

# **Monitoring Transit-Served Areas with Smartcard Data: A Brisbane Case Study**

## **Abstract**

A city can be divided into areas that are served by transit and those that are not. In this study, the former is referred to as “transit-served areas (TSAs)”. To quantify, monitor and visualise the TSAs of Southeast Queensland (SEQ), this study analyses half-year smartcard data between 2012 and 2013 from TransLink, the transit agency for SEQ. For scenarios are prescribed and four corresponding metrics (the minimum, actual, random and maximum travels) are calculated, which reflect transit riders’ different levels of elasticity of distance travelled (EDT) relative to the cost of travel within or between TSAs and how transit riders could possibly travel as EDT varies. The total trips generated by or attracted to TSA and the temporal and spatial variations of these metrics across days are used to monitor TSAs, especially transit trips within or between them. The results indicate that transit trips attracted to, and generated by TSA and transit trips between TSAs vary significantly over time and across space. Across the scenarios, the temporal variance tends to be larger as EDT becomes more inelastic. The above results provide useful references for decision-makers to understand better the ranges of transit demand (by TSA) across the space and time when EDT is a variable.

# 1. Introduction

A city can be divided into different subareas that are served by transit (used interchangeably with public transport hereafter) and those that are not. In this study, the former is referred to as “transit-served areas (TSA)”. TSA is not a new term that we first coined. In existing studies such as Calthorpe (1993), Cervero et al. (2002) and TCRP (2004), similar terms have been mentioned or defined. Calthorpe (1993), for instance, conceptualizes transit-oriented development as a transit-served neotraditional community. Cervero et al. (2002, p.75) regard Calthorpe (1993) as something “provocative”. Both Cervero et al. (2002) and Calthorpe (1993) favour TODs that are co-dependent and that are linked to each other by high-capacity fixed-guideway transit services. TCRP (2004, p.S10, S13, 139, 177, 322, 349) frequently uses terms such as “transit-served neighborhoods”, “transit-served subcity nodes” and “transit-served nodes and corridors”.

Traditional data sources such as maps, satellite images, censuses, household surveys and phone interviews have previously dominated decision-making processes related to TSAs. But collecting traditional data are costly and those data are generally updated infrequently. Most census bureaus, for instance, update their household survey data every five or ten years -- a relatively long period of time when many short-term decisions or policies have to be made. Smartcard data collected by transit agencies can generate timely information about transit riders over most time horizons. Much of this information is directly related to public transport planning, operation and modelling, for instance, where transit riders go, when and how often.

In theory, we can “paint” more dynamic and continuous pictures of TSA with smartcard data. We can visualise the distribution of transit riders by: corridor, minute, hour, day, month or season of a year. And with extra data, we can produce more than just those pictures. We can, for instance, interrogate various land-use and public transport plans and cross-reference it with smartcard data and other big data (e.g. cellular network data), to be able to identify where people reside, socialize, entertain and work. With the multimodal transport network files, we can map out the probable redistribution of transit riders when major bus stations undergo repairs. All of this information could help us make more informed short-term decisions about transit and urban systems, particularly on the extent to which we can be more proactive in terms of responding to disruptions such as extreme weather, natural disasters, unforeseen hazards or even terrorist threats, and on the extent to which we can better allocate resources or manage passenger traffic. This represents a significant paradigm shift of public transport planning and operations (c.f., Batty, 2013).

This research uses six months of anonymous smartcard data in one of the Australia’s capital cities collected between 2012 and 2013. The data are from TransLink, a division in Queensland’s Department of Transport and Main Roads, an entity tasked to deliver high-quality public transportation to quantify, monitor and visualise the TSAs of South East Queensland (SEQ). Four scenarios are designed and corresponding metrics for those scenarios across 172 days are prescribed and calculated: the actual, minimum, maximum and random travels. The temporal variations of these metrics across days are used to monitor different TSAs, especially the extent to which they are connected to one another, where the connections are considered as passenger flows between or within TSAs. The visual representations of these flows are also used to monitor and gain extra insights into TSAs, especially how trips attracted to or generated by TSA vary across the space, days and scenarios where transit riders’ elasticity of distance travelled (EDT) relative to the cost of travel are no longer a constant and varies across time and space.

## 2. Review of relevant literature

Traditionally, most existing studies have drawn data from traditional methods such as censuses and travel surveys to investigate travel behaviour and related phenomena such as suburbanization, jobs/housing separation and social exclusion (e.g., Zhou, 2014; Zhou et al., 2012; Li et al., 2012; Salas-Olmedo and Nogues, 2012; Engels and Liu, 2011; Healy and O’Connor, 2001). An emerging

number of scholarly work, however, illustrates the potential of smartcard data as an alternative source of travel behaviour information (e.g., Tao et al., 2014a, b; Zhou, et al., 2014a, b; Zhou and Long, 2016). Smartcard data can be used to derive the origin-destination matrix, which is a vital input to public transport planning and modelling (Alsger et al., 2015). They can be used to identify activity purpose, duration, route choice, segmentation, volume and location of transit riders (Tao et al., 2014b; Ma et al., 2012; Janosikova et al., 2014; Kim et al., 2014; He and Trepanier, 2015; Kieu et al., 2015; van der Hurk et al., 2015; van Oort et al., 2015; Tamblay et al., 2016). They can also be employed to evaluate the performance of metropolitan passenger railways with respect to on-schedule rates, occupancy and operational speeds (Eom et al., 2015).

Of those existing studies that are based on smartcard data, an increased number have begun to take advantage of the characteristics of smartcards such as a bigger sampling population, finer temporal and spatial granularity and continuous and longer duration of observations as compared to traditional data. This has enabled a more accurate data representation and interpretations to inform planning and policy making, for instance:

- Reconstructing and visualising a larger sample of transit riders' trajectories and associated corridors on a day (Tao et al., 2014b);
- Reproducing passenger flows along transit routes and developing models that can quantify the passenger volumes under different what-if scenarios (e.g., fare increases) (van Oort et al., 2015);
- Examining regularities and variations in trip generation by station by different periods of time across cities (Zhong et al., 2016);
- Quantifying the number of trips and their spatial distribution by day and comparing related patterns (Zhong et al., 2015);
- Deriving workplaces and residences of a large sample of transit riders and comparing their actual commute with some baselines (Zhou et al., 2014a, b; Zhou and Long, 2014).

Of these studies, however, few have looked at the dynamics of passenger trip generation, attraction and distribution by day and over months so as to examine the impacts of seasonality, to visualise what the extremes look like (e.g., when and how many maximum trips there are in a year and where these trips are within or across TSAs and how extensive they are and where they are) and how the extreme relates to the average or some comparable baseline. One exception is Morency et al. (2007). But unlike our SEQ data which capture both origin and destination information, their data only contain trip origin information and trip destination information had to be derived. Given the extra characteristics of smartcard data mentioned above, the studies of the longer-term dynamics of trip generation, attraction and distribution have now become feasible with the SEQ data. This study is an attempt to show how to translate feasibility into reality. It also shows that such dynamics can be quantified and visualized based on some simple metrics, which can be linked to different scenarios, for instance, transit riders' EDT relative to the cost of travel are no longer a constant and varies across time and space.

### **3. Methodologies**

In this study, we stipulate four scenarios and use four metrics to study TSAs. Our scenarios and metrics got inspirations from several seminal/existing excess-commuting studies, in particular, White (1988), Horner (2002), Charron (2007) and Murphy and Killen (2011). Specifically, our four scenarios designed to examine transit riders' EDT relative to the cost of travel within or between TSAs over time are:

- # 1: base, which is the status quo;
- # 2: transit riders are extremely elastic to the cost of travel;
- # 3: transit riders' EDT is randomly distributed;
- # 4: transit riders are extremely inelastic to the cost of travel.

We feel that those four scenarios are of interest to both scholars and decision-makers, as the scenarios show how transit riders could possibly travel as their EDT changes. Like other scholars using the excess-commuting framework before, we assume that the total number of transit trips for each day is fixed despite that EDT has changed across the scenarios. Unlike other scholars using the framework before who often quantify one or several days' metrics, we quantify four metrics: the minimum ( $T_{\min}$ ), actual ( $T_{\text{act}}$ ), random ( $T_{\text{rand}}$ ) and maximum ( $T_{\max}$ ) travels across 172 consecutive days and examine both temporal variations of these four metrics and spatial variations of the corresponding trip patterns. Given our much large numbers of observations or metrics, we believe that we should be able to see much salient variations in transit behaviours. Understanding these variations informs us about how to plan and operate our transit services in some extreme conditions, for instance, how transit riders would travel when the weather is extremely hot and few would like to travel by transit, that is, transit riders' EDT is extremely elastic.

On each of the 172 days, we calculate four metrics to facilitate our studies/comparisons of the scenarios. Each of the four metrics reflects how transit riders respond to the cost of travel within or between TSAs.  $T_{\min}$ , for instance, shows that if transit riders are extremely elastic to the cost (Scenario #2) and how small the total distance travelled for all transit riders can be. In this study, the temporal variations of these metrics across days are also used to monitor trips within and between TSAs, that is, connections between TSAs. We want to find the extreme of  $T_{\min}$  over time, for instance, what would be the smallest/largest total distance travel for Scenarios #2 of all the 172 days.

Ideally, we should define TSAs considering spatial variations and transit riders' preference. But to define TSAs that accurately would require an excessive amount of input data, for instance, local street network, safety, lighting, and transit riders' acceptable distance to a transit stop. In this study, we simply use Statistical Area Level 2 area (SA2) as a convenient proxy of TSA. We define TSA as a SA2 that has at least one transit stop within its boundaries. SA2 are a general-purpose medium-size area designated by Australian Bureau of Statistics (ABS). The top three criteria that ABS considers when delineating the SA2's boundaries are: population, functional and growth. A SA2 generally has a population of 3,000 to 25,000. Within the central urbanized area, a SA2 is usually 1-2 square kilometres in size and thus a resident therein can easily walk to a transit stop, if any. In terms of functionality, each SA2 usually has a centre that provides most services that residents need daily. Thus, in theory, most transit riders do not have to travel outside an SA2 to have their basic needs met. Regarding growth, SA2s contains regional towns or fringe areas around large cities that governments want to contain: the urban area, any immediately associated semi urban development and likely growth in the next 10 to 20 years so that the boundaries of SA2 remain relatively stable in 10 to 20 years (ABS, 2016)<sup>1</sup>. In light of the above, we shall see that SA2 has some features of TSA but it does not consider residents' transit usage characteristics and the typical service scope of transit stops therein. It is likely that in the urbanized core, as SA2s are so small that transit stops in one SA2 can serve residents therein as well as residents from neighbouring SA2s. If this the case, our estimated distinct transit riders by SA2 may not be as accurate as we expect. Outside Australia, counterparts of SA2 exist as well. Different transport planning agencies across countries, for instance, usually define and use Traffic analysis zone (TAZ). Therefore, our studies can be replicated outside Australia rather easily so long as TAZ boundaries, smartcard data and TSA data are available. Of course, regardless of the unit of analysis, spatial analysis, including analysis based on the excess-commuting framework, is often prone to the modifiable areal unit problem (MAUP) (for more information, please see Horner and Murray, 2002).

---

<sup>1</sup> For more info, please see: <http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/88F6A0EDEB8879C0CA257801000C64D9>, accessed 24 July, 2016.

When quantifying and visualising TSA, this study focuses on descriptive analyses of the 172 days' metrics, trip distribution between TSAs and trip generation/attraction by TSA. Regardless of whether it is  $T_{act}$ ,  $T_{min}$ ,  $T_{max}$  and  $T_{rand}$ , for a given day, the trip generation/attraction by TSA are fixed per the excess-commuting framework. In this study, this framework is slightly adapted and we assume the following:

- A city and region can be divided into different TSAs (in this study, SA2 is used as a proxy of TSA);
- Trips can occur within and between SA2s;
- All trip makers are homogeneous in terms of their locational preference and can be enticed to any SA2, that is, their trip destinations and origins are tradeable;
- The actual trip distribution within or between SA2s is only one of the many possible distributions between SA2s;
- Travel cost within an SA2 or between any two SA2s remain the same for each of the above four scenarios, e.g., the cost is always the linear distance between centroids of the two SA2, regardless of how many trips there are;
- Across the scenarios, even the cost of travel remains the same, transit riders' EDT can still vary because of reasons such as extreme weather and personal preferences;
- For trips occurring within an SA2, the travel distance is the radius of the SA2, which is assumed to be a circle with the same area as the SA2.

When transit riders' EDT is extremely elastic or inelastic, the minimum and maximum travels occur (Scenarios #2 and 4). When transit riders' EDT follows a random distribution, the random travel arises (Scenario #3, c.f., Charron, 2007 and Murphy and Killen, 2011). We could use the transportation problem algorithm (see White, 1988) and the hit-and-run algorithm (see Murphy and Killen, 2011) to quantify the minimum, maximum and random travels. Zhou et al. (2014) and Zhou and Long (2016) have developed methods to visualise trip generation/production and distributions when the minimum and maximum travels occur. In this study, we borrow these methods when visualising trip generation/production and distributions when the minimum and maximum travels.

Specifically, the minimum and maximum travels can be found by solving the following transportation problem:

$$\text{Min or Max: } Z = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n c_{ij} X_{ij} \quad (1)$$

$$\text{s.t. } \sum_{i=1}^n X_{ij} = D_j \quad \forall j = 1, \dots, m \quad (2)$$

$$\sum_{j=1}^m X_{ij} = O_i \quad \forall i = 1, \dots, n \quad (3)$$

$$X_{ij} \geq 0 \quad \forall i, j \quad (4)$$

where,

$m$  = number of origins;

$n$  = number of destinations;

$O_i$  = trips beginning at zone  $i$ ;

$D_j$  = trips destined for zone  $j$ ;

$c_{ij}$  = travel cost from zone  $i$  to zone  $j$ , in this study, we use the linear distance between  $i$  and  $j$  as the cost;

$X_{ij}$  = number of trips from zone  $i$  to zone  $j$ ;

$N$  = total number of trips.

The objective function (2) minimises average transport costs. Constraint (3) ensures that trip demand at each destination zone is satisfied while constraint (4) limits the number of trips leaving each origin zone to the number of trips originating there. Constraint (5) restricts the decision variables,  $X_{ij}$ , to non-negative values. It should be noted that travel costs,  $c_{ij}$ , may be expressed in terms of any measure of zonal separation, for example travel distance, travel time or indeed a generalised cost measure.

The approach for calculating  $T_{\text{rand}}$  is the Markov Chain Monte Carlo hit-and-run algorithm described in Murphy and Killen (2011, p.1261-62). It is given by the following:

$$N_T = N! \quad (5)$$

$$\text{s.t.} \quad \sum_{i=1}^n X_{ij} = D_j \quad \forall i = 1, \dots, n \quad (6)$$

$$\sum_{j=1}^m X_{ij} = O_i \quad \forall j = 1, \dots, m \quad (7)$$

$$X_{ij} \geq 0 \quad \forall i, j \quad (8)$$

where  $N_T$  is the number of possible trip configurations in a study area;  $N!$  is the factorial of the total number of trips in the study area. Constraints (6-8) are identical to those of the transportation problem and they limit trip distribution possibilities to those supported by the fixed distribution of origins and destinations by SA2. Finally,  $T_{\text{act}}$  was calculated from observed trip data and associated travel costs (In this case, it is the linear distance between centroids of SA2s or the “radius” of a SA2).

$T_{\text{act}}$ ,  $T_{\text{min}}$ ,  $T_{\text{max}}$  and  $T_{\text{rand}}$  show how different TSAs (SA2s) interact with one another and the interactions are characterized by the trip distributions when  $T_{\text{act}}$ ,  $T_{\text{min}}$ ,  $T_{\text{max}}$  and  $T_{\text{rand}}$  are achieved. Graphically, Figure 1 shows simple examples of  $T_{\text{act}}$ ,  $T_{\text{min}}$ ,  $T_{\text{max}}$  and  $T_{\text{rand}}$  where there are only four TSAs and four riders. In the figure, Circles A, B, C and D represent four TSAs that are not adjacent to one another. Each line with an arrow represent a trip unless otherwise stated. The arrow shows the direction of the trip. For those trips within a TSA, we assume that they all go from the border to the centroid of a TSA.

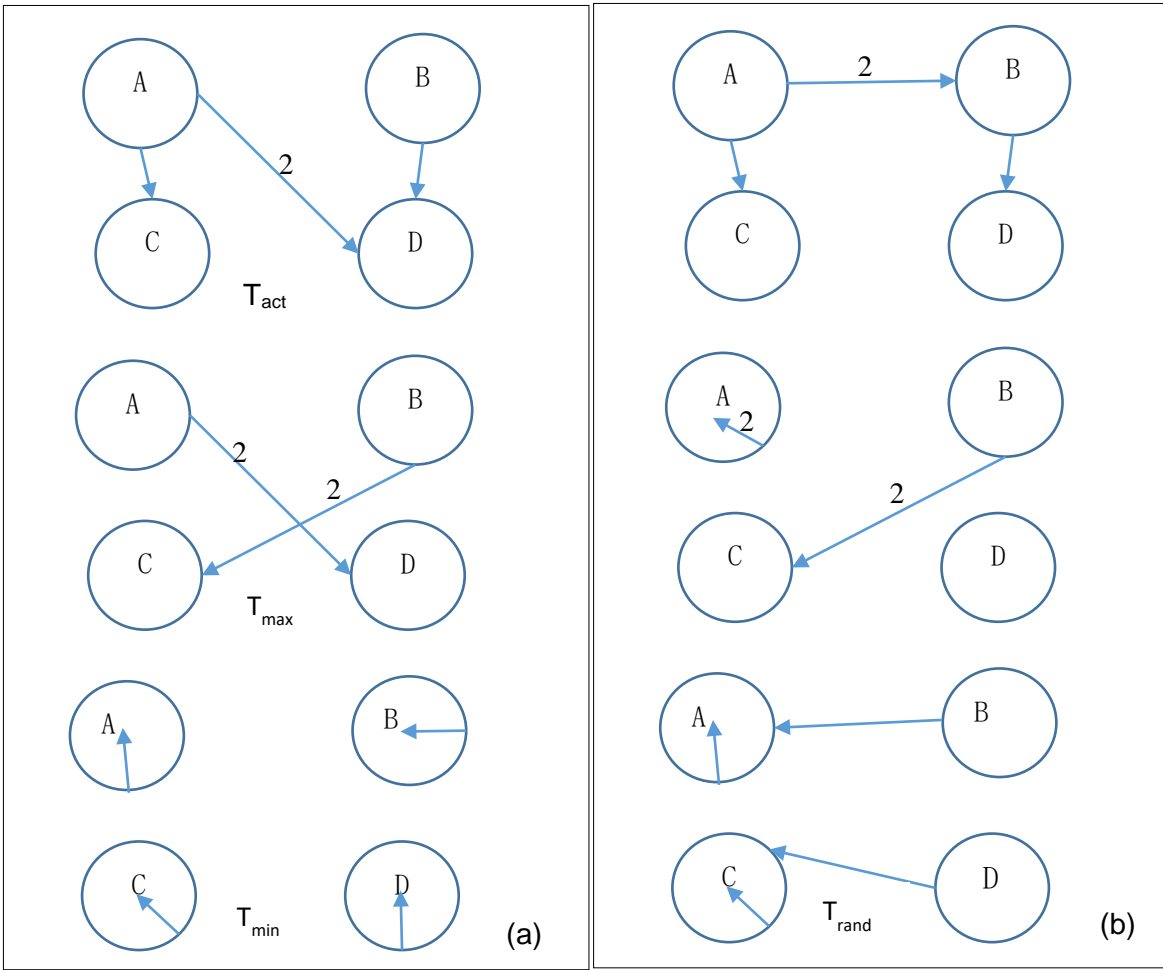


Figure 1: Examples of  $T_{min}$ ,  $T_{max}$ ,  $T_{act}$  and  $T_{rand}$

$T_{act}$  is how trip makers actually travel within or between different TSAs (Scenario #1, the upper portion in Panel (a) of Figure 1). It is also thus the actual trip distribution within or between TSAs. In most cases, it should not be identical to the  $T_{min}$ ,  $T_{rand}$  or  $T_{max}$  trip distributions, which never or rarely happen in reality.

$T_{min}$  is achieved as all the four trip makers' EDT is so elastic and they collectively minimize their overall trip costs (Scenario #2, the lower portion in Panel (a) of Figure 1). This can also occur in reality to some degree, for instance, when distance-based transit fare is so expensive or traffic congestion is so severe that all trip makers shorten their travel distance as much they could. It can also happen when land use mixture is so good that trip makers on average only need to travel within or to the closest TSA and have their respective needs met. In Figure 1,  $T_{min}$  emerges when all the four trip makers make internal trips, that is, they only travel within a TSA.

$T_{rand}$  is not about a single trip distribution and can be any trip distributions, including those for  $T_{act}$ ,  $T_{min}$  and  $T_{max}$  (Scenario #3, Panel (b) of Figure 1).  $T_{rand}$  thus in general means that trip makers' EDT follows some random distribution and they randomly decide whether/where they need to travel inside or outside a TSA. This could happen, for instance, when trip makers treasure amenities such as certain neighbourhoods and good schools so much that they always maximize access to those amenities first before considering the cost of travel. As a result, their trips, especially commuting trips, can be longer or more expensive than what  $T_{min}$  can offer. In other words, trip makers make trade-off between travel costs and other benefits (or costs) and their route choice is not necessarily the one that minimizes the overall travel cost.



$T_{max}$  is achieved when trip makers are extremely inelastic to the cost of travel (Scenario #4, the central portion in Panel (a) of Figure 1). This can occur, for instance, when transit fare is free or heavily subsidized and all trip makers thus indulge themselves in traveling to faraway and even the furthest destination so as to save other costs such as entertainment, food, mortgage and rent. It can also happen in reality more or less when different land uses or crucial facilities are so separated that trip makers must travel a long distance to satisfy their respective needs.

In light of the above, the four scenarios and metrics  $T_{min}$ ,  $T_{max}$ ,  $T_{act}$  and  $T_{rand}$  can inform transit planners and operators about how people’s demand for transit (EDT in this study) is related to transit cost (in this study, the distance of travel), land use and amenities. They reflect how the transit demand/trips would be allocated to different transit routes or even be translated into bicycle or walk trips (e.g., when trips occur within a TSA and they can be bicycle or walk trips). Of course, the four scenarios also show how different TSAs can be connected (via passenger flows) when EDT changes (See Figure 1 for examples). When looking across days, the temporal variations of the four metrics for each policy scenario can show us how parsimonious, random, or indulgent transit riders can be in terms of their demand (measured by  $T_{min}$ ,  $T_{max}$ ,  $T_{act}$  and  $T_{rand}$  values and the total number of transit riders across days) and related route choice.

In SEQ, local smartcard data record nearly 90 percent of all transit trips (TransLink, 2016). Therefore, this study captures most of the population that used ferries, buses and rail in SEQ over a six-month period. This is an important difference from previous studies (since 1998), which are mostly based on extrapolated and sampled data (see Table 2 for a list of representative existing studies we identified) and typically only cover a few days of the year (except LEHD, which are based on continuous administrative data from multiple sources). Of course, unlike previous studies, this study looks at transit trips of all purposes and thus the interpretations of related metrics and related results are different too, which have been described above.

Table 2: Selected Previous Studies

Source	Data	Sample size(s)	Metrics
Frost et al, 1998	Worktravel data of UK Census	Unreported	$T_{act}$ ; $T_{min}$
Ma and Banister, 2006	Census of Population and Housing	2% of the population	$T_{min}$ ; $T_{act}$ ; $T_{max}$
Horner, 2007	Census Transportation Planning Package (CTPP)	Unreported	$T_{min}$ ; $T_{act}$ ; $T_{max}$
Yang and Ferreira, 2008	CTPP	Unreported	$T_{min}$ ; $T_{act}$
Murphy, 2009	Traffic simulation model	Unreported	$T_{min}$ ; $T_{act}$ ; $T_{max}$
Horner and Schleith, 2012	Longitudinal Employer Household Dynamics (LEHD), US Census	Unreported	$T_{min}$ ; $T_{act}$ ; $T_{max}$
Loo and Chow, 2011	Census and Statistics, Information Services, and Transport Departments' data	Unreported	$T_{min}$ ; $T_{act}$ ; $T_{max}$

Zhou and Long, 2014	Smartcard data	216,844	$T_{\min}; T_{\text{act}}; T_{\max}$
Schleith and Horner, 2014	LEHD	Unreported	$T_{\min}; T_{\text{act}}; T_{\max}$

## 4. Site and Data

SEQ, Australia is the site for this research. SEQ occupies an area of about 22, 240 square kilometres in the eastern portion of Australia. Within SEQ, there are 11 local government areas with an estimated population of 3.27 million, 70.2% of the total Queensland population (Queensland Treasury, 2015). Since August 2012, TransLink, a division of the Queensland Department of Transport and Main Roads, has been responsible for the delivery of passenger transport services across Queensland, including SEQ. Public transport remains a popular way to get around in SEQ, with over four million transit trips each week (TransLink, 2016). Nearly all local passengers and most visitors to SEQ use a smartcard, the “Go card”, to pay their fare when using the system, comprised of buses, ferries and trains. The Go card is a more popular option amongst local and foreign passengers because TransLink charges up to 63% extra for those using a paper ticket. To accurately charge their fares, TransLink requires that Go-card users tap their card for both boarding and alighting, which is the same as Transport for London. Few other transit agencies, however, ask transit riders to tap in and out. Thus, the Go card data are somewhat special. These Go card taps automatically generate the following information:

- Trip ID (a unique “leg” of a journey) and journey ID (stop of origin-stop of destination, based solely on fare rules),
- The stop information for trip origin and destination,
- Boarding and alighting time, and
- Unique card number and card type (concession card or regular card).

The information is instantly sent to and stored on a central server. For this study, six months of data were requested from TransLink. TransLink defines a journey as the set of trips taken under one fare basis. TransLink considers two consecutive transactions linked into a single journey if the time gap between them (alighting to subsequent boarding) is less than 60 minutes. The fare is discounted for transfers, and a maximum of three transfers are allowed for a journey.

The data used in this study covers a six-month period from 1 November 2012 to 30 April 2013 across SEQ. The total number of raw trip transaction records was 69,194,428 with 49,159,474 journey transaction records identified after data processing. One thing should be noted is that the journeys identified are all kinds of trips, not just work trips, many of which we must assume are discretionary. To complement the dataset derived from smartcard taps, data from other sources were used as well, including Google Transit Feed Specification (GTFS) data and ABS’ SA2 zonal data.

The framework for our Go card data processing approach is shown in Figure 2. It is composed of three parts: data preparation, demand calculation and distance calculation. The database provides information on passenger trips and geometry information.

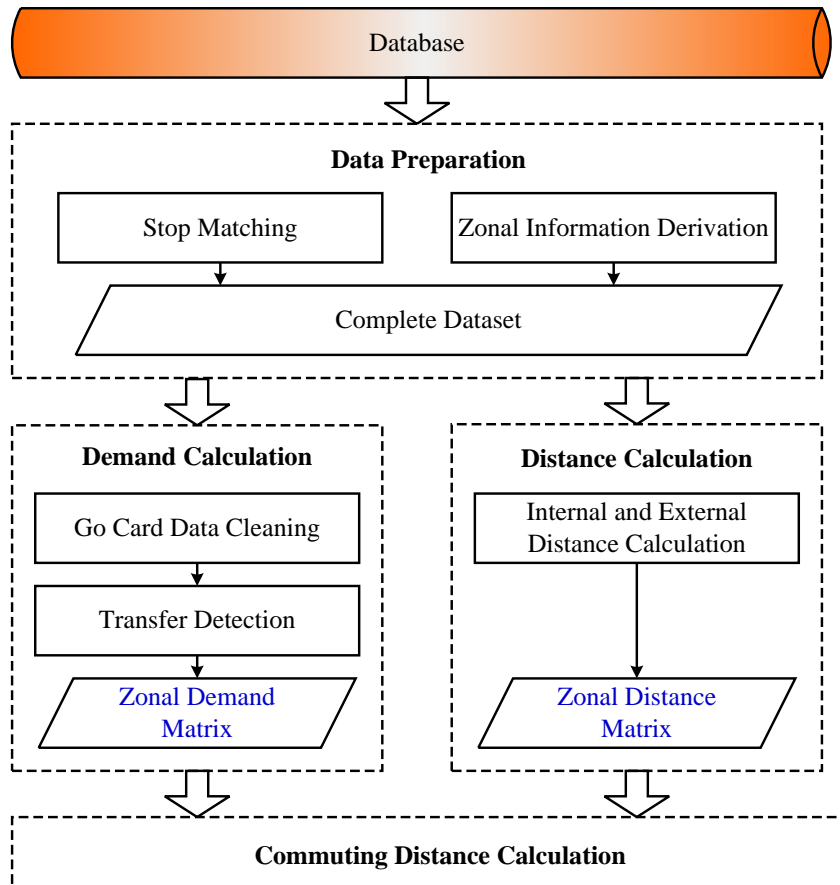


Figure 2: Framework for data processing

Data Preparation: this step matches stops among *Go* card, GTFS and SA2 zonal databases, and derives zonal information (centroid and area radius). The outputs are *Go*-card stop ID, GTFS stop ID, stop coordinate (Latitude and longitude), zone (SA2) ID, zonal centroid coordinate, and zonal area radius.

In practice, the stop IDs are not necessarily consistent across different operators. The *Go* Card dataset does not include the latitude and longitude of the stops, nor do they always match the stop names that TransLink specified in the local transit schedules. A “off-optimality” heuristic was first developed to distinguish the transfer interchanges from the activity locations and to single out different journeys (Nassir et al., 2015). To calculate fare, Translink defines journey as a set of related trips. Each trip produces one “tap on” and “tap off”, which together are recorded as a transaction in the smartcard data. If a journey contains more than one trip, TransLink has its own fare rule. It only charges one fare if two or more consecutive transactions occur less than 60 minutes and consider those trips as a single journey. When processing smartcard data, we adopted a slightly different definition of “journey” so as to more accurately distinguish the transfer interchanges from the activity locations. Based on local household travel data, we established the following rules to single out the journeys that we define: if the successive transactions are on the same route, then there could be one or two journeys, depending on the boarding times of two trips or the distance between the two boarding stops: (a) if the temporal gap between two boarding times is larger than 40 min, then there are two journeys; (b) if the spatial gap between two boarding stops is larger than 400 meters, then there are two journeys (Alsger et al., 2015).

Once individual journeys are singled out, we use Application Programming Interface (API) to access the local GTFS data. The purpose of this is to match the journey’s origins and destinations recorded by the smartcard data with those of the local GTFS data, adding extra information such as transit

stops' coordinate information to the former. API is a set of functions and procedures that allow the creation of applications which to access the features or data of an operating system, existing application, or other services. TransLink authorized us to access its GTFS data via API. Our API uses journey ID, trip origin and destination stop names, boarding and alight time as input, which the smartcard data record. After “communicating” with TransLink’s GTFS data, our API returns a series of text files, which contains extra information such as coordinates of stops, fares, routes, stops, transfers and trips. Given that the transits stops in SEQ are rather stable, the text files enable us to retrieve 90% of the journey’s origins and destinations recorded by smartcard with that of the local GTFS data. The text files can be fed into ArcGIS 10.2 to create .shp files for mapping origin and destination of each journey and to calculate the linear distances between TSAs. The .shp files of the origin and destination (points) and of the SA2 boundaries (polygons) are used together to assign each origin and destination the unique ID of SA2, again in ArcGIS 10.2. In ArcGIS 10.2, a spatial join can assign the attributes such as unique ID of the polygon directly to the attribute table of the output point file.

**Zonal Information Derivation:** The zonal (SA2) centroid and zonal area radius was derived using the ArcGIS 10.2’s Production Editing function and the built-in mathematical function of attribute tables. The input data are SA2 .shp files. Much of the above work is not difficult but tedious. However, given that many can reuse the heuristic/procedures to complete the work, it is recommended that TransLink in the future offers some standard add-ins about those heuristic/procedures when it release *Go* Card data. It is even better that TransLink offers identifiers for different transit stops by spatial units that are commonly used by local transport planners and modellers such as SA2 and TAZ.

**Demand Calculation:** this step aims to further clean *Go*-card transactions, perform transfer detection and calculate zonal (SA2) origin-destination demand matrix. The outputs are zonal based internal and external transit travel demand, that is, transit journeys within or between SA2s in our case study.

Cleaning and refining the data are important steps in data processing. Raw data usually contain erroneous records caused by system failure or human error. The archived *Go* card data were screened to minimize the possibility of erroneous data by setting different rules and using schedule information from GTFS system. The results of the cleaned data indicated that 17 percent of trip records (n=11,763,053) were excluded from all the raw trip records (n=69,194,428) due to checking and fixing of erroneous data with different types of errors. These errors are summarized in Table 1.

Table 1: Description of *Go*-card Transaction Errors

Error types (%)	Description (Causes)
System failure (1%)	No boarding information, and no alighting information (unknown)
<i>Go</i> card reload (3.4%)	An additional transaction record for the same passenger trip (top up <i>Go</i> card in a vehicle)
Extremely large interval transaction (3.5%)	The difference between the alighting time and boarding time for a transaction is larger than 2 hours or across several days (forget to touch off from the last trip)
Ticket evasion (2%)	Boarding stop equals to alighting stop (tap in front door and tap again in the back door)
Driver faulty operation (7%)	The service direction in the transaction is wrong, e.g. the sequence of the inbound stops are actually the outbound stops (driver forget to change the route display information)
Abnormal stop	The boarding or alighting stops are not in the stop list of the recorded service route

---

(0.1%)	(the vehicle does not stop at the designated stop due to bus bunching)
Total=17%	All errors
(n=11,763,053)	

---

The zonal based internal and external demand was calculated by assigning each journey transaction record to the origination and destination zone matrix based on boarding and alighting stop coordinates.

Distance Calculation: this step aims to calculate the zonal OD distance (that is, distance between SA2s) matrix. Internal distance (that is, the distance within TSAs/SA2s) is the radius of the zone (SA2, calculated assuming a circle area) and external distance is the Euclidean distance between the corresponding centroids of two SA2s. In our studies, we do not use fare to represent travel cost ( $c_{ij}$ ) due to data unavailability, as we were not given the fare information for the base scenario and we are not sure how much fare a transit rider would pay in Scenarios 2 to 4. In Scenarios 2 and 4, transit riders would take some routes or route segments that they did not travel in the base scenario. Thus, even we know the fare of the base scenario, we still have to figure out the fare of the former in order to estimate the fare of Scenarios 2 to 4.

## 5. Empirical Results

### 5.1 Descriptive analysis results

Table 3 presents the means and standard deviations (SD) for the four metrics of interest. The means and SD were based on 50 weekends and 50 randomly drawn weekdays from 1 November, 2012 to 30 April, 2013. The four metrics listed in Table 3 can represent how different SA2 (TSAs) could interact with each other when transit riders' EDT vary.

Values in Table 3 indicate that a different number of riders patronize TSAs across days and the numbers vary quite notably (also see Figures 5 and 6 for visuals of the changing shapes and sizes).

Table 3: Four Metrics of Interest for 172 Days

		$T_{min}$	$T_{act}$	$T_{rand}$	$T_{max}$
EDT		Extreme elastic	Status quo	Randomly distributed	Extreme inelastic
Mean (km)	All days	1.47	10.90	26.20	31.60
	Weekdays (n=50)	1.43	11.08	24.40	29.23
	Weekends (n=50)	1.55	10.24	31.14	38.19
SD (km)		0.14	0.67	3.8	5.00
SD/Mean		9%	6%	15%	16%

Figure 3 shows how the four metrics vary across days over the 172 days. Paired sample student-t tests indicate that all the four metrics' weekend mean and weekday mean are significantly different. The  $T_{act}$ , for instance, has a weekend mean of 10.24 (Saturday and Sunday, n=50) and a weekday mean of 11.08 (n=50, randomly drawn from 122 weekdays, excluding Australian public holidays). The two samples are significantly different,  $t(49)=-4.560$ ,  $p=0.000$  (2-tailed).

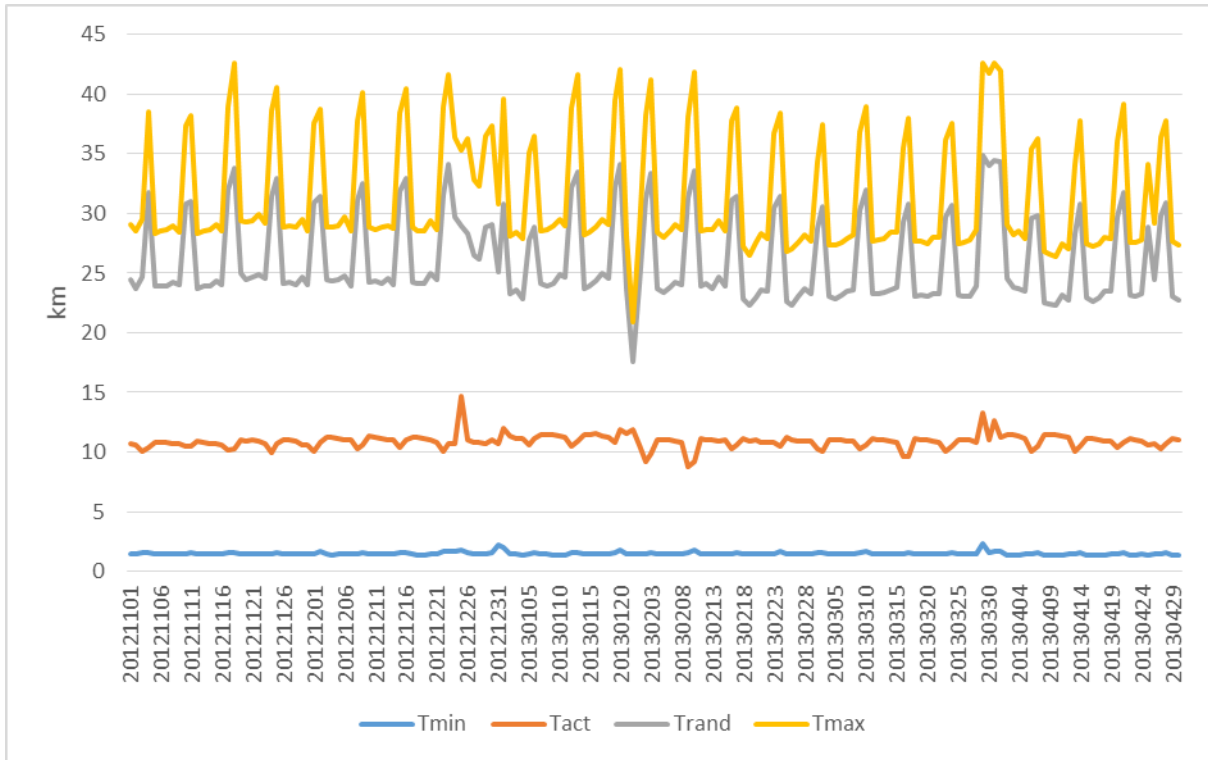


Figure 3: Four Metrics Variations over 172 Days

Figure 4 shows how the number of transit riders (that is, demand for transit) vary by day. More or less, who are these riders and where they are travelling to and from by day contribute to the variations of the four metrics. The number of transit riders can be as small as 20,000ish and as big as 250,000ish. This partly shows why it is challenge for a transit system to offer a right suite of transit services to meet the demand of such a great variation across days.

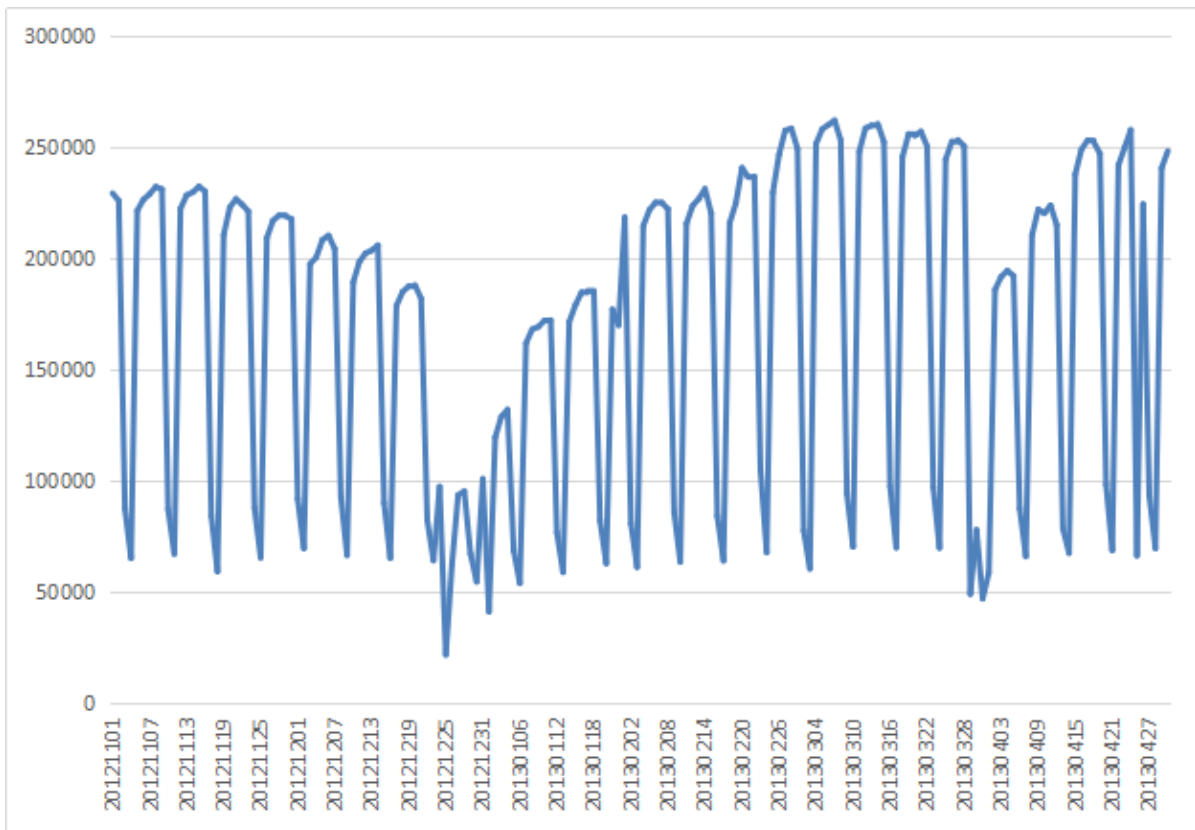


Figure 4: The Number of Transit Riders over 172 Days

Again, a paired sample student-t test indicates that the mean of riders traveling on weekends (Mean=87,483) and weekdays (Mean=185,308) are significantly different,  $t(49)=-8.864$ ,  $p=0.000$  (2-tailed). This finding more or less justifies why many transit agencies have different schedules for their services on weekdays and weekends.

Except the maximum of  $T_{min}$  and  $T_{max}$ , the minimum and maximum for  $T_{act}$ ,  $T_{min}$  and  $T_{max}$  occur on different days. Table 4 presents these days and the corresponding values of  $T_{act}$ ,  $T_{min}$  and  $T_{max}$ .  $T_{rand}$  is not studied and discussed here as it would vary, depending on how many random EDT distributions one generates when calculating the mean of  $T_{rand}$  on each day. Unlike the other three metrics, the  $T_{rand}$  value for a given day is an estimated mean of many possible trip distributions rather than a fixed value for one trip distribution.

$T_{act}$ 's maximum value occurs on Christmas Day (25 December 2012) and  $T_{min}$  and  $T_{max}$ 's maximum values occur on Good Friday (29 March 2013) if we do not treat holidays separately during the study period. But if we take these holidays into account and separate them from other days,  $T_{act}$ ,  $T_{min}$  and  $T_{max}$  reach their respective maximum values on three different days. But  $T_{act}$ 's maximum value still tends to occur on a special day, which is the last day of 2012. Frequency of transit services tends not to influence the extreme values of  $T_{act}$ ,  $T_{min}$  and  $T_{max}$ .  $T_{act}$ 's two extreme values, for instance, happened on two weekends, when the frequency of transit services was significantly lower than on weekdays. Thus, holidays tend to contribute to the minimum or maximum of  $T_{act}$ . In other words, we may need some adjustments for transit services on those days. Our data currently do not allow us to deduce what causes the minimum and maximum. But it can be interesting to find extra data to study it.

**Table 4: Extreme Values for  $T_{act}$ ,  $T_{min}$  and  $T_{max}$**

Extreme and characteristics		$T_{min}$	$T_{act}$	$T_{max}$
Minimum	Date	10 Apr 2013	09 Feb 2013	22 Jan 2013
	Day of Week	Wed.	Sat.	Tue.
	Distance (km)	1.35	8.80	20.93
	Riders*	220,372	86,017	170,102
Maximum	Date	29 Mar 2013** (31 Dec 2012***)	25 Dec 2012** (31 Mar 2013***)	29 Mar 2013** (18 Nov 2012***)
	Day of Week	Friday (Mon.)	Tuesday (Sun.)	Friday (Sun.)
	Distance (km)	2.27 (2.23***)	14.64 (12.58***)	42.62 (42.60***)
	Riders*	49,337 (101,043)	22,048 (47,263)	49,327 (59,480)

\*This is the number of riders when  $T_{min}$ ,  $T_{act}$  and  $T_{max}$  occur, not the minimum or maximum numbers of riders that TransLink served over the 172 days.

\*\*All days are treated the same.

\*\*\*Holidays such as Christmas Day, Good Friday and New Year's Day are excluded.

## 5.1 Visuals

Figures 5 and 6 show the trips attracted per resident by SA2 when  $T_{act}$  reaches its minimum and maximum during the 172 days and corresponding trip distributions. Actually, one can also make similar figures for  $T_{min}$ ,  $T_{max}$  and  $T_{rand}$ . These figures, of course, more or less resemble Figures 5 and 6 as trip generation and attraction by TSA are assumed to be fixed for a given day across all the four scenarios. But they have quite different meanings, as described earlier. Figures of  $T_{min}$ , for instance, show when trip makers are extremely elastic to cost of travel (See Figure 1's panel (a)). In light of length constraints of a manuscript, we do not present all the figures here because there are as many as  $172 \times 6 = 1,032$  after we made one distinct figure for each metric (i.e., one trip-generation and one trip-attraction figure for all the four metrics, one trip-distribution figure for each of the four metrics) per day. But they are available upon request. Our discussion and conclusions, nevertheless, are based on eyeballing of all the 1,032 figures about  $T_{act}$ ,  $T_{min}$ ,  $T_{max}$  and  $T_{rand}$ .

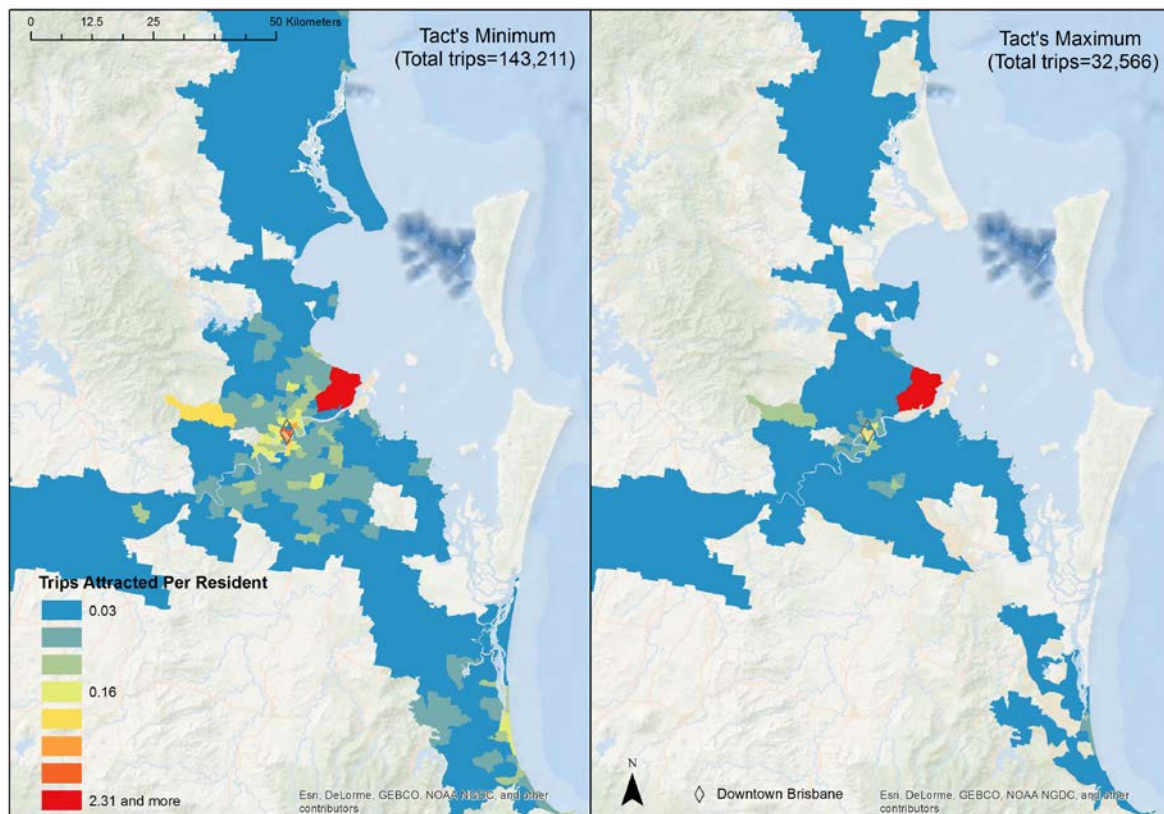


Figure 5: Trip Attractions per Resident When  $T_{act}$  Reaches the Maximum and Minimum



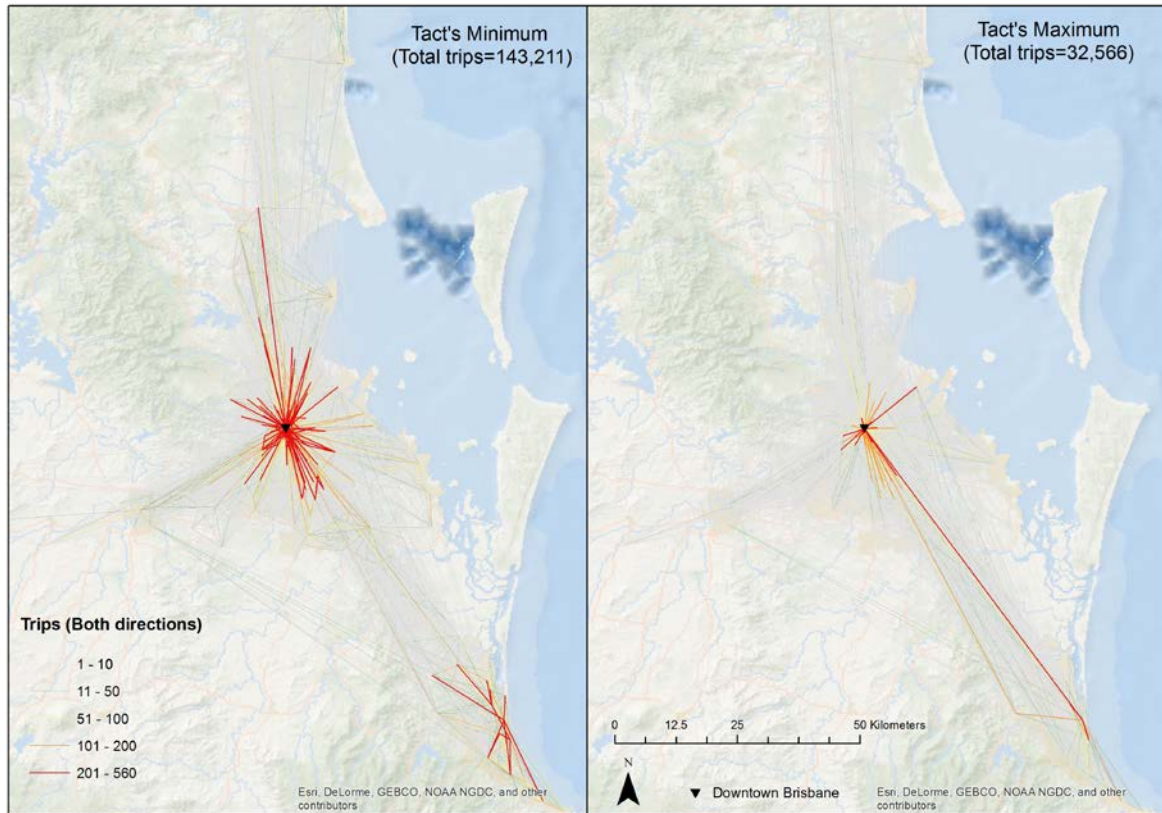


Figure 6: Trip Distributions When  $T_{act}$  Reaches the Minimum and Maximum

## 6. Discussion and conclusions

Overall, the quantitative results demonstrate that there are great variations in TSAs when we assume that transit riders' EDT varies, which can result in four different metrics:  $T_{act}$ ,  $T_{min}$ ,  $T_{max}$  and  $T_{rand}$  as well as different numbers of riders, trips generated/attracted and trip distributions over time and across space. Our results show that the extremes of  $T_{act}$  occur on special days such as Saturday or holidays such as Christmas Day. Interestingly, on Christmas day, transit riders ( $n=22,048$ , about one ninth of the 172 days' average), on average, travel to the furthest destination than any other day. It is a Saturday ( $n=86,017$ ) that, transit riders are most inactive by making on average the shortest trip over the 172 days. If we assume that transit riders' EDT is extremely elastic,  $T_{min}$ 's minimum (1.35 KM) can represent an extreme where transit riders make on average the shortest trip among across the four scenarios. In SEQ,  $T_{min}$ 's minimum occurs on a Wednesday and there are as many as 220,372 transit riders, 20% more than the average number of riders that TransLink serves daily over the 172 days. This extreme and the corresponding number of transit riders together imply that SEQ could significantly reduce its average trip length for transit riders. Its current destination/origin distribution or land use can also in theory support this. On the other hand,  $T_{max}$ 's maximum (42.62 KM) can represent another extreme where transit riders' EDT is so inelastic and they must on average make the longest trip to maximize their utility. Luckily, there are not that many transit riders then ( $n=49,327$ , about half of the 172 days' average ridership). However, this extreme reflects that in a worse scenario than the status quo when complementary origins and destinations are more spread out, transit riders in SEQ can expect on average a much longer trip than now. Thus, the four metrics:  $T_{act}$ ,  $T_{min}$ ,  $T_{max}$  and  $T_{rand}$  and corresponding numbers of riders, trips generated/attracted and trip distributions over time and across space provide some new references for transit agencies to gauge and evaluate their efficacy in optimising service supply and examine how different factors such as seasonality, land use and fare change may influence such efficacy. This efficacy, for instance, can be measured by a ratio of

passenger\*kilometers utilized by riders and the total passenger\*kilometers offered by the transit system as a whole.

The visual representations show that TransLink does have TSAs that vary greatly from day to day and across space, whether we visualise TSAs (SA2s) by transit trips generated/attracted by SA2 or transit trips within or between SA2s. The numbers of total trips generated/attracted by SA2 and riders served by TransLink also change notably across days. Weekends and holidays see much less trips and riders than weekdays. Given that some SA2s (e.g., some SA2s in the west of Downtown Brisbane) have on average 20 trips for the entire 172-day period and that each trip on average costs nearly \$7 subsidy for TransLink (TransLink, 2016), it is worthwhile for TransLink to explore service provision options, e.g. whether a vanpool or paratransit program can be more cost effective for riders to and from these SA2 than fixed route bus services. In Los Angeles, for instance, an employer subsidized and ran a vanpool program and can make a vanpool trip for its commuting employee as cheap as US\$3 (Zhou et al., 2012). What's more, the vans in the program do not have a fixed route and thus avoid traffic congestion by choosing a less congested route where van drivers, who are also vanpooling employees, see fit. With the rise of shared-ride and car-sharing services such as Uber and Zipcar, transit agencies like TransLink should consider how to partner with them to create more win-win situations for transit riders and even travelers at large. Those who vanpool or carpool because of reduced or consolidated transit services, for instance, could use Uber or Zipcar for emergency rides, ad hoc needs or first/last mile trips. Uber can "rent" a spare driver and a bus from a local transit agency when there is a group of customers. With continuous provision of and the provision of a more comprehensive picture of TSAs, which is further enabled by smartcard data, transit agencies like TransLink should re-think their service operations and planning strategies. Based on a full-year's trip distribution by day, for instance, they could optimise their resource allocation, taking into account factors such as seasonality, spatial regularity/variability of demand, land use changes and fare adjustments.

The above metrics and visual representations have shown how smartcard data and novel analytics/visual representations can together improve and even revolutionise transit system planning, monitoring and operation in the future. Pelletier et al. (2011) argue that smartcard data can help us accomplish transit-related tasks at three levels: (a) strategic (e.g. long-term planning); (b) tactical (e.g. services adjustments and network development); (c) operational (e.g. ridership statistics and performance indicators). But they do not provide many examples. This study has provided concrete examples and some transferrable metrics at all three levels:

At the operational level, these metrics and visual representations of over 172 days can show where transit riders are most likely or unlikely to use transit services continuously, where we can intensify or consolidate transit services and where there could possibly be transit-service gaps or oversupply based on a long period of time (e.g., 1 year).

At the tactical level, the mean of daily ridership for a long period of time (e.g., 1 year) and corresponding TSA maps based on smartcard data can serve as something similar to average annual daily traffic to transportation engineers (Caltrans, 2015), which can be used as guidelines for transit fleet planning, services adjustments and resource/service optimization.

At the strategic level, the combined quantitative and visual representation results about TSAs,  $T_{min}$ ,  $T_{act}$  and  $T_{max}$  can be used to integrate long-term transit and land use planning.  $T_{min}$  and corresponding visual inform us about how short the trip length of transit riders could possibly be on average and where trip length can be shortened (relative to  $T_{act}$ ).  $T_{max}$  and corresponding visual show how much longer the trip length of transit riders could possibly be on average and where trip length may be lengthened (relative to  $T_{act}$ ). Knowing the "where" for a long term can help us better plan our land use so as to shorten average travel distance of transit riders,

**increase transit ridership or capacity along certain corridors and improve riders' overall travel experience.**

**Last but not least, transit supply in SEQ, in contrast with auto travel, places strong spatial constraints on people's travel. Network structure and service availability determine travel to a much higher degree for transit users than for the general population. Our studies of TSAs thus only disclose a small part of the complex picture of local travel in SEQ.**

## References

- ABS 2016, Statistical Area Level 2 (SA2). Available at:  
<http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/88F6A0EDEB8879C0CA257801000C64D9>, accessed July 24, 2016.
- Alsger, AA, Mesbah, M, Ferreira, L & Safi, H 2015, 'Use of Smart Card Fare Data to Estimate Public Transport Origin-Destination Matrix', *Transportation Research Record*, 2535, pp.88-96.
- Batty, M 2013, 'Big data, smart cities and city planning', *Dialogues in Human Geography* vol.3, no. 3, pp. 274-279.
- California Department of Transportation (Caltrans), State of California, 2015, 'Highway Design Manual', available at:  
[http://www.dot.ca.gov/hq/oppd/hdm/pdf/english/HDM\\_Complete\\_07Mar2014.pdf](http://www.dot.ca.gov/hq/oppd/hdm/pdf/english/HDM_Complete_07Mar2014.pdf)
- Calthorpe, P 1993, 'The Next American metropolis: Ecology, community, and the American dream', Princeton, NJ: Princeton Architectural Press.
- Cervero, R, Ferrell, C, and Murphy, S 2002, 'Transit cooperative research program, research results digest number 52.' Washington, D.C.: Transportation Research Board, National Research Council.
- Charron, M 2007, 'From excess commuting to commuting possibilities: more extension to the concept of excess commuting', *Environment and Planning A*, 39, pp. 1238–1254.
- Devillaine, F, Munizaga, M & Trepanier, M 2012, 'Detection of activities of public transport users by analyzing smart card data', *Transportation Research Record*, 2276, pp.48-55.
- Engels, B & Liu, G 2011, 'Social exclusion, location and transport disadvantage amongst non-driving seniors in a Melbourne municipality, Australia', *Journal of Transport Geography*, vol.19, pp.984-996.
- Eom, J K, Song, JY & Moon DS 2015, 'Analysis of public transit service performance using transit smart card data in Seoul', *KSCE Journal of Civil Engineering*, vol. 19, no. 5, pp.1530-1537.
- Frost, M, Linneker, B & Spence, N 1998, 'Excess or wasteful commuting in a selection of British cities', *Transportation Research A-Policy Practice*, vol. 32, no. 7, pp.529-538.
- He, L & Trepanier M 2015, 'Estimating the destination of unlinked trips in transit smart card fare data', *Transportation Research Record*, 2535, pp.97-104.
- Healy, E & O'Connor, K 2001, 'Jobs and housing location in Melbourne, 1986-1996', *Australian Planner*, vol.38, no. 1, pp.9-15.
- Horner, MW & Murray, AT 2002, 'Excess commuting and the modifiable areal unit problem', *Urban Studies*, vol. 39, no. 1, pp. 131-139.

- Horner, MW & Schleith, D 2012, 'Analyzing temporal changes in land-use transportation relationships: a LEHD-based approach', *Applied Geography*, 35, pp.491–498.
- Horner, MW 2007. 'A multi-scale analysis of urban form and commuting change in a small metropolitan area (1990–2000)', *Annals of Regional Science*, vol. 41, no. 2, pp.312-332.
- Janosikova, L, Slavik, J & Kohani, M 2014, 'Estimation of a route choice model for urban public transport using smart card data'. *Transportation Planning and Technology*, vol. 37, no.7, no.638-648.
- Kieu, LM, Bhaskar, A & Chung, E 2015, 'Passenger segmentation using smart card data', *IEEE Transactions on Intelligent Transportation Systems*, vol.16, no.3, pp.1537-1548.
- Kim, K, Oh, K, Lee, YK, Kim, S & Jung, JY 2014, 'An analysis on movement patterns between zones using smart card data in subway networks', *International Journal of Geographical Information Science*, vol.28, no.9, pp.1781-1801.
- Li, T, Corcoran, J & Burke, M 2012, 'Disaggregate GIS modelling to tract spatial change: exploring a decade of commuting in South East Queensland, Australia', *Journal of Transport Geography*, 24, 306-314.
- Loo, BPY & Chow, ASY 2011, 'Jobs-housing balance in an era of population decentralization: an analytical framework and a case study', *Journal of Transport Geography*, vol. 19, no. 4, pp.552-562.
- Ma, K.R. and Banister, D. 2006. 'Extended excess commuting: A measure of the jobs-housing imbalance in Seoul'. *Urban Studies*, vol. 43, no.11, pp.2099-2113.
- Ma, XL, Wang, YH, Chen F & Liu, JF. 2012, 'Transit smart card data mining for passenger origin information extraction', *Journal of Zhejiang University-Science C-Computers & Electronics*, vol.13, no.10, pp.750-760.
- Morency, C., Trepanier, M. & Agard, B. 2007. 'Measuring transit use variability with smart-card data', *Transport Policy*, vol. 14, pp. 193-203.
- Murphy, E & Killen, JE, 2011, 'Commuting economy: an alternative approach for assessing regional commuting efficiency', *Urban Studies*, vol.48, no.6, pp.1255-1272.
- Murphy, E 2009, 'Excess commuting and modal choice', *Transportation Research A-Policy Practice*, vol. 43, no. 8, pp.735–743.
- Nassir, N, Hickman, M & Ma, ZL 2015, 'Activity detection and transfer identification for public transit fare card data', *Transportation*, vol. 42, pp. 683-705.
- Pelletier, MP, M. Trepanier, M & Morency, C 2011, Smart card data use in public transit: A literature review, *Transportation Research Part C-Emerging Technologies*, Vol.19, pp. 557-568.
- Queensland Treasury 2015, *Population growth highlights and trends, Queensland regions, 2015 edition*. Queensland Government, Brisbane, Australia.
- Salas-Olmedo, MH. & Nogués, S 2012, 'Analysis of commuting needs using graph theory and census data: a comparison between two medium-sized cities in the UK', *Applied Geography*, vol.35, no.1, pp.132-141.
- Schleith, D & Horner, M 2014, 'Commuting, job clusters, and travel burdens: analysis of spatially and socioeconomically disaggregated longitudinal employer-household dynamics data', *Transportation Research Record*, 2542, 19-27.
- Tamblay, S, Galilea, P, Iglesias, P, Raveau, S & Munoz, JC 2016, 'A zonal inference model based on

- observed smart-card transactions for Santiago de Chile', *Transportation Research Part A-Policy and Practice*, 84, pp.44-54.
- Tao, S, Corcoran, J, Mateo-Babiano, I & Rohde, D 2014a. 'Exploring Bus Rapid Transit passenger travel behaviour using big data', *Applied Geography*, 53, pp.90-104.
- Tao, S, Rohde, D & Corcoran, J 2014b, 'Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap', *Journal of Transport Geography*, 41, pp.21-36.
- TransLink 2016, '*TransLink Tracker. October-December Q2*'. Brisbane, Australia: TransLink.
- The Transit Cooperative Research Program (TCRP) 2004, 'Transit-oriented development in the United States: Experiences, challenges, and prospects,' Washington D.C.: Federal Transit Administration.
- van der Hurk, E, Kroon, L, Maroti, G & Vervest, P 2015, 'Deduction of passengers' route choices from smart card Data', *IEEE Transactions on Intelligent Transportation Systems*, vo.16, no.1, pp.430-440.
- van Oort, N, Brands, T & de Romph, E 2015, Short-term prediction of ridership on public transport with smart card data', *Transportation Research Record*, 2535, pp.105-111.
- White, MJ 1988, 'Urban commuting journeys are not "wasteful"', *Journal of Political Economy*, 96, pp. 1097–1110.
- Yang, J & Ferreira, JJ 2008, 'Choices versus choice sets: a commuting spectrum method for representing job-housing possibilities', *Environment and Planning B*, 35, pp. 364-378.
- Zhong, C, Batty, M, E. Manley, E, Wang, JQ, Wang, ZJ, Chen, F & G. Schmitt, G 2016, 'Variability in regularity: mining temporal mobility patterns in London, Singapore and Beijing using smart-card data', *PLOS One*, vol.11, no. 2.
- Zhong, C, Manley, E, Arisona, SM, Batty, M & Schmitt, G 2015, 'Measuring variability of mobility patterns from multiday smart-card data', *Journal of Computational Science*, 9, pp.125-130.
- Zhou, J & Long, Y 2014, 'Jobs-housing balance of bus commuters in Beijing exploration with large-scale synthesized smart card data', *Transportation Research Record*, 2418, pp. 1-10.
- Zhou, J & Long, Y 2016, 'Losers and Pareto optimality in optimizing commuting patterns', *Urban Studies* (In Press).
- Zhou, J 2014, 'From better understandings to proactive actions: housing and commuting mode choices among university students', *Transport Policy*, 33, pp.166-175.
- Zhou, J, Murphy, E & Long, Y 2014a, 'Commuting efficiency in the Beijing Metropolitan Area: An exploration combining smartcard and travel survey data', *Journal of Transport Geography*, 41, 175-183.
- Zhou, J, Wang, Y. & Schweitzer, L 2012, 'Jobs-housing balance and employer-based travel demand management program returns to scale: evidences from Los Angeles', *Transport Policy*, 20, pp.22-35.
- Zhou, J., Murphy, E. and Long, Y 2014b. 'Visualizing the minimum and maximum solutions of the transportation problem of linear programming for Beijing's bus commuters', *Environment and Planning A*, 46, pp.2051-2054.