DEVELOPING PUBLIC TRANSPORT OD MATRICES FROM GPS AND SMARTCARD DATA

Laurence Chittock

Suzanne Callighan

Ben Mackley

Mott MacDonald, UK

ABSTRACT

Public transport Origin-Destination (OD) matrices are a key component of transport demand analysis and forecast, and are typically derived from manually collected survey data. Data of this form is often difficult and expensive to collect as it requires the manual counting and interviewing of passengers at multiple sites throughout a city, and the data is normally only collected on a single weekday during months where no events or school holidays take place. Furthermore, these surveys only collect a sample of travel patterns and are potentially subject to misrepresentation as a result. At the same time, the use of smartcards as a method of payment for public transport is becoming commonplace throughout many European and global cities, with patronage of these systems replacing other means of payment. Automated data collection, where boarding events are recorded for individual smartcards, allow analysts to derive large scale and longitudinal travel patterns which can be used to replace traditional and expensive surveys. However, smartcard data alone can often only provide partial information with trips recorded at the fare stage and when a passenger boards.

In this paper, and building on the work of others (Munizaga and Palma, 2012), we demonstrate how GPS data can be used to determine buses' locations in relation to their stops, and how this can be matched with smartcard data to provide boarding locations for most records. Using data from the city of Tbilisi in Georgia, where all publicly run buses are fitted with GPS devices, and smartcards are used for ~80% of bus and metro trips, we analyse individuals' trip chains to infer where they alighted using a similar approach to Gordon (2012). We demonstrate how this method can be applied to large datasets, encompassing the whole city across several days. This information is then used to build an OD matrix for bus and metro travel in the city which can be used as an input into a transport model, or used by planners to understand public transport usage in more depth than would be possible via survey data.

1. INTRODUCTION

Face-to-face interviews and manual, or video-based, passenger counts have traditionally been used to assess public transport level of service and understand where people are travelling from and to. These assessments help feed service provision and can be used to develop public transport Origin-Destination (OD) matrices, which form the building blocks of a transport model. A survey approach, typically where passengers are interviewed while waiting for a service, allows true origins and destinations to be asked for, and supplementary information such as journey purpose can also be ascertained. However, as a means of determining level of service across a whole city, it can be costly and time-consuming. Because of this, the data is normally based on a small sample and extrapolated to represent movements and demand across a whole year.

In recent years, smartcard ticket systems have been introduced by transport authorities across many European cities to standardise payment and speed up passenger boarding. In addition, these systems collect vast amounts of data, allowing analysts to derive large scale and longitudinal travel patterns which can be used to replace traditional and expensive surveys. The advantage of using this data source is that it is 1) cheaper (the data is already being collected), and 2) it is available over months and years, meaning the output results are less prone to sampling error, or results can be analysed during particular dates or events such as on public holidays. In some cases, smartcard data is recorded both at point of entry into the system, and at point of exit. If the locations of these transactions are known, as in the case of Jang's (2010) analysis of smartcard data in Seoul, South Korea, then it is relatively straightforward to develop matrices from this boarding and alighting information.

In many cases however, only partial information is recorded, with trips sometimes only recorded at the fare stage and only when a passenger boards (their entry point). Locations defined by fare stage are often complex to infer, since the stages tend to vary by route and bus company, and where they are known, their spatial granularity can be poor (i.e. there may only be 4 or 5 fare stages for a route, meaning a stage can represent distances well over 1km). To overcome this lack of information, it is often necessary to combine the ticket data with GPS traces from buses to determine where the bus was at a particular time. Due to the spatial accuracy of GPS (often between 5 – 10 metres), processes are usually required to match traces to a particular route. Cortes et al., 2011 employed spatial path rectification techniques to help assign GPS points along a road, where the specific route of the bus was not known. Their method assigned each GPS point to the nearest road, and then, using the full set of points from one device, calculated the most likely route based on the path taken by the majority of points. A similar method was also developed by Vanajakshi et al., 2009 who used intermittent GPS data to infer a bus's speed at various points throughout a day by matching the traces to the route and calculating the averaged time and distance difference between them.

For cities which employ an entry-only smartcard validation system, it is also necessary to try and infer the alighting location such that bus capacity profiles and full OD matrices can be developed. Initially Wang (2010), and latterly Gordon (2012) used Underground (train) and bus smartcard ticket data from London, UK which contained boarding and alighting information for all Underground trips, but only boarding information for bus trips. They developed a data analysis approach to help tackle this problem by filtering through the data and looking at individuals' smartcard records separately and in sequential order. Their process analysed data from a series of routes and aimed to find an alighting location based on a closest-stop rule, which assumes that a passenger alighted from their journey at the closest point to where they boarded their next journey. Testing this approach on a set of routes with good data, Wang (2010) presented destination inference successes between 57 – 66% (where a success is defined as the subsequent boarding stop being within 1km of a feasible alighting stop on the previous route). Advancing these techniques, Gordon (2012) reports a destination inference of 74% by introducing a daily symmetry rule, which assumes that the last trip of a day ends near the origin of the first trip of the day.

Trip-chaining and alighting inference techniques are also developed by Devillaine et al., 2012 and by Munizaga et al., 2014 who validate their method against exogenous survey data. Their approach assumes the closest-stop rule and centres on data that is directly comparable to their survey data.

2. CONTEXT

The data studied in this paper is collected by TTC (Tbilisi Transport Company) which runs all public buses and the metro system in the city of Tbilisi, Georgia. This system serves a population of over 1 million people, running buses on over 180 routes and an underground rail service with 2 lines and 23 stations (one running North to South, and the other East to West connecting at a central station). A network map and route planner, showing stop information for a particular route is shown in Figure 1. Since 2006, the city has employed a smartcard system, called 'Metromani' which requires passengers to validate the start of their journey upon boarding a bus, and entering the metro system. According to revenue data, TTC report that 80% of journeys are carried out using smartcards (with the rest being cash fare).

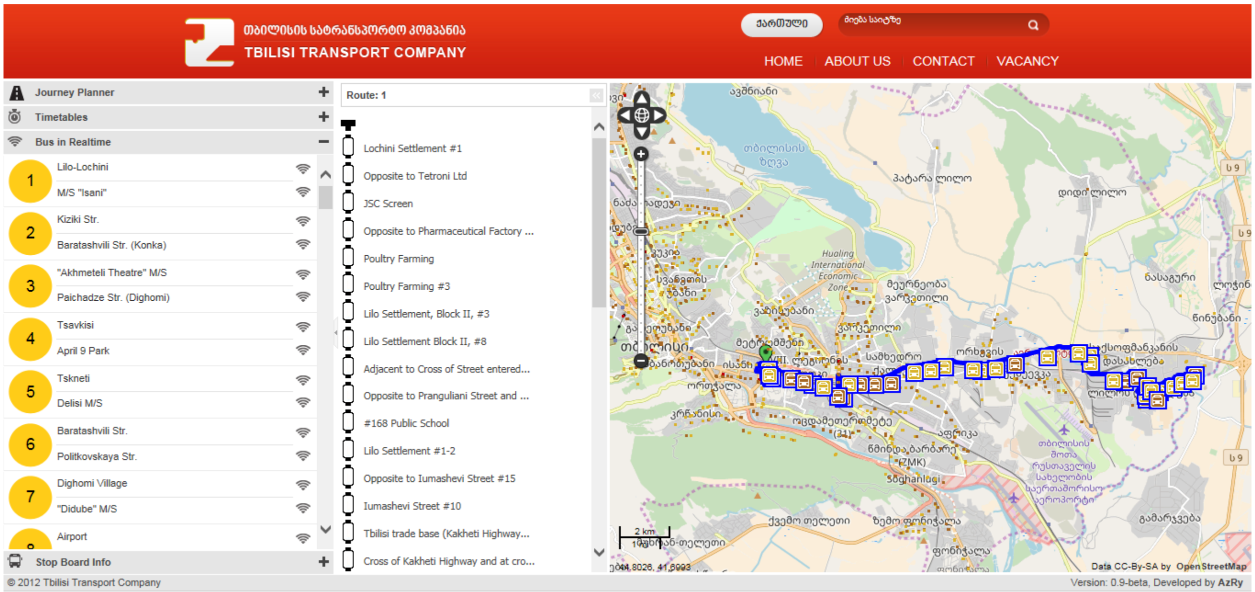

Figure 1 - Tbilisi Transport Company route planner (TTC, 2012)

For this study, TTC provided us with three key datasets: smartcard ticket data for bus and metro, GPS data which tracks the location of the city's buses, and complementary data including bus line and stop information in a GIS format and an ordered list of all stops for each route.

2.1 Smartcard data

For metro journeys, the smartcard system records the start time/date of an event and the boarding metro station. The resulting metro line used is not known, although in Tbilisi there are only 2 lines and 1 interchange station, and the alighting station is also unknown as validation is not required upon exit. Bus smartcard data also collects time/data information, but the boarding location is not specified, with only the bus number and the route number given. We were provided with 8 days of ticket data from 23rd May – 1st June 2016. Data for 28th and 29th was not provided as the brief was only to investigate weekday travel. Smartcard ticket data is collected at the weekends however, so in principle could be studied alongside weekday data. The 8-day dataset contains trip information for ~5.6 million (5,592,751) boarding events from ~600,000 (597,771) distinct devices. An example of the smartcard data, with the key fields shown is given in Table 1.

Table 1 - Unsorted bus smartcard data

Date Time	Device ID	Route number	Bus number	Fare	Balance on card
22/05/2016 00:01	1228356197	73	3086	0.5	4.7
22/05/2016 05:39	3811283834	140	2129	0	
22/05/2016 05:39	1835855498	88	2129	0	
22/05/2016 05:42	1004001755	9	1146	0.5	10.45
22/05/2016 05:46	1205627171	34	1145	0	

2.2 GPS data

For each of TTC's buses a GPS unit records the time and location when the bus is in service. We were also provided with a list of all of the bus IDs and the route they were serving on a given day. Data was provided from every public bus route throughout the city. The GPS data was recorded every 20 seconds and was accurate to ±10m.

2.3 Supplementary data

Further to the ticket and GPS data, we were also provided with a GIS layer containing all of the bus and metro routes, all of the stops, and a stop list for each route detailing which bus stops each service calls at and in what order.

3. METHOD

In this section, we detail the methodologies developed to determine where a smartcard record took place and, in turn, to which zone of the matrix a boarding or alighting event should be allocated. Since the bus smartcard data in Tbilisi is not supplemented with a stop location as part of the boarding record, it was necessary for us to infer the location of each event based on the GPS traces of the bus at the corresponding time. An example of GPS data for one route, plotted in GIS, is shown in Figure 2.

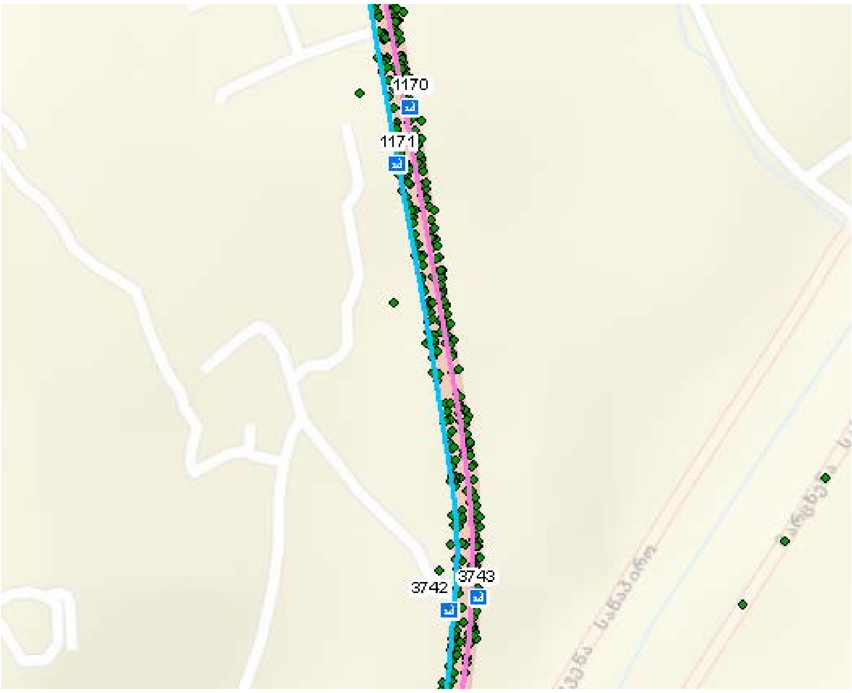

Figure 2 - GPS traces for 1 bus route on 1 day, between stops

To determine when a bus was at a particular stop, we employed a method similar to Cortez et al., 2011, who matched GPS traces to bus routes. For every GPS record we knew the bus's ID number and the route it was serving. Using GIS, we plotted these GPS points alongside the route layer and its associated bus stops. Around each stop we determined that the bus was potentially *at* the stop if there were one or more GPS traces within 50 metres. In addition, we needed to know the direction the bus was travelling so that we could assign it to the stop on the correct side of the road, and ignore nearby stops on the other side of the road (the GPS measurements alone were not accurate enough for us to determine this based purely on location). Since we knew the expected sequence of bus stops along each route, we developed an algorithm to determine the direction the bus was travelling based on the order of the stops it passed. For instance, in the timestamped sequence of

GPS traces, a point recorded at time *t* is near bus stop 4 (on the outbound route) *and* bus stop 42 (on the inbound route). Later in the sequence, a GPS point at time *t + 2 minutes* is located near bus stop 5 and bus stop 41. Since the sequence of stops is increasing in the outbound direction and decreasing in the inbound direction, we determine that the observed points were travelling in the outbound direction (from stop 4 to stop 5). This process is highlighted in Figure 3, where the GPS traces are plotted for one route alongside that route's associated stops in their sequential order.

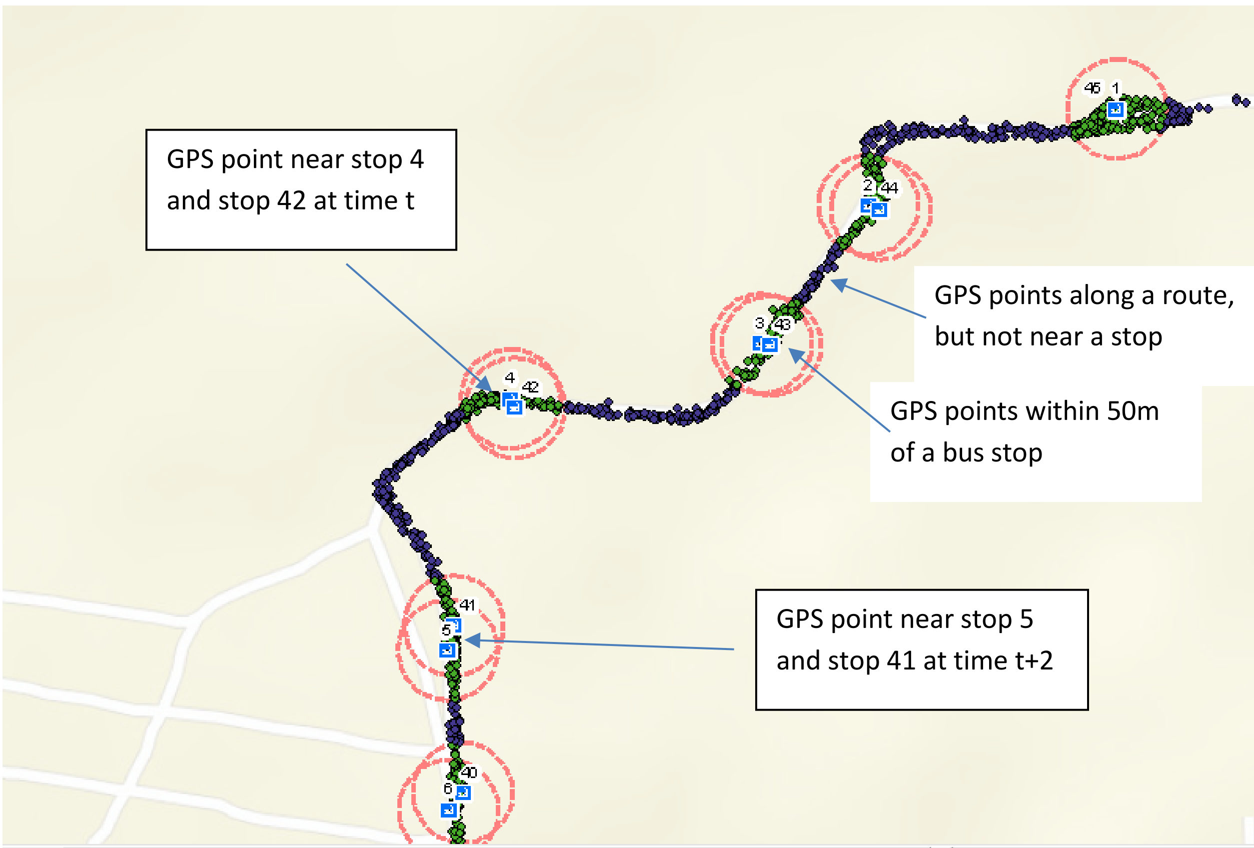

Figure 3 - GPS traces assigned to nearby bus stops

Using this approach, we determined the number of times each bus traversed its route during a day, and to which stop the nearby GPS traces should be assigned. For each leg of a bus's route we assigned the time it arrived at every stop using the first GPS point seen near that stop. Further, the last GPS point seen near the stop was determined to be the time it left the stop. The difference between these two timestamps was calculated as the bus's dwell time (within the constraint of the 50-metre buffer). The daily sequence of a bus's operation on a route, with its arrival and departure time at each stop, was output in a series of tables. An example of this output is shown in Table 2.

Table 2 - Bus operation table for route 14

Date	Vehicle ID	Vehicle Leg ID	Route #	Dir-ection	Stop Sequence	Current Stop #	Next Stop #	Distance along route (m)	Arrival Time	Dep-arture Time	Next Arrival Time	Average Speed km/h
25/05/2016	412	1	14	1	1	1188	1190	0	08:41:04	08:49:48	08:50:47	21.26
25/05/2016	412	1	14	1	2	1190	3316	348.5	08:50:47	08:51:58	08:52:29	34.11
25/05/2016	412	1	14	1	3	3316	1191	642.1	08:52:29	08:52:29	08:53:11	32.70
25/05/2016	412	1	14	1	4	1191	3314	1026.7	08:53:11	08:53:11	08:54:14	26.46
25/05/2016	412	1	14	1	5	3314	1192	1490.3	08:54:14	08:54:14	08:55:17	12.42
25/05/2016	412	1	14	1	6	1192	1193	1704.9	08:55:17	08:55:17	08:56:21	39.16
25/05/2016	412	1	14	1	7	1193	1194	2409.8	08:56:21	08:56:21	08:57:02	39.18
25/05/2016	412	1	14	1	8	1194	2506	2851.8	08:57:02	08:57:21	08:58:47	33.52
25/05/2016	412	1	14	1	9	2506	1196	3656.2	08:58:47	08:58:47	08:59:50	29.92
25/05/2016	412	1	14	1	10	1196	1197	4173.1	08:59:50	08:59:50	09:01:15	28.61
25/05/2016	412	1	14	1	11	1197	1198	4852.9	09:01:15	09:01:15	09:01:57	18.55
25/05/2016	412	1	14	1	12	1198	1199	5066.6	09:01:57	09:01:57	09:03:21	15.92
25/05/2016	412	1	14	1	13	1199	2497	5441.0	09:03:21	09:03:21	09:04:03	28.93
25/05/2016	412	1	14	1	14	2497	1200	5774.2	09:04:03	09:04:03	09:04:45	40.21
25/05/2016	412	1	14	1	15	1200	3697	6247.1	09:04:45	09:05:06	09:06:50	28.94

3.1. Assigning a bus stop to a smartcard boarding record

Using our calculated tables of bus routes and timestamped locations of the bus at stops, we attempted to assign each smartcard boarding location to a bus stop on that route using a matching process. This match was based on the bus route, the bus vehicle number, and the timestamp. For instance, in a bus route sequence, if the bus is recorded arriving at stop A at 17:20:10 and arriving later at Stop B at 17:24:50, all matching smartcard records between these times are assigned to stop A. This assumes that a smartcard validation record can take place at any time *after* the bus arrives at stop A but *before* it arrives at the next stop B. This technique is highlighted diagrammatically in Figure 4. In Tbilisi, tickets must be validated using one of the self-validation ticket machines upon boarding the bus, thus it is assumed that most smartcard transactions are recorded before the bus reaches the next stop.

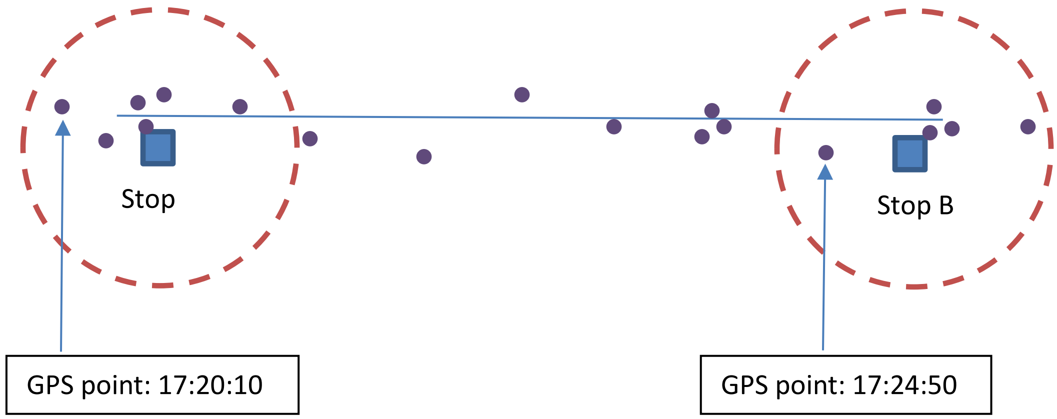

Figure 4 - Assigning smartcard boarding events to a bus stop

Based on the methods described above, smartcard boarding events were assigned to a bus stop based on a matched timestamp, bus number, and the bus's route. In total, more than 97% of the smartcard records were successfully matched to a boarding location. Upon closer inspection of the ~3% of unmatched records, most these took place when buses were in the depot (presumably by TTC employees) and often in large batches in a short timeframe. With these records dismissed a 99% match rate was achieved, the remaining unmatched records a result of a presumed error in the bus ID → bus route lookup table (i.e. a bus was logged as running on route 24, but was actually operating on another route).

3.2 Inferring Alighting stops from smartcard boarding records

In addition to the bus smartcard data, which requires the assignment of boarding location, TTC also collect and maintain smartcard data for the city's metro. This dataset names the metro station where a boarding event took place, and hence this location is already known. Because both bus and metro are operated under a unifying smartcard system, the same device ID can be observed in both datasets. Thus, when combined and sorted sequentially, each device's trips can be identified by time and mode. An example of this combined data is shown in Table 3.

Table 3 - Combined smartcard data with inferred boarding locations

Date Time	Device ID	Route number	Boarding location (Stop ID)	Direction	Stop sequence number
23/05/2016 06:01	762411082	73	3086	I	4
23/05/2016 06:39	762411082	M	144	Unknown	-
24/05/2016 05:39	762411082	73	3086	I	4
24/05/2016 06:12	762411082	9	1146	I	7
24/05/2016 18:46	762411082	34	1145	O	36
24/05/2016 20:00	762411082	M	258	Unknown	-

So that full OD matrices can be generated, it is necessary to infer an alighting record for each boarding event. Similar to Gordon (2012), we have developed an algorithmic approach to assign a trip's alighting location based on the location of the next observed boarding using the closest-stop rule. This is achieved by searching for all the bus or metro stops within 1km of the next boarding event. From these nearby stops, it is assumed that the passenger alighted at a stop on the previous trip's route. If more than one stop is nearby, the closest to the subsequent boarding is chosen. This process is shown in Figure 5 where a series of boarding events are known for smartcard *x*. For an event recorded at time t, and assumed to be boarding at stop 5 on route A, we know that the passenger alights at a later stop along route A. In passenger x's smartcard transaction history, a subsequent trip is recorded at stop 18 on route B. Since stops 8 and 9 are near (within 1km) of stop 18, it is assumed that passenger x alighted at the nearest of these stops; in this case stop 9.

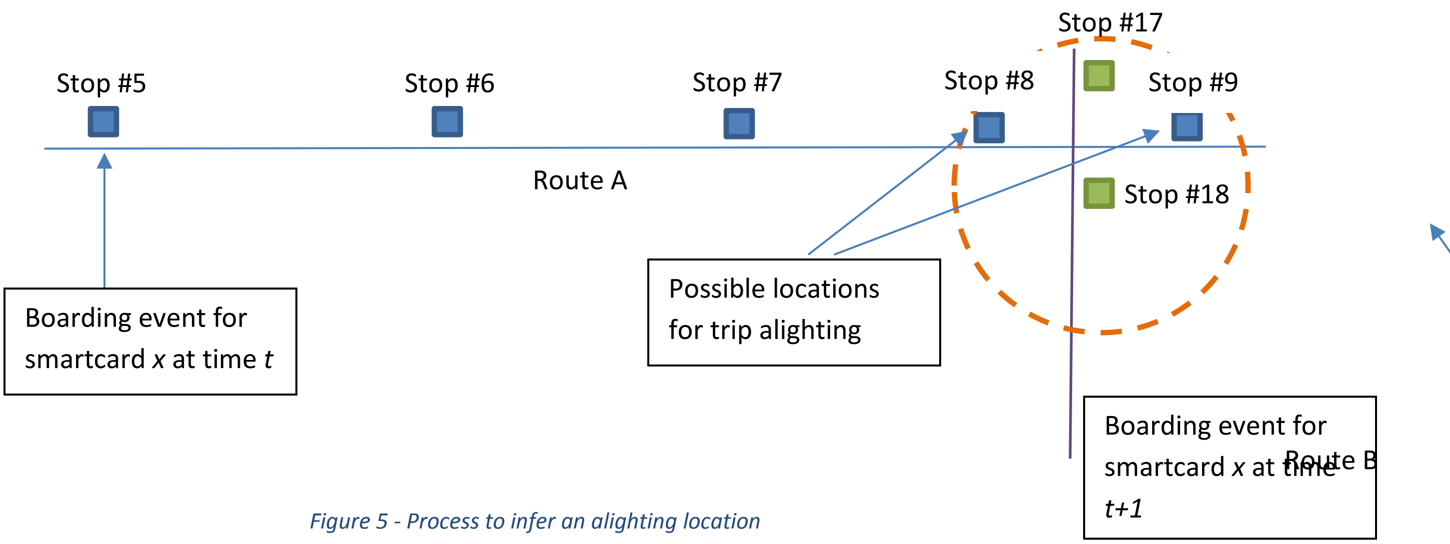

Figure 5 - Process to infer an alighting location

This process is scripted algorithmically and run through all of the trips for each smartcard device. For every boarding event, an alighting record is created and designated as either an alight, a transfer, or location unknown. The alighting is deemed to be unknown if the boarding event is not proceeded by another on the same or next day, or if no matching stops are returned in the algorithm process (i.e. their next trip did not start near where their last trip could have ended). In these cases, it is assumed that the passenger travelled intermediately via some other, non-public transport, mode. Transfers are assumed to have occurred if a boarding is recorded less than 30 minutes after the previous boarding. In addition, records which occur within 30 seconds of the next one are discarded since multiple duplicated records were observed in bunches within the dataset (assumed to be a recording error from the ticket validation machine). This is also based on the fact that Metromani cards are not transferrable, and so cannot legitimately be scanned for an accompanying passenger. For all other records that produce a stop match, it is assumed that the passenger has completed their journey and alighted at their destination. This procedure is shown schematically in Figure 6.

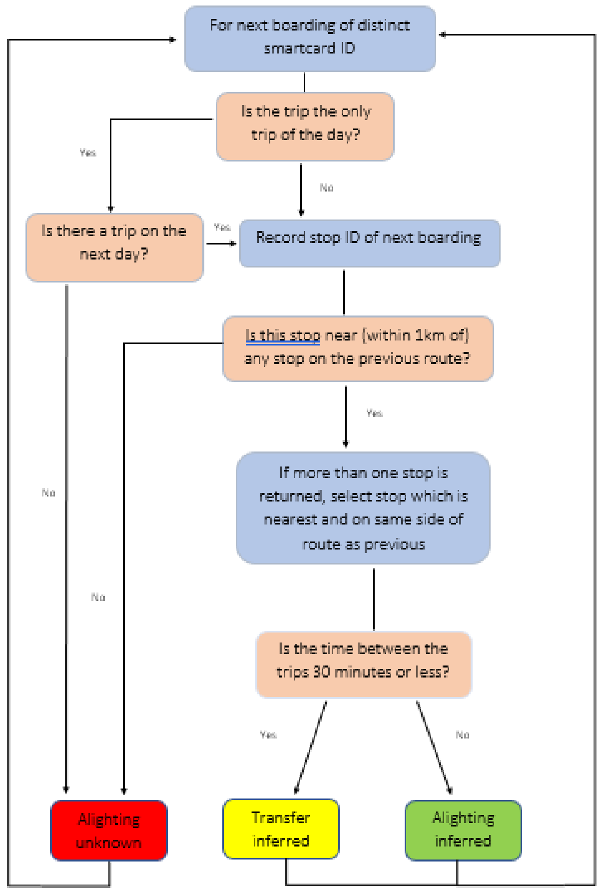

Figure 6 - Schematic of algorithm to run through all smartcard data and assign alighting locations

4. RESULTS

GPS data was provided for each of the city's buses over an 8-day period, totalling over 11 million individual GPS traces. Coupled with GIS layers of the bus routes and their associated stops, each bus's service could be plotted, analysed, and matched to corresponding stops along its route to determine when it arrived and departed from each one. Smartcard data, available over the same 8 days, contained ~5.6 million boarding events from ~600,000 passengers. Based on the assumption that smartcards are validated at any point after a passenger boards at one stop, but before the bus arrives at the next, 97% of transactions were assigned a boarding location (99% once transactions at bus depots were removed). Each boarding event was assigned an associated alighting designation, either the destination of a trip, a transfer, or an unknown location if a match could not be achieved. From the 5.6 million smartcard records, ~4.6 million records were assigned a distinct boarding location (the majority of those records discarded due to multiple transaction records within a short time frame). Of the 4.6 million boarding locations, 13% (~520,000) were not proceeded by another trip on the same, or next, day. For the remaining trips, using the method described in section 3 which inferred a passenger's alighting location based on their next boarding location, ~2.9 million records were assigned an alighting location that would have been feasible assuming a walk of <1km. Approximately 800,000, or 27%, of the alighting records are defined as transfers, since a subsequent record was noted within half an hour of the previous boarding. These results are summarised in Table 4, which presents data from all 8 days (23rd May – 27th May and 30th May – 1st June inclusive), and also with counts from the 27th May and 1st June removed. In these cases, since a subsequent day of data was not available, the last boarding record of the day would not have been assigned an alighting location because no later trips were known for that passenger. Ignoring these days improves the alighting match success from 71% to 73%.

Table 4 - Boarding and alighting inference results

	All days	6 days (excluding days without subsequent day)
Total smartcard transactions	5,592,751	4,128,748
Discarded 'duplicate' boarding records	17.1%	16.8%
Single record day and subsequent day	9.3%	9.4%
Distinct boarding records	73.6%	73.8%
Total	100%	100%
From distinct boarding records:		
Unmatched alighting locations	29.4%	27.3%
Alighting location found	70.6%	72.7%
Of which defined as a transfer	27.1%	26.3%

Using records where a boarding and alighting location has been successfully inferred, it is possible to assign the trips to an OD matrix. Since smartcard data and our methodology only deals with public transport patronage, strictly it is only possible to generate an access-egress matrix showing where passengers entered and departed the public transport network. Without further information on subsequent modes (such as walking), the assumption is that a boarding bus stop represents a passenger's origin, and the alighting location represents their destination. Further to this, every bus stop location was assigned to a zone; the zoning system having been developed as part of additional modelling work for Tbilisi City Council. This zoning system was created to allow for integration with a 2014 household travel survey and 2011 Census data. In total, there are 345 zones across the city, with the smallest zones concentrated in the corridor which follow the city's main river, and the metro line which extends to the western university district. Using this zoning system, each of the matched boarding and alighting sequences are assigned to the zonal matrix. Because the smartcard records are timestamped, the matrix can be filtered to only show trips at certain times of day, as in Figure 7 which shows sum of all alighting trips per zone in the AM peak (07:30 – 09:30). Although not represented in this figure, it would also be possible to factor the matrix to account for missing trips and cash-fare passengers (if it is assumed that the distribution of these trips is the same).

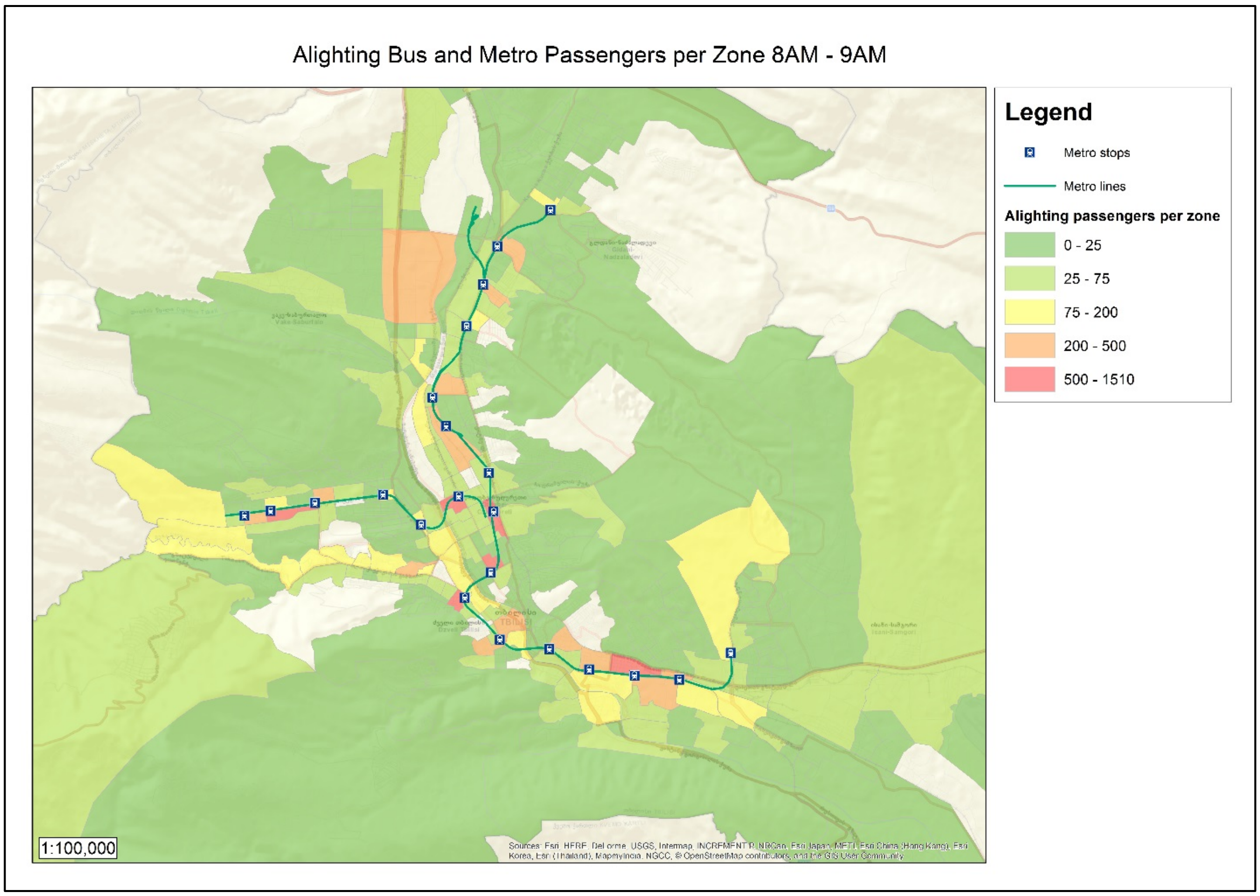

Figure 7 - Inferred alighting of passengers per zone between 8AM - 9AM

5. DISCUSSION AND CONCLUSION

In this paper, we describe a series of processes to develop an OD matrix for public transport use in the city of Tbilisi, Georgia. The data analysed is generated via a smartcard system that is used for 80% of public bus and metro journeys in the city. Unlike some datasets of this kind, which are stamped with a boarding stop or fare stage, the bus smartcard data in Tbilisi is only provided with a route number, bus ID, and timestamp. Therefore, using GPS data recorded from every bus, a method was developed to determine when buses arrived at each stop along their route. Although not explored in this paper, the bus route tables produced in this process could be used to understand detailed level of service indicators such as bus operating speeds throughout the day, dwell times, and route operating frequency. Combined with smartcard transactions, assigned to each bus, it would also be possible to infer capacity profiles based on the developed boarding and alighting information. This could provide the public transport authority with in-depth information about which of their services could benefit from improvement, whether that be frequency variation or routeing adjustment.

The alighting inference process employed in this paper assumes a closest-stop rule, that is, where a passenger gets off their bus or metro at the point closest to where they next board a service. Whilst this assumption is likely to be true in a lot of cases, clearly if any intermediate mode (including walking) is used, the inferred alighting stop may not be found, or will be mis-assigned to an incorrect stop. In Tbilisi, although TTC operate all public transport routes, semi-regulated marshruktas (mini-buses) are also in operation and serve many of the same routes as TTC, albeit not to defined timetables. Thus, an inbound journey could be taken in the morning using a smartcard , and the outbound journey later in the day via a marshrukta for a cheaper price. Validation surveys could potentially be used to understand how many public transport trips are proceeded by another mode, but even this is likely to underestimate the number of 'unplanned' trips where someone uses a marshrukta because it arrives at the stop before a scheduled bus. Similar to Munizaga and Palma, 2012 but in contrast to Gordon (2012), we apply the closest stop rule to trips starting on subsequent days as opposed to employing a daily symmetry rule. Validation surveys in this case would also be useful to determine which approach is more accurate, or whether a combination of correlated results could be used.

Our approach was developed and scripted in Python to facilitate integration with GIS processes and to handle data across the whole city, and to use multiple weeks' worth of data. This has enabled us to process over 11 million GPS traces and 5 million smartcard transactions and could be expanded to handle data over a longer period, or shortly after it is collected, providing the bus company with consistently up-to-date information. In addition, since these methods make use of already collected, passive data, the developed matrix can be produced at a much lower cost than traditional passenger surveys and interviews. While providing large cost savings, this can also mean that the use of interview techniques becomes more targeted and is used more for validation rather than as a primary means of data collection. For instance, although the 80% patronage of the smartcard system provides a good sample size, the travel patterns of the remaining 20% are unknown. Without further information, it may be sufficient to assume travel patterns are the same, but this is unlikely to be the case with cash payers probably less frequent bus users (i.e. not commuters) or transient passengers visiting from another location. In this case, it may be worthwhile conducting surveys that specifically target cash payers. Similarly, because our alighting method relies on, preferably repeated, patterns it is likely that the trip behaviour of unmatched records is different to those where a match was found.

The methods developed in this paper are applicable in any location where smartcard ticketing and location information is prevalent (be that via bus GPS loggers or location stamped ticket records). With systems being rolled out, or already well-established, in many European cities analyses of this type can provide a comprehensive understanding of public transport usage, capacity, and operation. For more in-depth studies involving transport modelling, the production of OD matrices is possible. Compared to manual survey derived matrices, these can therefore be developed at a lower cost, yet provide more representative assessments of a city's public transport network, enabling transport authorities to make better informed improvements in their city.

BIBLIOGRAPHY

Cortés, C.E., Gibson, J., Gschwender, A., Munizaga, M., Zúñiga, M., 2011. Commercial bus speed diagnosis based on GPS-monitored data. Transportation Research Part C: Emerging Technologies 19, 695–707. doi:10.1016/j.trc.2010.12.008

Devillaine, F., Munizaga, M., Trépanier, M., 2012. Detection of Activities of Public Transport Users by Analyzing Smart Card Data. Transportation Research Record: Journal of the Transportation Research Board 2276, 48–55. doi:10.3141/2276-06

Gordon, J.B. (Jason B., 2012. Intermodal passenger flows on London's public transport network : automated inference of full passenger journeys using fare-transaction and vehicle-location data (Thesis). Massachusetts Institute of Technology.

Jang, W., 2010. Travel Time and Transfer Analysis Using Transit Smart Card Data. Transportation Research Record: Journal of the Transportation Research Board 2144, 142–149. doi:10.3141/2144-16

Munizaga, M., Devillaine, F., Navarrete, C., Silva, D., 2014. Validating travel behavior estimated from smartcard data. Transportation Research Part C: Emerging Technologies 44, 70–79. doi:10.1016/j.trc.2014.03.008

Munizaga, M.A., Palma, C., 2012. Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. Transportation Research Part C: Emerging Technologies 24, 9–18. doi:10.1016/j.trc.2012.01.007

TTC, 2012. Tbilisi Transport Company [WWW Document]. URL http://ttc.com.ge/?lang_id=ENG (accessed 9.5.17).

Vanajakshi, L., Subramanian, S.C., Sivanandan, R., 2009. Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses. IET Intelligent Transport Systems 3, 1–9. doi:10.1049/iet-its:20080013

Wang, W., 2010. Bus passenger origin-destination estimation and travel behavior using automated data collection systems in London, UK (Thesis). Massachusetts Institute of Technology.