study involved the completion of personal interviews using intercept surveys and online survey methods. These personal interviews were designed to collect travel and demographic information from working individuals and individuals making school trips. The intercept surveys were aimed to extend geographically over the entire Cape Town Metro and were designed to represent the travel behaviour of the working population of Cape Town. A sample size of 1.94% of the working population was achieved.

The following data sources were identified that contain travel information of individuals and could be used as an alternative source to the traditional household travel survey. The extent of the data source is described briefly:

- Probe data (Source 1): 3 separate months in 2020 and 2022, across Cape Town.
- Probe data (Source 2): 5 days in 2021, across Cape Town.
- Golden Arrow Bus (GABS) on-board surveys: all trips across Cape Town in 2016.
- Minibus Taxi (MBT) on-board surveys: 3 trips per route for all routes across Cape Town during 2017 and 2018.
- Anonymised MyCiTi on and off-electronic taps: 1 month in 2019, across the MyCiTi network.
- Anonymised cellphone data: 2 consecutive months in 2020 and 2022, across Cape Town.

- Cordon counts around CBDs, some industrial areas, low-, middle- and high-income residential areas (2016, 2017, 2018 and 2021).

The methodology employed for the City of Cape Town study was to compare results derived from the alternative data sources, with the results obtained through the personal interviews. This comparative analysis aimed to identify correlations and discrepancies between the diverse datasets, shedding light on the reliability and consistency of the information contained in the alternative data sources, and whether it is suitable for use in transport models.

The assessment of these independent data sources aimed to identify the available data and its suitability for informing origin-destination (O-D) patterns across various transport modes. However, it is essential to acknowledge and understand the limitations and restrictions associated with these independent data sources, which include the following:

Sampling Bias: Several alternative data sources represent samples of a larger population. For instance, probe data reflects individuals with vehicles equipped with tracking devices, potentially introducing biases toward specific income groups and residential areas.

Public Transport Data Constraints: Public transport data sources, such as MyCiTi, GABS, and MBT, do not capture the entirety of a public transport passenger's trip. These datasets focus on the vehicle trip, omitting details about modal transfers or same-mode transfers.

A thorough examination, including rigorous analyses and tests, was conducted to understand these limitations and biases inherent in the alternative data sources. Additionally, O-D patterns from the alternative data sources were compared with those from personal interviews. Most alternative data sources are mode-specific except for the cellphone data that represents O-D information irrespective of the mode. The objective of these comparisons was to identify correlations between data obtained from personal interviews and alternative data sources.

Instances, where strong correlations were identified between personal interview data and alternative data sources, demonstrated the viability of using the latter to obtain O-D information for specific modes. Conversely, weak correlations indicated that the alternative data source, in its current state, is not suitable, and data from personal interviews should be prioritised. This methodology is summarised in Figure 1 below.

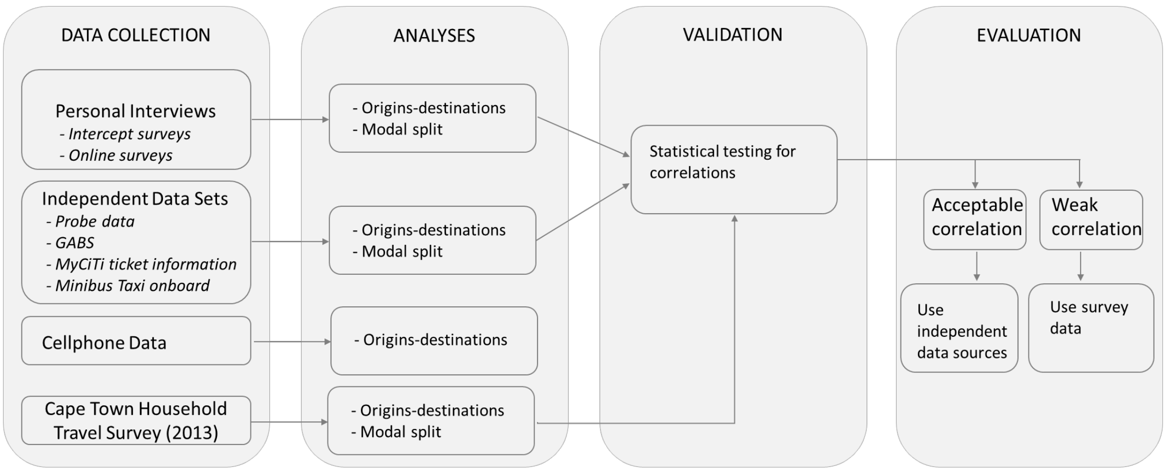

Figure 1: Methodology flow diagram

3.1 Data Collection and Analysis

The alternative data sources were gathered and analysed. The product of the analysis was an O-D matrix for each dataset, being either representative of a specific mode or representing all modes of transport. The matrices represent the number of commuter and/or school trips to and from the 48 Macro Zones (MZs) in Cape Town. This section will delve into an examination of these datasets, whether a representative O-D matrix could be developed and shedding light on the respective strengths and weaknesses of each alternative data source.

3.1.1 Probe Data, Source 1

In evaluating travel patterns, access to reliable and spatially analysable data is paramount. Probe data, collected by GPS-enabled on-board units for insurance and anti-theft purposes, is such a data source. It provides coordinates and timestamps of vehicles en-route and enables the spatial analysis of travel patterns. While anonymised by service providers, the data includes information on vehicle makes and models, enabling the inference of vehicle types such as light vehicles, buses, and motorcycles.

In the City of Cape Town study, travel patterns for three types of trips were extracted from the probe data. These trips are commuter trips, all trips and Minibus Taxi (MBT) trips. For commuter trips, a set of criteria was developed and applied to the data identified individuals frequently travelling between residential and business zones, revealing specific patterns for the a.m. and p.m. periods. Despite a lower sample of p.m. commuters, strong similarities in Origin-Destination (O-D) patterns were observed, suggesting the method's acceptability.

Comparisons extended to all private vehicle trips, revealing similarities in O-D patterns with commuter trips, validating the approach's effectiveness. A statistical correlation test affirmed the correlation between the O-D results of commuters and all private vehicle trips where the dataset for all private vehicle trips was much larger. It was concluded that the strict criteria for identifying commuters were not yielding different results than from using all private vehicle trips and that it was more beneficial to use the larger dataset for all private vehicle trips for O-D analyses.

Additionally, the study investigated MBT vehicles, initially aiming to derive O-D patterns but later discovering biases in the data. The analysis of these vehicles revealed that even with a substantial sample size, the data is biased towards specific Taxi Associations that have permanent tracking devices installed. This was confirmed by comparing the probe data with the number of MBT operating licenses on known MBT routes. In some cases, there were noticeable differences between routes that are supposed to carry high MBT traffic according to the operating licenses, but that show very little activity for MBT vehicles in the probe data. The conclusion reached was that the O-D patterns of MBT probe vehicles would not be a true representation of real MBT movement across Cape Town.

In summary, the City of Cape Town study showcased the usefulness of probe data in extracting travel patterns, especially for commuter and private vehicle trips. However, limitations in representing MBT movements underscore the need for cautious interpretation and consideration of biases in the data.

3.1.2 Probe Data, Source 2

The assessment extended to the O-D matrix derived from probe data (Source 2) data. While the precise structure of the trip data is undisclosed, it is presumed to share a similar format with the probe data from Source 1.

The O-D matrix comprises data accumulated over five weekdays in October 2021, predating the implementation of Protection of Personal Information Act (POPIA) restrictions. Representative of all trips and vehicles with journey start times between 06:30 a.m. and 08:00 a.m., the matrix includes multiple O-D pairs for the same vehicle during the same peak period. This arises from intermediary stops not being eliminated before calculating the O-D matrix. The same vehicle would also be present multiple times in the dataset because of adding multiple trips for the same vehicle over five days. The number of trips is therefore cumulative in the matrix.

A statistical correlation test was conducted between this dataset and the personal interview result, indicating a strong correlation.

3.1.3 Minibus Taxi On-Board Surveys

The MBT on-board surveys that were conducted in 2018 provide valuable insights into MBT trips as well as boarding and alighting events. The MBT surveys are useful to determine the exact O-D information of individual passengers since the boarding and alighting locations of individual commuters are recorded in the survey. The City-wide sample size in the peak periods is 194 891 daily passengers.

Most MBT routes in Cape Town were surveyed, but only a sample was surveyed on routes. Furthermore, the sample is not necessarily statistically significant as only 1-3 trips per route were undertaken. This implies that although it is known where MBTs drive and where they stop to pick up and drop off passengers, the actual number of passengers on a route cannot be established from this dataset.

MBT rank surveys have been conducted, but they are not yet comprehensive enough to calibrate the O-D numbers derived from the on-board survey sample. It was also attempted to calibrate the on-board survey data with the MBT vehicles in the probe data, but this was unsuccessful due to the biased nature of the MBT probes, as mentioned in Section 3.1.1.

Modal transfers may also occur at Public Transport Interchanges (PTIs) which is not reflected in the data. The data implies that the destination of the passenger is the PTI, whereas the passenger journey typically will continue. O-D matrices could therefore not be extracted from the Minibus Taxi on-board surveys.

3.1.4 Golden Arrow Bus Service On-Board Survey

The on-board surveys conducted for the Golden Arrow Bus Service (GABS) offer a more extensive dataset, encompassing all GABS bus trips. Completed between 2016 and 2017, these surveys involved surveyors boarding all buses operating on every GABS route. However, while the surveys captured the overall bus trips, specific details on the boarding and alighting of individual commuters were not recorded. Consequently, extracting the O-D information for individual bus passengers from the database is not possible. To mitigate this, the challenge was simplified by assuming that all passengers boarding, originate from the same Macro Zone (MZ) as the bus origin. This simplification enabled some O-D data extraction, facilitating the calculation of the percentage distribution of passengers departing from one MZ and arriving at another within the City.

It is important to note that the data does not account for modal transfers at PTIs, impacting the accuracy of passenger destination assumptions, which are assumed to be the PTI. In reality, the passenger journey extends beyond this point.

The sample size for GABS passengers during peak periods is 198,481, representing the total population captured in the dataset.

3.1.5 MyCiTi

The MyCiTi bus service proves to be a valuable information source for analysing the O-D patterns of its passengers. This stems from the system's inherent capability to automatically track specific passengers throughout their journey, from boarding to alighting. Upon boarding, each passenger's unique card is scanned and registered in the system with a station name and timestamp. The same process is repeated when the passenger alights. Knowing the geographical locations of all stations, these can be assigned to Macro Zones (MZs) within Cape Town, enabling the determination of O-D patterns for MyCiTi passengers.

For analysis purposes, a dataset from February 2019 was obtained, capturing the period when the N2 Express routes were still operational. The City-wide sample size during peak periods comprises 34 571 passengers on a typical weekday, serving as a representative snapshot of the total MyCiTi passenger population.

3.1.6 Cellphone Data

Cell phone data serves as a valuable source of O-D information, registering a person's position each time their phone connects to a new cell tower. This data, stored by service providers with timestamps, allows for the establishment of commuter patterns based on assumptions about a person's likely location at different times of the day.

To extract O-D patterns, a methodology was developed in collaboration with the service provider, involving data extraction for specific periods and zone assignment based on the service provider's base stations. Sensitivity to base station locations led to the redistribution of data at the suburb level and the derivation of O-D matrices per unique weekday.

Considering privacy concerns, the service provider refrained from sharing statistics for zones with fewer than 10 unique users on any given day. To address this, neighbouring zones were merged until the minimum threshold of 10 unique users was reached. The outcome was an O-D matrix of work-to-home trips per suburb for individual weekdays, averaged across all weekdays, and assigned to macro zones to derive an average daily device count O-D matrix.

3.1.7 National Household Travel Study

The third National Household Travel Survey (NHTS) conducted in 2020 aimed to understand work-related travel patterns in the Cape Town metropolitan region. However, several concerns were identified during the assessment of the dataset:

- The dataset for Cape Town included responses from only 0.10% of the working-age population, with information on work-related travel available for a small portion of individuals who travel to work or an educational facility.
- The NHTS sampling frame, based on 18 Transport Analysis Zones (TAZs) in Cape Town, lacks homogeneity in terms of land use and demographics. The small sample size per TAZ limits its representativeness, especially for the 48 macro zones defined in the transport model.
- The coarse zoning of TAZs hinders detailed analysis of people's mobility patterns, and the overall small sample size restricts TAZ-level analysis.

- City-wide analysis is possible to some extent, but modal split analyses may be misleading due to the exclusion of Bus Rapid Transit (BRT) as a main mode option in the NHTS questionnaire.
- High rates of non-responsiveness in certain sections of the NHTS limit the overall usability of the data.

Given these limitations, further analysis of the NHTS for Cape Town was deemed impractical.

3.1.8 Cordon Counts

Traffic volumes have been consistently recorded at various cordon locations in Cape Town in recent years, with eighteen out of the nineteen cordons counted in 2021. These counts detail traffic volumes categorised by vehicle type, direction, and in 15-minute intervals, encompassing both vehicle and occupancy/passenger data. The cordon dataset enables the determination of modal splits at the survey locations.

It is important to note that the spatial definitions of most cordons and MZs differ. Typically, cordons are confined to specific portions within an MZ. Due to these distinctions, caution should be exercised when comparing modal split data for several reasons:

- **Aggregation Levels:** The personal interviews are aggregated at the MZ level, representing large spatial areas and diverse land uses, while cordons are more focused and smaller in scale.
- **Categorisation Discrepancies:** Cordon counts are categorised by vehicle types, while survey data is categorised by mode. This creates discrepancies, for example, where 'Light Vehicles' in cordon counts include more vehicle types than 'Private Car' in survey data.
- **Difficulty in Vehicle Differentiation:** Some challenges arise in distinguishing between vehicles like 'bakkies, LDVs, and small trucks,' making it difficult to determine whether they are private use vehicles (private cars) or LDVs/commercial vehicles. These are collectively classified as 'Light vehicles'.
- **Inclusion of Non-Commuter Vehicles:** Cordon counts may encompass a significant portion of non-commuter vehicles, unlike the modes referred to in personal interviews, which are specific to commuter vehicles.
- **Influence of Through Traffic:** Traffic volumes at certain cordon locations can be influenced by traffic, impacting the specificity of the modal split to the cordoned area.

3.1.9 Personal Interviews

Personal interviews in the form of interception surveys and self-interviews through an online questionnaire were completed. The interception surveys were undertaken by a survey team of 38 surveyors over the course of 57 days. 28 000 personal interviews were completed, resulting in a sample size of 1.94% of the working population in Cape Town. The intercept surveys were conducted at locations of congregation that are not biased to a certain mode of transport, like retail centres as people from all income groups are expected to congregate around these areas and would not be biased towards a mode of transport. The locations of the intercept surveys were monitored and adjusted throughout the survey period to obtain a good spatial representation across Cape Town and also to get statistical significance in certain areas.

3.2 Validation

To validate the various datasets against personal interviews and understand the correlation between them, two types of statistical analyses were conducted. The first set of analyses compared the personal interviews (referred to as "survey count") with the "O-D count" of external data sources (GABS on-board survey, MyCiTi on-board survey, and the probe database) on a mode-specific basis. The second set of analyses included all modes and tested the O-D correlations between different sources.

The evaluation involved assessing Pearson and Spearman R-values, which are correlation coefficients indicating the strength of the relationship between two sets of data. The results showed correlations ranging between 0.44 and 0.79, indicating positive correlations. Two examples of correlation plots are shown in Figure 2 below and all R-values are summarised in Table 1.

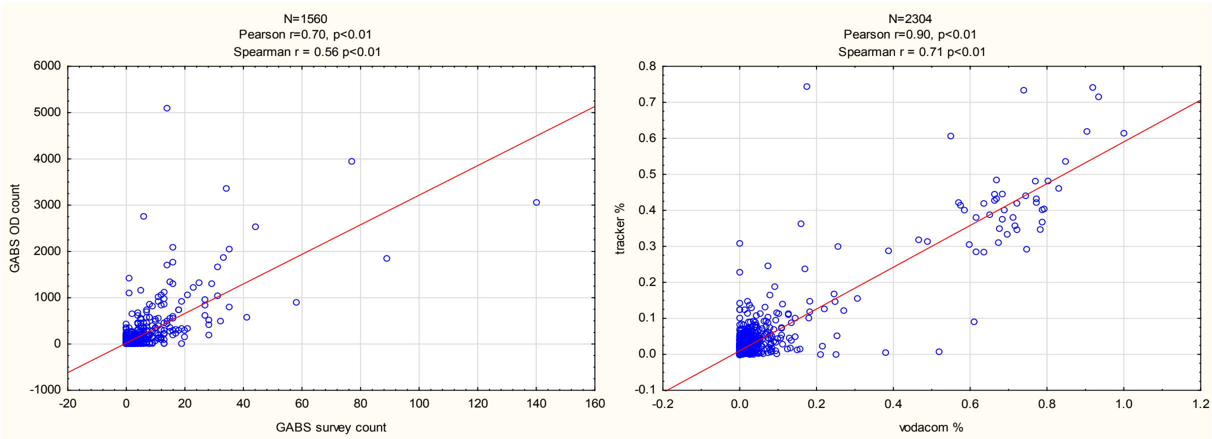

Figure 2: Examples of R-value plots indicating correlations between data sets

Table 1: Summary of correlations (R-values) between different datasets

Mode Specific Comparisons		Unique OD pairs	Pearson R-value	Spearman R-value
GABS from Personal Interviews	GABS	1560	0.70	0.56
MyCiTi from Personal Interviews	MyCiTi	210	0.52	0.56
Private from Personal Interviews	Private	2025	0.72	0.69
Comparison including all modes				
Vodacom	Probe: Tracker	2304	0.90	0.71
Vodacom	Probe: TomTom	2304	0.91	0.64
Probe: Tracker	Probe: TomTom	2304	0.93	0.74
Personal Interviews	Vodacom	1968	0.76	0.73
Personal Interviews	Probe: Tracker	1968	0.78	0.77
Personal Interviews	Probe: TomTom	1968	0.72	0.67

The statistical evaluation concluded that "In light of the fact that the external data (which is available continuously over time) would be used for long-term planning, the statistical results showed enough correspondence to view the external data also as representative of the population." This underscores the reliability of external data sources for long-term planning purposes.

3.3 Evaluation

3.3.1 Suitability of the Approach

In conducting the Cape Town Travel Study, we concurrently examined and analysed alternative data sources alongside personal interviews, resulting in positive correlations through statistical testing. While acknowledging some limitations, the adopted approach was deemed suitable for the City of Cape Town study's objective to extract O-D information from data sources, other than the traditional household travel surveys.

The study included alternative data sources from different periods with the earliest data set in 2016 (GABS) and the most recent data sets in 2022 (probe data and cellphone data). The personal interviews were completed at the end of 2021 and the beginning of 2022. Although the study recognises the temporal disparity, the analysis of travel behaviour aims to understand patterns and trends, and not absolute values. It is expected that the correlations between alternative data sets and the personal interviews could have been stronger should the data be from the same period, and even with this disadvantage of temporal disparity, the correlations proved to be satisfactory.

3.3.2 Origin-Destination and Modal Split Patterns

The investigation revealed that O-D patterns can be extracted from the GABS on-board survey, two probe datasets, MyCiTi data, and cellphone data, with notable correlations. However, the MBT on-board survey does not align with passenger O-Ds and is not suitable.

3.3.3 Demographic Data

The personal interviews provided comprehensive and reliable demographic information. However, concerns were raised about the sample size and spatial granularity of the National Household Travel Survey, limiting its overall usability. The NHTS was therefore not used for comparison with the demographic data acquired through the personal interviews.

3.3.4 Transfer Patterns

The data collected from personal interviews regarding multi-modal trips and transfer patterns were deemed unsatisfactory. GABS on-board surveys, MBT on-board surveys, and MyCiTi data lack data on modal transfers and no transfer patterns could be derived.

3.3.5 Limitations of Data Sets

While probe data for private vehicles and MyCiTi data proved valuable for O-D patterns, the MBT probe data displayed biases related to specific Taxi Associations’ vehicles having a higher probability of having an on-board GPS. GABS on-board data necessitates assumptions on the origin and destination of passengers, impacting O-D accuracy. In the cellphone dataset, challenges were noted about spatial sample size per zone and redistribution.

3.3.6 Impact of POPIA

Privacy concerns were addressed, with none of the data sources including identifiable personal information. The cellphone data provider ensures privacy, making anonymised data available only for samples greater than 10.

4. CONCLUSION

In response to the evolving challenges associated with traditional household travel surveys and the task of maintaining an accurate transport model for informed decision-making, the City of Cape Town embarked on a comprehensive exploration of alternative data sources to inform transport models. Faced with obstacles such as health concerns, security issues, and financial constraints related to survey methodologies, the study delved into the effectiveness of diverse alternative data sources in capturing travel behaviour information.

Advancements in technology opened avenues to data sources that were traditionally absent, presenting both opportunities and challenges. This paper examined the successes and limitations of various data sources, including probe data, on-board surveys, ticket information, cellphone data, the National Household Travel Study, and cordon surveys. Each data source underwent a detailed evaluation, shedding light on its applicability in informing Cape Town's transport models.

The investigation commenced with a literature review highlighting global trends in utilising alternative data sources, emphasising the potential of integrating existing, readily available data into transport modelling. The subsequent methodology section explained the approach taken, which involved personal interviews, using online and intercept methods and the comparison of results with alternative data sources. Statistical analyses validated the external data sources, affirming their representativeness for long-term planning.

The study's findings underscored the adopted approach's suitability, revealing positive correlations between personal interviews and alternative data sources. O-D patterns could be successfully extracted from several alternative sources, including GABS on-board surveys, probe data, MyCiTi, and cellphone data. However, challenges surfaced in deriving transfer patterns.

The positive statistical correlations between datasets, with R-values ranging between 0.44 and 0.79, proved alternative data sources to be satisfactory alternatives to traditional household surveys in obtaining travel behaviour data for transport planning.

Despite the limitations and challenges encountered, the study positioned alternative data sources as valuable contributors to transport planning for the City of Cape Town. From the success of MyCiTi's automatic tracking system to the insights gathered from on-board surveys and probe data, the City now comprehend the use of alternative sources to understand commuter behaviour. The impact of the Covid-19 pandemic on travel behaviour was acknowledged, emphasising the need for adaptive planning.

In conclusion, this paper contributes to the conversation on urban travel behaviour studies and transport planning by showcasing the potential and challenges of using alternative data sources.

5. REFERENCES

Calabrese, F, Colonna, M, Lovisolo, P, Parata, D & Ratti, C. 2011. Real-time Urban Monitoring Using Cell Phones: A Case Study in Rome. *IEEE Transactions on Intelligent Transportation Systems,* 12(1):141-151.

City of Cape Town, City-wide Travel Study, Draft Report 3, March 2023.

Hawelka, B, Sitko, I, Beinat, E, Sobolevsky, S, Kazakopoulos, P & Ratti, C. 2014. Geo-located Twitter as Proxy for Global Mobility Patterns. *Cartography and Geographic Information Science*, 41(3):260-271.

Koutsopoulos, HN. 2003. The Use of Global Positioning Systems to Monitor Travel Patterns and Elicit Activity-Travel Behaviour Dynamics. *Transportation Research Part C: Emerging Technologies*, 11(3-4):241-263.

National Household Travel Survey. Available at: www.statssa.gov.za/publications/P0320/P03202020.

Pelletier, MP, Trepanier, M & Morency, C. 2011. Smart Card Data Use in Public Transit: A Literature Review (2011).

Zheng, Y, Capra, L, Wolfson, O & Yang, H. 2010. Urban Computing: Concepts, Methodologies, and Applications. *ACM Transactions on Intelligent Systems and Technology*, 1(1):1-27.