A Method for Determining Alighting Locations of Bus Passengers

I. Gökaşar[1], A. Fidanoğlu[2]

[1]Department of Civil Engineering, Boğaziçi University, Istanbul, Turkey
[2]Department of Transportation and Logistics, Istanbul University, Istanbul, Turkey

Abstract

Origin and destination (OD) matrices are considered as a key point for the transportation and city planners. With the development of advance transportation technologies installed in the public transit, transportation agencies can make necessary adjustments and implement dynamic measures much easier. Automated Data Collection (ADC) systems installed to the vehicles, which are more efficient and economic than other conventional data collection methods, are especially helpful to store data regarding passengers and their trips. Although the origin of a passenger be deduced much easier, the challenging part is to detect the passengers' destination properly, especially for the bus passengers due to the lack of data when the passenger exits from the system. The objective of this study is to provide a method for inferring the alighting locations of the bus passengers in order to construct more reliable OD matrices. This method is tested by conducting a case study for a single bus route in Istanbul. The overall results show that by using the proposed algorithm over 65% of all recorded trips' alighting locations were derived. By excluding instances where boarding information is missing or the same card in repetitively used, the success rate reaches up to 70%. For future studies it is suggested that unique characteristics of the transportation systems must be delved into in order to eliminate misleading outputs of any proposed algorithm.

Keywords: *Intelligent Transportation Systems, Origin and Destination Matrices, ADC Sytems*

1 Introduction

Automated Data Collection systems become widely used in public transportation systems with the implementation of the technological innovation over the past decades (Cui 2006). ADC systems have become popular as they provide effective and cheaper alternative to the conventional data collection methods. Commonly used examples of ADC systems are namely Automated Fare Collection (AFC), Automated Vehicle Location (AVL) and Automated Passenger Counting (APC) systems.

Production of OD matrices by using ADC data has several advantages over traditional surveys as:

- Significant reduction in cost of obtaining OD matrices,
- Obtaining individual trip information of passengers by the help of passenger smartcard,
- Easiness to update the data and run the process more frequently.
- Providing continuous trip information of passengers in transit system in a larger scale.
- Conveniently utilizing in more comprehensive studies.

Barry et al. (2002) proposed a methodology that is used to estimate origin and destination of the passengers by using MetroCard information in New York City. He applied a set of straightforward methods to each set of MetroCard to assign a destination for every origin station. He validated his assumptions at very high rate with the travel diary information stored by the New York Metropolitan Transportation Council.

Cui (2006) aimed in his study to create a model to estimate a network level bus passenger origin OD matrix. He used the ADC data from Chicago Transit Authority (CTA) to make OD estimation for the bus network in Chicago. For the inferences of origins and destinations, he applied the methodology at single route level and network level. His study for transit rides in public transportation of Chicago was based on the trip chaining OD estimation method. He achieved to infer the high portion of the origins and destinations in his study.

Zhao et al. (2007) developed a method to estimate the origin and destination locations of the rail passenger trips with the automated fare collection (AFC) data supplied by the Chicago Transit Authority (CTA). During his study, he also generated a software to assist the application of his proposed algorithms. He suggested the integration of the automated fare collection data of CTA which stores the trip transactions and the automated vehicle location data of CTA which records the vehicle location to infer the boarding station ID of the passenger. In his study, both the rail to rail trip sequence and rail to bus cases are studied with the help of integration of AVL and AFC data.

Wang (2010), made OD estimations with the case studies for several routes in London. In the inference of boarding location, similar to Zhao (2007), he combined the iBus data of the buses and the Oyster data of the passengers. These two data consist of the AVL and AFC data, respectively. He also used the similar algorithm with previous studies to estimate the alighting locations in London. After the inference of alighting and boarding location of the passengers using the studied routes, he further analyzed the interchange times. He questioned the appliance of a fixed temporal threshold for the consecutive trips to be identified as linked trips. He stated importance of in-vehicle travel time and route headways in determining of interchange time and study the interchange times of the trips in his London case study.

With the new technologies used in our daily lives, further opportunities become feasible for estimating the OD matrices. Iqbal et al. (2014) used mobile phone call detail records and some traffic count data to construct the OD matrices. They use mobile phone data of over 2.5 million of users in Dhaka, Bangladesh. Travel characteristics of the passengers are derived from the mobile phone data while the traffic scenario is constructed upon the traffic count data. Alexander et al. (2015) also used phone call detail records to estimate OD matrices for the passengers live in Boston metropolitan. They validated their technique through local and national surveys. Jiang et al. (2013) and Schneider et al. (2013) showed that daily travels of the passengers deducted from the mobile phone data give similar results with the household surveys at individual level. Bonnel et al. (2015) developed OD matrices upon the dataset produced by mobile phone operators. The dataset includes communication events, handover and location area up-date information of mobile phone users in Paris Region. From the spatio-temporal trajectory of the users, travels of the passengers are derived.

2 Methodology

In this analysis, origin and destination matrices are constructed for the selected bus route only. The boarding and alighting locations of the passenger in the studied route are inferred in this study. For global origin and destination matrix the proposed algorithm should be used for every single trip of that passenger.

The municipality records the boarding locations in AFC data; therefore, the boarding locations of the particular passenger were derived from AFC data only. Cui (2006), Zhao (2007) and Wang (2010) made the origin inference for the passengers by integrating the AVL data and the related AFC data recorded in the systems in their studies. The AVL data of the buses operate in the studied route could be provided only for a small portion of trips. Therefore, AVL data is used in this study only when a direction error is found for the inferences made. One example is illustrated in Table 1. For the specific bus (C-1722) at the given time, the boarding location is recorded as Kısıklı in the AFC dataset. On the other hand, when the relevant recorded coordinates in AVL data is located in the map it is found that the closest stop for this coordinates is Dostluk Parkı bus stop.

Table 1. Recorded Boarding Location in AFC Data and Calculated Boarding Location from AVL Data.

Date	Stop ID	Bus No	Stop Name	Stop Name from GPS of Bus
15.09.2014 18:40:23	A0291A	C-1722	KISIKLI	DOSTLUK PARKI
15.09.2014 18:40:26	A0291A	C-1722	KISIKLI	DOSTLUK PARKI
15.09.2014 18:40:28	A0291A	C-1722	KISIKLI	DOSTLUK PARKI

Zhao (2007), Cui (2006), Trepanier *et al.* (2007) and Wang (2010) all made the same assumptions for the inference of destination methodology as:

- Passengers don't use private transportation modes between the recorded trips.
- The distance between alighting location of the previous trip and the boarding location of the next trip cannot exceed predetermined level for these consecutive trips to be considered as transit trips.
- Passengers return to boarding location of their first trip with their last trip on that day.

First two assumptions are made in the interchange method which is used to infer the alighting location of the trips with interchanges to other routes. With the help of these two assumptions, boarding location of the next trip is considered as the alighting locations for the previous trips. This method is called as Interchange or Next Trip

Method. On the other hand, for the last trips of the days, the third assumption is also taken into consideration and the first trip of the day is taken as the next trip of the studied trip. By this way, passengers are assumed to come back to the location where they start their first trip on the day. For the last trips of the day, other assumptions used in the interchanges above are also still valid. This method is named as Last Trip Method.

Wang (2010) showed the process for destination inference in his study with Figure 1.

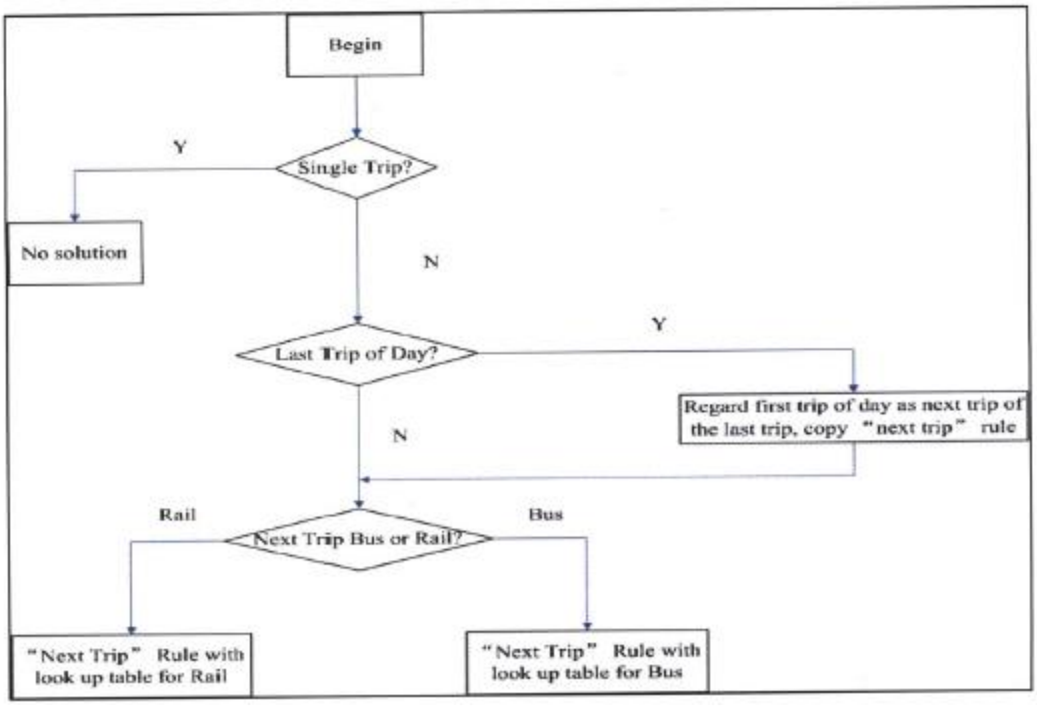


Figure 1. Process for Destination Inference (Wang 2010).

In this study; however, the methods used in the previous studies for both interchanges and last trips were modified in the proposed methodology. The process carried out in this study for the inference of alighting locations can be summarized as:

- Next trips of the studied trips are found.
- In terms of their next trips, trips were classified into groups.
- For the trips which have no other trip on the analyzed days, no alighting location is inferred.
- For the trips which have a next trip in 2 hours, boarding location of the next trip is checked in terms of the distance to the stops of studied route. If distance to the closest stop of the studied route is below some specified limit or the next trip is made at one of the stops in the route of studied trip, then these stops are considered as the alighting location. In this group, destinations of the main possible routes which is assumed to be missing in the ADC data are also taken into account and the closest stop of the studied route to the origins of these trips are taken as the alighting location if the next trip is made at the location near the destination of these main routes. For this cluster, if at the end of alighting inference, direction error occurs then recorded direction of the studied trip is changed. These additional two assumptions are made for the trips in this group because of the short time interval between the consecutive trips.
- For the interchanges made beyond 2 hours, the same procedure with the previous item is followed. However, for the trips in this group the missing trip assumption and the correction in direction are not made. This is because the time interval between the trips are relatively long and the passenger can reach to his/her recorded next boarding by using different routes.
- For the interchanges made in the same buses are studied in a different group to detect the repetitive use of the same smartcard. If the passenger makes his/her next trip in 60 minutes at the same direction and bus then it is assumed that the cardholder use his/her smartcard for another passenger. Hence for these records no alighting location is inferred.

• For the trips which are the last trips of the day, boarding location of the first trip is checked whether it is one of the studied route's stations or close enough to the stations of the studied trips. If it is below limits, as in the previous studies also, the boarding location of the first trip of the day is taken as alighting location of the last trip.
• For the trips which are last trips of the day and no result found with the methods described in the previous item another method is proposed in this study. Since the next trip of all trips are found at the beginning of the study, next trips of these trips are analyzed in terms of their day. If the next trip is made on a day close enough to the day of studied trip then the same procedure used for trips in interchange after 2 hours cluster is applied to infer the alighting location. By this method, many of the trips which are not studied in the methods suggested in the previous studies can be analyzed further. For example, as it is seen in the process for destination inference in Wang's (2010) study for single trips which have no other trips on that day, no result can be inferred. However, with the help of proposed method, further analysis can be made for these trips also. But, it should be noted that the number of days between the day of the studied trip and next weekday is taken as the limits for the difference between the day of studied trip and the next trip. The next trip day must satisfy this rule.

3 Istanbul Case Study

3.1. Characteristics of 11 L (Bulgurlu-Uskudar) Route

11 L (Üsküdar-Bulgurlu) bus route which runs between Üsküdar and Bulgurlu in the Anatolian part of İstanbul is selected for the analysis. Total length of the route is 9-10 km for each direction with 12 minutes headways during early mornings and 15 minutes daytime headways. 11 L route is analyzed in this study because of the following reasons:

• This route runs between the location, mainly consisting of residences and the location which is the one of the most commonly visited places in İstanbul. Therefore, in this route not only the commuters but also the irregular users of this route are expected to be recorded.
• The route has intersections with Metrobüs BRT line, Marmaray subway and the ferries runs in the Bosphorus between the Anatolian and the European part of the İstanbul. These are commonly used public transportation systems in İstanbul. By analyzing the 11 L route, the interchanges to these main routes can be studied.
• Since Üsküdar is a location which has many historical places and shopping centers, it is quite possible to make comments about the interchanges in the location with these features after the analysis of 11 L route.

Buses in 11 L route start their trips in Esatpaşa to Üsküdar direction. After reaching the Üsküdar they turn back to Bulgurlu direction and finish their trip in Esatpaşa. They make a ring trip which means they don't stop and wait in Üsküdar stations. Kültür Merkezi bus stop is not included in the bus stop list for Üsküdar direction. However, it is known that Kültür Merkezi bus stop is located in the both direction of 11 L and the 11 L buses stops at this stop in the trips to Üsküdar direction. For this reason in this study inference of Kültür Merkezi bus stop as alighting location in Üsküdar direction is taken as a valid inference. However, since there is no boarding records for Kültür Merkezi in ADC data to Üsküdar direction, boardings made at this stop couldn't be inferred.

3.2.ADC Data Analysis

In this study, ADC data of 11 L for September 15-23, 2014 is used. A sample from data is shown in Table 2. ADC data mainly contains;

• Date; which is the date and time in second precision.
• Route: is the name of the line that ADC data records.
• ID; is the unique ID number for each smartcard.
• Type of Ticket; is the type of smartcard given to the passenger according to his/her status in terms of age, education, etc.
• Stop ID; is the unique ID of each bus stops for each direction.
• Gate No; is the unique ID of bus travels in that route. Since in ferries, Metrobüs and subways fares are collected in stations, "Gate No" refers to stations in these transportation systems.
• Name of Stop; is the name of each bus stop. There is no "Stop ID and Name of Stop" information for ferries, Metrobüs and subways because of the explained reason.

Table 2. An Example ADC Data.

English	Turkish	Data
Date	Tarih	15.09.2014 06:01:33
Operation Group	Operatorgrubu	Özel Halk Otobüsü
Operator	Operator	Otobüs A.Ş.
Route	Hat	11L
ID	medyaserino	046**************
Type of Ticket	BiletTipi	İndirimli Bilet / Discount Ticket
Type of Fare	GecisTipi	Kontürlü / With Credit
Type of Interchange	AktarmaTipi	Normal / Normal
Stop ID	DurakId	A0424B
Gate No	KapiNo	C-1709
Name of Stop	NoktaAdi	AÇAN SOKAK

ADC data of 11 L for September 15-23, 2014 have 13.530 records while ADC data which contains the information about 11 L and other trips have 69.195 records for that period. Both datasets are used for this study.

3.3.Inference of Alighting Locations

To start the analysis, first very next trip of every single 11 L trips is extracted from "All Trips" dataset. The information about the date, stop name, stop ID, route name and gate no are taken from the dataset and named as "Next Trip" of the relevant 11 L trips. After extracting all next trips, 11 L ADC dataset is clustered to perform the analysis with different assumptions.

Table 3. Type of 11 L Trips and Their Percentages.

Type of 11 L Trips	Number of Trips	%
Interchange in that day	6849	51%
Last trip of the day	5067	37%
Single trip in that day	1533	11%
Single trip in all days	81	1%
Total	13530	100%

4 Proposed Algorithm

Methodology used in this study is based on the several assumptions described in the previous studies like Cui (2006) and Wang (2010). Some additional assumptions are made in this study to improve the previous methods. The proposed algorithm is deemed to give accurate results in the inference of alighting locations of the trips with the consideration of key points stated below:

- In the search for closeness of the next boarding in the interchanges made within 2 hours, look up table for the closest stops should be generated with the consideration of possible missing trips.
- Threshold for the maximum walking distance should be determined for the studied routes before the analysis.
- For the steps of the algorithm introduced to detect the repetitive use of the same smartcard, temporal threshold (1 hour in this study) should be determined according to the minimum duration of the route for a single round.

The proposed algorithm starts with the initial step that checks whether there is any other trip of the passengers being evaluated. If there is not any other trip of those passengers on studied days, then no inference can be made.

For repetitive use of the same smartcard, the algorithm checks the first next trip to determine whether it is a repetitive use of the same smartcard. If it is so, then the algorithm goes back in the process and finds the next record of that passenger. This goes until finding the real next trip that passenger made after the studied trip. The proposed algorithm also checks in its second step whether the studied trip is actually the repetitive use of the same smartcard. For this, algorithm asks the previous trip of the studied trip with the same conditions. If the previous trip of the studied trip made under the given conditions, then it is concluded that the studied trip is the repetitive use of the same smartcard and no inference is made for the studied trip. This step is introduced to eliminate the misleading alighting inferences for the passengers who used other passengers' smartcard.

After pinning down the repetitive uses of the same smartcard, the algorithm checks whether it is the last trip of the day or there is any trip after that. According to the type of the studied trip algorithm goes whether last trip part or interchange part.

Both in the last trip and interchange inferences, the algorithm first checks the closeness of the inferred alighting location. If it is close enough to the studied route then the algorithm proceeds and checks whether the boarding stop and the inferred alighting stop are the same or not. If they are the same, there is obviously no result. If not, the algorithm further checks for the direction error. The direction error means that inferred alighting location is not on the route of recorded direction.

In the last trip section, the proposed algorithm goes to the step where the day of next trip is checked for the trips that no inference could be made from the first boarding of the passenger on that day. After that, the same procedure as described in the previous paragraph is followed.

The interchange portion of the proposed algorithm is different from the previous algorithms in the literature in that it changes the recorded direction of the studied trip if a direction error occurs for the trips which have an interchange in 2 hours.

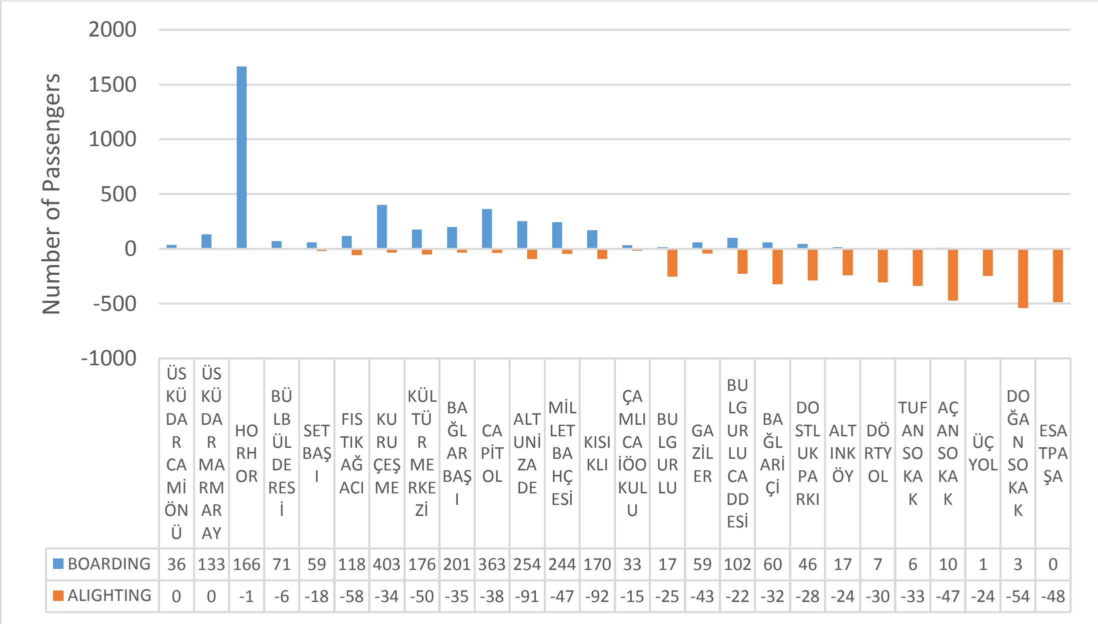
5 Inferred Origin & Destinations

With the modifications to the previous algorithms proposed in this study, the inference process gives very fulfilling results as shown in Table 4. The total number of studied 11 L trips is 13,304. This number reaches to 13,530 with the inclusion of the examined 11 L trips of the passengers that have no other trips in the studied days and the 11 L trips which have no boarding information in their next trip. Total of 8,812 out of 13,304 11 L trips, the alighting locations are inferred. For 65.13% of all 11 L trips ODs are successfully inferred. This rate increases to 66.24% when the 11 L trips, not studied, are excluded.

Table 4. Overall Inference Results.

Reason of Error	Total	Percentage
Direction Error	1,229	9.24%
Missing STOP ID In 11 L Trip	624	4.69%
Same Stop	440	3.31%
Interchange Stop Is Not Close	1,329	9.99%
Repeated Records	507	3.81%
First Stop Of The Day Is Not Close	29	0.22%
Day Of Next Trip Is Not Close	334	2.51%
Results		
Inference OF Alighting	8,812	Success Rate
Total Studied 11 L Trips	13,304	66.24%
Total 11 L Trips	13,530	65.13%

As an example of the outputs of the study, Figure 2 is given, where the number of boardings and alightings in every stop in Bulgurlu Direction are shown. The figure shows the aggregated numbers of passengers in every stop. However, from the proposed algorithm, boarding and alighting location of each passenger can be derived at individual level.



	ÜSKÜDAR CAMİÖNÜ	ÜSKÜDAR MARMARAY	HORHOR	BÜLBÜLDERESİ	SETBAŞI	FISTIKAĞACI	KURUÇEŞME	KÜLTÜR MERKEZİ	BAĞLARBAŞI	CAPİTOL	ALTUNİZADE	MİLLET BAHÇESİ	KISIKLI	ÇAMLICA İÖO KULU	BULGURLU	GAZİLER	BULGURLU CADDESİ	BAĞLARİÇİ	DOSTLUK PARKI	ALTINKÖY	DÖRTYOL	TUFAN SOKAK	AÇAN SOKAK	ÜÇYOL	DOĞAN SOKAK	ESATPAŞA
BOARDING	36	133	166	71	59	118	403	176	201	363	254	244	170	33	17	59	102	60	46	17	7	6	10	1	3	0
ALIGHTING	0	0	-1	-6	-18	-58	-34	-50	-35	-38	-91	-47	-92	-15	-25	-43	-22	-32	-28	-24	-30	-33	-47	-24	-54	-48

Figure 2. The Number of Boarding and Alighting at Each Stop of 11 L Route in Bulgurlu Direction

5.1.Inference Rates of Methods

In this study, mainly two clusters are introduced, namely 11 L trips with interchanges in that day and the 11 L trips as the last trips of that day. Inference rate of the interchange cluster is found to be much higher than the last trip cluster, because it is easy to track the passenger in interchange cluster with his/her next trip. On the other hand, the last trip rule applied at the first step of last trip method is based on the assumption that a passenger returns his/her first boarding location with his/her last trip. This requirement couldn't be met for most of the 11 L trips in the last trip cluster. Hence, to analyze these trips another method used at second step is introduced. This method uses the next trip of the passengers on the upcoming days to infer the alighting location for the last trip of the day with certain assumptions.

It is hard to track the passenger after his/her last trip of the day. Also, the possibility of the passenger to be recorded in his/her next trip on the next days in a different location is quite high. Therefore, the proposed algorithm successfully worked for only small portion of the trips.

6 Conclusions

OD matrices with wide-range of usage can help transit agencies to improve the quality of transportation systems, in several aspects. With the information of origins and destinations of the passengers, transit planners can detect the critical and mostly used interchange stations. With the help of this information, required improvements can be introduced into these locations.

If the OD matrices are generated on the network level, they also give very useful outputs. For instance, passenger flows during any day can be explained by the results found in the generation of ODs for transit ridership. These results can be used by not only transit planners but also the city planners. To monitor the origin and destination of the passengers directly gives the very informative data to detect the residential locations and the locations mainly consisting of the work places. With this information, city planners can improve their decision making process about the possible locations of the planned new industrial zones or the new residential areas.

In this analysis OD estimation for a single route in İstanbul is made. With the help of previous studies about this topic, basic assumptions were taken to infer the alighting locations of the studied trips of September 15-23, 2014.

With the proposed algorithm, using new methods and assumptions, the trips which cannot be studied under the assumptions of the previous studies, were also analyzed.

The overall results showed that it is quite possible to estimate the destinations of passengers for the transit agency of İstanbul with only the analysis of recorded ADC data. Using the proposed algorithm, over 65% of all recorded trips alighting location was inferred. If the ADC data with missing boarding information or the repetitive use of the same card were excluded from the dataset, this success rate would reach to 70 percent.

As it is seen in this study, unique characteristics of the transportation systems should be further analyzed to eliminate the misleading outputs of any proposed algorithm. Since the passengers' public transportation usage habits are different from each other in different metropolises, the most appropriate assumptions and methods should be set with the consideration of these information. It is also known that transportation agencies in different cities apply different rules and systems. Therefore, properties of the studied public transportation systems should be analyzed carefully before the OD inference study is started.

With the comparison made between the result of this study and the observations made in the studied route, it is concluded that even if the assumptions made in the analysis are quite consistent and rational, it is highly possible to see different results in practice. This shows the difficulty of analyzing a system whose main component is the passengers. Human factors sometimes cannot be explained even by the accurate assumptions and methods. Therefore, it is suggested to further continue to conduct surveys in transportation systems not to collect data about the alighting and boarding location of the passengers but to understand the behavior and priorities of the passengers when using public transportation systems.

References

Barry, J., R. Newhouser, A. Rahbee, and S. Sayeda, 2002, "Origin and Destination Estimation in New York City Using Automated Fare System Data", *Transportation Research Record,* Vol. 1817, pp. 183-187.

Cui, A., 2006, "Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems", M.S. Thesis, *Massachusetts Institute of Technology*.

Trépanier, M., N. Tranchant, R. Chapleau, 2007. "Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System", *Journal of Intelligent Trasnportaton Systems: Technology, Planning and Operations,* Vol. 11, pp. 1-14.

Wang, W., 2010, "Bus Passenger Origin-Destination Estimation and Travel Behavior Using Automated Data Collection Systems in London, UK", M.S. Thesis, Massachusetts Institute of Technology.

Zhao, J., A. Rahbee, N. H. M. Wilson, 2007, "Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems", Computer- Aided Civil and Infrastructure Engineering, Vol. 22, No. 5, pp. 376-387.

Jiang, S., Yang, Y., Fiore, G., Jr., J.F., Frazzoli, E., González, M., 2013. "A Review of Urban Computing for Mobile Phone Traces: Current Methods, Challenges and Opportunities". Proceedings of the ACM SIGKDD International Workshop on Urban Computing.

Schneider, C., Belik, V., Couronné, T., Smoreda, Z., González, M.C., 2013. Unraveling Daily Human Mobility Motifs. J. Roy. Soc. Interface 10.

Iqbal, M.S., Choudhury, C.F., Wang, P., González, M.C., 2014. "Development of Origin-Destination Matrices Using Mobile Phone Call Data". Transportaion Research Part C, pp. 63–74.

Alexander, L., Jiang, S., Murga, M., Gonzalez, M., 2015. "Origin–Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data", Transportaion Research Part C, pp. 240–250.

Bonnel, P., Hombourger, E., Olteanu-Raimond, A.-M., Smoreda, Z., 2015. "Passive Mobile Phone Dataset to Construct Origin-Destination Matrix: Potentials and Limitations." Transportaion Research Procedia, pp. 381–398.