

Use of Mobile Ticketing Data to Estimate an Origin-Destination Matrix for New York City Ferry Service

Subrina Rahman
City College of New York
160 Convent Avenue, New York, NY 10031
+1-(917)-963-4602
srahman007@citymail.cuny.edu

James Wong, AICP
Vice President/Director of Ferries
New York City Economic Development Corporation
110 William Street, New York, NY 10038
+1-(212)-312-3688
jwong@edc.nyc

Candace Brakewood, PhD
City College of New York
160 Convent Avenue, New York, NY 10031
+1-(212)-650-5217
cbrakewood@ccny.cuny.edu

The views and opinions expressed in this article are those of the authors and do not necessarily represent those of New York City Economic Development Corporation or The City of New York.

ABSTRACT

One of the fundamental components of transit planning is understanding passenger demand, which is commonly represented using origin-destination (OD) matrices. However, manual collection of detailed OD information via surveys can be expensive and time consuming. Moreover, data from automated fare collection systems, such as smart cards, often include only entry information without tracking where passengers exit the transit network. New mobile ticketing systems offer the opportunity to prompt riders about their specific trips when purchasing a ticket, and this can be used to track OD patterns during the ticket activation phase. Therefore, the objective of this research is to utilize backend mobile ticketing data to generate passenger OD matrices and compare the outcome to OD matrices generated with traditional onboard surveys. Iterative proportional fitting (IPF) was used to create OD matrices using both mobile ticketing and onboard survey data. Then, these matrices were compared using Euclidean distance calculations. This was done for the East River Ferry service in New York City, and the results show that during peak periods, mobile ticketing data closely match survey data. In the off-peak and during weekends, however, when travelers are more likely to be non-commuters and tourists, matrices developed from mobile ticketing and survey data are shown to have greater differences. The impact of occasional riders making non-commute trips is the likely cause for these differences since commuters are familiar with using the mobile ticketing product and occasional riders are more likely to use paper ticket options on the ferry service.

INTRODUCTION

Urban ferry service plays an important role in the transit network in New York City. The New York City Economic Development Corporation (NYCEDC) manages the East River Ferry (ERF), a privately-operated ferry route connecting Manhattan, Brooklyn and Queens in New York City. According to the NYCEDC, the eight-stop route carried 1.3 million passengers in 2014. The ferry service was initially launched as a three-year pilot in 2011 and was turned into permanent service in 2014 following high ridership and extensive public support. During its pilot phase in 2012, the ERF became the first urban transit service in the United States to launch a mobile application for fare payment (1). This mobile ticketing application, or “app”, allows passengers to buy tickets directly on their smartphone using a credit or debit card (2). This new mobile ticketing system creates a rich source of data about where and when passengers utilize the ferry service. In light of this, the purpose of this study is to conduct an exploratory analysis of this backend mobile ticketing data in an anonymized format to assess its potential for planning/operations applications. Specifically, this project aims to assess if mobile ticketing data can be used to create origin-destination (OD) matrices, which are a fundamental input to the transit planning process.

PRIOR RESEARCH

This section provides a brief review of relevant prior research. First, a short summary of literature pertaining to OD estimation using data from automated fare collection (AFC) systems is presented, and this is followed by a recent study about AFC data from a ferry service.

AFC systems, such as those with smart cards, provide vast quantities of data about where and when passengers travel on transit systems, which can be used for transportation planning in many different ways (3). One common use of smart card data is to estimate transit passenger origin-destination (OD) matrices (*e.g.*, 4,5,6,7,8,9,10). However, one of the primary challenges in OD estimation using smart card data is that most transit systems only require the passenger to tap-in with their smart card, which means that only origin information is collected and destinations must be inferred (11). Therefore, opportunities for improved datasets exist when new fare collection systems collect both the origin and destination of transit passengers.

Although AFC systems are very common in urban bus and rail systems, they are infrequently used on ferry services. However, one recent study of a linear urban ferry service in Brisbane, Australia investigated travel behavior of passengers using more than a million smart card records over a six months period (12). This large dataset included numerous variables, such as the date and time of the trip, trip ID, route direction, boarding stop ID, and alighting stop ID. The results of the analysis showed aggregate-level trends in travel behavior; however, the authors did not use the dataset to estimate OD matrices for the ferry service.

In summary, there is a large literature pertaining to the use of AFC data for transit planning, particularly OD estimation. However, to the best of the authors’ knowledge, none of the prior studies have used data from new mobile ticketing systems, which are capable of collecting both origin and destination information. Moreover, there are few prior studies that utilize new data sources from urban ferry services. This research aims to help fill these gaps in the literature.

OBJECTIVE

New mobile ticketing systems have the potential to provide large amounts of automatically collected data about passenger travel patterns. Therefore, the objective of this research is to

utilize backend mobile ticketing data to create transit passenger OD matrices, and this is done for a single ferry route in New York City. To assess the validity of the mobile ticketing data, it is compared with the results of an onboard OD survey. Both datasets are used as to create seed matrices, which are then expanded using iterative proportional fitting. The mobile ticketing OD matrices are then compared to the survey data using Euclidean distance. Last, a brief analysis is conducted to identify the pattern of travel behavior of the ferry users.

BACKGROUND

This section provides background information about the East River Ferry service, as well as their mobile ticketing application.

Background about the East River Ferry

The NYCEDC is a non-profit corporation that works on behalf of the City of New York. The NYCEDC manages the ERF, a privately-operated ferry route providing transit service in New York City. The ERF service connects neighborhoods in Brooklyn and Queens to Manhattan. Figure 1 shows the route map, which runs from Pier 11 to East 34th Street in Manhattan, with four stops in Brooklyn and one stop in Queens. Additionally, the ERF provides seasonal weekend access to Governors Island between May and September. The service is provided using three vessels, and there is a peak hour headway of 20 minutes. The off-peak headway varies between 30-80 minutes based on the season.

Background about Mobile Ticketing

In January of 2012, the operator of the ERF launched a mobile ticketing smartphone application known as “NY Waterway” for payment on both its Hudson River and East River Ferry services. The app was developed by the company Bytemark, and it can be downloaded for free from iTunes and the Google Play store. Using the app, passengers can buy a ticket at any time and then activate the ticket before they are ready for travel on the ferry. The process of purchasing and then activating a ticket creates a record in a backend database, which provides a rich source of data about where and when passengers are traveling on the ferry service.

Because riders can purchase tickets for multiple ferry routes, the app prompts users for their origin and destination in order to determine the ticket type, as shown in Figure 2. Notably, however, the ERF route has a flat fare for all origin-destination pairs. Because the price is not tied to the origin or destination, passengers could input arbitrary origin-destination information and still purchase a valid ticket for any trip on the ERF. Therefore, this analysis aims to assess the reliability of this new data source by comparing it to a traditional origin-destination paper survey.

DATA SOURCES

The analysis presented in this paper relies on three different sources: an onboard origin-destination survey, mobile ticketing data and onboard ridership counts. These three sources are described in the following paragraphs, beginning with the onboard survey.

Onboard Origin-Destination Survey

An onboard survey was conducted on the ERF during six days in October 2014. The survey was distributed using paper cards that combined OD information and three questions on travel behavior including trip purpose, frequency of riding the ERF, and access mode to/from the ferry.

Survey cards were color-coded for each of the seven stops of the ERF route, and a colored card corresponding to the boarding stop was handed to each passenger as they entered the boat. Passengers were asked to return the card to the survey staff when they reached their destination; therefore, passengers did not have to answer questions about boarding and alighting because this was collected via the color-coded cards.

Ridership counts were used to calculate a baseline number of responses necessary for a sample size that would provide a 95% confidence level in four different time periods: AM peak (7-9:30AM), PM peak (4:30-7PM), midday (9:30-4:30), and weekend. The goal was to sample passengers on each ferry trip during the previously mentioned periods over the course of six days. It should be noted that there was an incident on one of the PM peak trips, during which passengers were transferred to another boat; however, most passengers still completed the survey, so this data are included in the analysis. In summary, a total of 1,367 responses were collected, and these were used to create a baseline OD matrix to compare with the mobile ticketing data.

Mobile Ticketing Data

Mobile ticketing data were provided by the app developer for the month of October 2014, which was the same period when the onboard survey was conducted. In total, 37,486 tickets were activated using the NY Waterway app during that month. The data included the origin-destination pair that each passenger inputted into the application, the ticket type (e.g., single ride) that was purchased, and the date and time that the ticket was activated. No personally identifiable information was included in the dataset. The date and time fields were used to select only the transactions that occurred during the same time periods as the onboard survey, which reduced the sample size to 3,544 observations for the four specified time periods that were used in the following OD estimation process.

Table 1 shows the mobile ticketing data sample size for each time period compared with the onboard survey sample size for the same time period. The mobile ticketing activation counts are larger than the number of survey responses in most periods because all mobile ticketing activations were included in the table when the onboard survey was conducted during multiple days in a specific time period (e.g., October 7 and 8 for the AM peak period).

Ridership Data

The ferry operator provided boarding and alighting ridership data for the month of October 2014. In order to comply with Coast Guard regulations related to the number of passengers on a vessel at any one time, the ferry operator conducts manual on- and off-counts of passengers at each stop on all trips on the ERF. This manually collected data required minor adjustments; specifically, on-counts were assumed to be true and off-counts were adjusted accordingly based on the assumption that all passengers disembark at the last stop for each run. In the following analysis, the average monthly on-and off-counts during each of the four time periods were used as the marginal values to create the OD matrices, which will be discussed in more detail in the methodology section.

METHODOLOGY

The analysis presented in this paper is comprised of two sets of calculations. First, both the survey and mobile ticketing data were adjusted using iterative proportional fitting (IPF) to estimate OD matrices for the full ridership based on on/off counts during four time periods.

Second, the chi-squared Euclidean distance was calculated to compare the OD matrices from the mobile ticketing data and the onboard survey data. This procedure is summarized in Figure 3 and discussed in more detail in the following paragraphs, beginning with the IPF process.

Iterative Proportional Fitting

Iterative proportional fitting (IPF) is a commonly used method to adjust the cells in a matrix to fit the rows and columns under given constraints (14). The original values of a two dimensional table are adjusted gradually by repeating calculations to meet row and column constraints (referred to as the marginals), similar to a weighting system. In the case of transit passenger OD matrices, a seed matrix from baseline OD data is gradually adjusted to match passenger boarding and alighting volumes that are shown on the marginals of the matrix (15,16,17). This allows a small sample of OD information, such as from an onboard survey, to be expanded to meet aggregate on/off counts.

The formula used in this study is presented below (18). Each cell has value $P_{ij(k)}$, where i designates the row number, j represents the column number, and k is the iteration number. To make a row adjustment for iteration $k+1$ (shown in equation [1]), each cell ($P_{ij(k)}$) is divided by the sum of that row of cells, and then this value is multiplied by the marginal row total (Q_i). Columns are adjusted in a similar manner, as shown in equation [2]. This iterative process is continued until the values converge to the marginal values (Q_i, Q_j).

$$P_{ij(k+1)} = \frac{P_{ij(k)}}{\sum_j P_{ij(k)}} * Q_i \quad [1]$$

$$P_{ij(k+2)} = \frac{P_{ij(k)}}{\sum_i P_{ij(k)}} * Q_j \quad [2]$$

Where

$P_{ij(k)}$ = a single element of OD matrix,

i, j, k = row, column, and iteration serial number, respectively, and

Q_i, Q_j = marginal row totals and marginal column totals, respectively.

Comparison of Matrices using Euclidean Distance

After OD matrices were calculated using IPF for both the mobile ticketing and onboard survey data, the Euclidean distance was used to compare the probability distribution of matrices, which were represented using the percent of passengers traveling between any given OD pair. Though there are a number of techniques to measure distances between matrices (19), Euclidean distance is one of the simplest and most widely accepted methods. The formula used in this study is shown in equation [3], where D is the Euclidean distance between matrix X and matrix Y .

$$D = \sqrt{\sum_{n=1}^n (X_n - Y_n)^2} \quad [3]$$

Where

D = the Euclidean distance, and

X_n, Y_n = the n th elements of matrices X and Y , respectively.

RESULTS

The results of the analysis are discussed in four parts. First, OD estimation results using the IPF procedure are summarized for each time period (AM peak, PM peak, midday, and weekend) using both survey and mobile ticketing data. Second, the distance calculations are presented that compare the mobile ticketing and survey data to one another. Third, an additional analysis of other travel behavior questions from the onboard survey is presented. Finally, an analysis of the ticket types from the mobile ticketing data is presented.

IPF Calculations

In order to conduct the IPF procedure, the boarding and alighting numbers from the ridership counts were used as the marginal values for the row and column totals for each of the four time periods (AM peak, PM peak, midday and weekend). Then, the onboard survey responses were used to create seed matrices for each of the four time periods, and similarly, the mobile ticketing data were used to create separate seed matrices for each time period. However, the seed matrices of both the survey and mobile ticketing data had missing values for a small number of OD pairs, which could skew the results when used in the IPF procedure. A common way to solve this problem is to add arbitrarily small values (for example, 0.001, 0.1, or 0.5) to all cells with missing records (20,21), and in this study, 0.1 was added to all cells having missing passenger movements. Then, the corrected seed matrices were used in an IPF procedure to meet the total boarding and alighting counts from the ridership data, and the results are briefly discussed in the following four sections for each time period.

AM Peak

The results for the AM peak period (7-9:30AM) are shown in Table 2, which includes four tables displaying the seed and final matrices for both the survey and mobile ticketing data. The average ridership count during the AM peak period for the month of October was 1,345, and this is shown in Table 2 as the total marginal value. The number of onboard surveys collected during the AM peak was 370 (not shown), which occurred over two days (October 7 and 8), whereas the corresponding number of mobile ticketing activations during the same two days was 1,476 (not shown).

As can be seen on the left side of Table 2, the most common origin in the seed matrices was North Williamsburg in Brooklyn (38% of survey data; 32% of mobile ticketing data), whereas the most prevalent destination was either Pier 11 or East 34th Street in Manhattan (85% combined total in the survey data; 68% in the mobile ticketing data). The results of the IPF procedure are shown on the right side of Table 2, and the most common origin-destinations pairs were North Williamsburg-East 34th Street (24% in the survey data and highlighted in dark blue; 22% in the mobile ticketing data and highlighted in dark red) and North Williamsburg-Pier 11 (13% of survey data and highlighted in light blue; 15% of mobile ticketing data and highlighted in light red). These origin-destination patterns seem reasonable given local land use patterns, as North Williamsburg is a popular residential area, whereas Pier 11 serves the financial district in Manhattan and East 34th Street is in midtown Manhattan. However, the two datasets did have some minor differences. For example, the mobile ticketing seed matrix had some ridership between Pier 11 and North Williamsburg (5%), as well as East 34th Street and North Williamsburg (6%), whereas the survey did not; the mobile ticketing data for these OD pairs were scaled down to 0% and 1%, respectively, in the final matrix.

Midday

During the midday period (9:30AM-4:30PM), the average ridership count for the month of October was 1,806, which was used for the total marginal value in the IPF calculations (results not shown). The total number of onboard surveys collected during the midday was 329, which occurred on October 8, 9, and 30. The total number of mobile ticket activations during the midday periods on those three days was 967. After conducting the IPF calculations, the most common OD pair was DUMBO and Pier 11 for both survey (15%) and mobile ticketing (12%) data at midday. Interestingly, Greenpoint to East 34th Street ranked as a popular midday pair (6% for survey data and 3% for mobile ticketing data), indicating unique commuting patterns and/or an emergence of Greenpoint as a tourist locale.

PM Peak

In the PM peak period (4:30-7PM), the average ridership count for October was 1,296, which was used as the total for the marginals in the IPF calculation (results not shown). The total number of onboard surveys collected during the PM peak period was 338, which occurred on October 8 and 22. A total of 785 mobile tickets were activated on those two days in the PM peak period.

In the seed matrices, the highest percentage of passengers boarded the ferry from Pier 11 (42% for survey data and 34% for mobile ticketing data) and East 34th Street (34% for survey data and 23% for mobile ticketing data) in Manhattan, and the most common destinations were the residential neighborhoods of North Williamsburg in Brooklyn (30% for survey data and 25% for mobile ticketing data) and Long Island City in Queens (20% for survey data and 12% for mobile ticketing data). This pattern is nearly opposite that from the AM peak period, as was expected. Conversely, DUMBO in Brooklyn had strong PM productions (18%) compared to AM attractions, perhaps due to the emerging job centers in DUMBO that do not necessarily abide by typical peak hour commute patterns (including more tech-oriented companies). The results of the IPF calculations for both the onboard survey and the mobile ticketing data were similar, and the most popular OD pairs were Pier 11-North Williamsburg and East 34th Street-North Williamsburg.

Weekend

For the weekend, the average ridership count in the month of October 2014 was 4,116 for Saturdays and Sundays (results not shown). The total number of onboard surveys collected on the weekend was 330, and these were collected on Saturday, October 18. On that date, there were a total of 316 mobile ticket activations, which was relatively few compared to weekday periods.

The weekend seed matrices for the onboard survey and mobile ticketing data had some differences. For example, the onboard survey seed matrix revealed that Pier 11 had the highest percentage (29%) of alighting passengers, while the mobile ticketing data had only 14% of alighting passengers. Despite these differences, the largest OD pairs were identical to the midday period, with Dumbo-Pier 11 being the most popular OD pair (13% for the survey and 8% for the mobile ticketing data).

Comparison of Matrices

After the matrices were created for the survey data and mobile ticketing data, a comparison of both the seed and final IPF matrices was conducted. The metric used for the comparison was the chi-squared Euclidean distance. It should be noted that other measures of distance (e.g., the

Hellinger distance) were also considered but are not presented because the results were similar. The Euclidean distance was calculated for each time period (AM, PM, midday and weekend) in the following four ways:

1. Between the seed matrix of the mobile ticketing data and the seed matrix of the onboard survey data;
2. Between the seed and final IPF matrices for the mobile ticketing data only; and
3. Between the seed and final IPF matrices for the onboard survey data only; and
4. Between the final IPF matrix of the mobile ticketing data and the final IPF matrix of the onboard survey data.

The four different distance calculations are shown for each time period in Figure 4. The range of Euclidean distances is between 0.04 and 0.18. In general, the AM and PM peak periods have smaller distances compared to midday and weekend periods. The most important result is the comparison of IPF adjusted matrices for mobile and onboard survey data, which is at the bottom of Figure 4. This shows that the AM and PM peak distances are very small (0.048 for the AM and 0.039 for the PM), whereas the midday and weekend distances are much larger (0.086 for the midday and 0.088 for the weekend). These results suggest that origin-destination information from mobile ticketing data most closely aligns with the onboard survey data during the peak periods.

Behavior Analysis from Survey Questions

In addition to the origin-destination information, the onboard survey also contained a small number of travel behavior questions, including trip purpose and frequency of travel on the ERF. The results from these two survey questions are summarized in the Table 3. As can be seen in the table, the total sample sizes are slightly smaller than the origin-destination responses during each time period because approximately 1% of the surveys had inconsistent or missing responses to these questions. The top part of Table 3 shows the results from the trip purpose question, which revealed that a majority of passengers were commuting during the AM peak (92%) and PM peak (83%). On the other hand, only 40% of passengers commuted during the midday period, while 44% took the ERF for leisure or fun. However, on the weekend, approximately 69% of riders used the ferry for leisure or fun.

The bottom part of Table 3 shows the results to a question asking riders how many trips they make on the ERF in a typical week, and the responses were categorized based on frequency of use. Approximately 71% of riders in the AM peak and 57% of riders in the PM peak take the ferry 4 to 10 times in a typical week, which is likely indicative of commuting behavior. On the other hand, 35% of midday passengers and 45% of weekend riders rode the ferry for the first time, which may represent a large percentage of tourists.

Analysis of Mobile Ticketing Data

The backend data from the mobile ticketing application also includes information about what ticket type was purchased (e.g., one way), which is required to complete the transaction and correctly charge the fare to the passenger. In the app, the rider can select one of eleven different ticket types, and the percent of passengers who bought each type of ticket during the days and time periods when the onboard survey were conducted are shown in Table 4. This analysis reveals that all four time periods have similar patterns, and the majority of the mobile ticketing users bought one way tickets (either weekday or weekend depending on the period). Since a

large number of trips in this sample were made during AM peak period, there is a good chance that many of the users were commuters. Only a small percentage of mobile ticketing users bought 30 day or monthly tickets due to the limited discount available; similarly, a small number of users purchased bike passes (a \$1 fee for bringing a bike onboard the ferry).

CONCLUSIONS

One of the fundamental requirements for transit planning is to understand passenger demand in the form of origin-destination (OD) matrices. However, manual collection of detailed OD data can be expensive and time consuming. New mobile ticketing systems have the potential to provide large amounts of automatically collected data with passenger OD information. This research demonstrated that backend data from a mobile ticketing application used on New York's East River Ferry (ERF) service can be utilized to estimate OD matrices. The ERF mobile ticketing dataset was compared to the results of a recent onboard OD survey, and both datasets (mobile and survey) were expanded using iterative proportional fitting (IPF) to meet on/off ridership counts for four different time periods (AM peak, PM peak, midday and weekend). These OD matrices were compared using the Euclidean distance, and the results revealed that distances between the mobile ticketing and survey datasets were very small for the AM and PM peak, whereas the midday and weekend distances were larger. This suggests that OD information from mobile ticketing data most closely aligns with the survey data during the peak periods.

Additionally, the recent onboard survey included a small number of questions pertaining to travel behavior. These questions showed that the majority of peak period passengers were commuters (92% during the AM and 83% during the PM peak), and that most peak period passengers were regular riders (82% during the AM and 67% during the PM peak typically take the ferry at least 4 times per week). Based on the results of these survey questions combined with the OD estimation results, it can be inferred that most mobile ticketing users during the AM and PM peak periods are likely commuters and/or regular ERF riders. Therefore, mobile ticketing systems are likely to provide the most reliable travel behavior information during peak periods when travel patterns are more consistent compared to the midday or weekend. An important area for future research is to examine mobile ticketing purchases over time; however, this would require an identifier for each person to assess individualized travel behavior, which was not available in this anonymized dataset.

An additional analysis was performed with the mobile ticketing data to understand time of day trends for ticket types. Most riders purchased one way tickets regardless of the time period. This may be due to the price of the one way tickets compared to the 30 day or monthly passes on the ERF, which offer little incentive to purchase a period pass over single tickets. These findings may differ in other transit systems with mobile ticketing systems where there is a greater price differential between ticket types.

In summary, this research demonstrated that mobile ticketing systems provide origin-destination information that can be used to supplement and in some cases (such as the peak periods) perhaps substitute for manually collected data. Practitioners must exercise caution, however, to account for sample biases that result from commuters who are more likely to use mobile ticketing apps due to their familiarity of the systems compared with occasional riders who are more likely to use traditional fare media, such as paper tickets. These results suggest that this new fare collection technology has significant potential to provide large quantities of valuable travel information about where and when passengers are traveling, which can be used for transportation planning.

ACKNOWLEDGEMENTS

The authors acknowledge the Billy Bey Ferry Company for its operation of the ERF and cooperation with this research, especially Paul Goodman, Donald Liloia, and Larry Vanore; and Bytemark, the company that developed the app and provided access to data for research, especially Jesse Wachtel. The authors also thank Susan Liu for her help during her internship at NYCEDC.

REFERENCES

1. Tavilla, E. *Transit Mobile Payments: Driving Consumer Experience and Adoption*. Federal Reserve Bank of Boston, February, 2015.
2. Brakewood, C., et al. Forecasting Mobile Ticketing Adoption on Commuter Rail. *Journal of Public Transportation*, Volume 17, Issue 1, 2014, pp. 1-19.
3. Pelletier, M.-P., Trépanier, M., and Morency, C. Smart Card Data Use in Public Transit: A Literature Review. *Transportation Research Part C: Emerging Technologies*, 19(4), 2011, pp. 557–568.
4. Barry, J., Newhouser, R., Rahbee, A. and Sayeda, S. Origin and Destination Estimation in New York City Using Automated Fare System Data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1817, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 183-187.
5. Cui, A. Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems. Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2006.
6. Farzin, J. Constructing an Automated Bus Origin-Destination Matrix Using Farecard and Global Positioning System Data in Sao Paulo, Brazil. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2072, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 30-37.
7. Frumin, M. Automatic Data for Applied Railway Management: Passenger Demand, Service Quality Measurement, and Tactical Planning on the London Overground Network. Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2010.
8. Wang, W., Attanucci, J., and Wilson, N. Bus passenger origin-destination estimation and related analyses using automated data collection systems. *Journal of Public Transportation*, 14.4: 7, 2011.
9. Munizaga, M. and Palma, C. Estimation of a Disaggregate Multimodal Public Transport Origin-Destination Matrix from Passive Smartcard Data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*. Volume 24, 2012, pp. 9-18.
10. Gordon, J., Koutsopoulos, H., Wilson, N., and Attanucci, J. Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2343, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 17-24.
11. Zhao, J., Rahbee, A., and Wilson, N. H.M. Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems. *Computer-Aided Civil and Infrastructure Engineering*. Volume 22, No. 5, 2007, pp. 376-387.
12. Soltani, A., et al. *Travel Patterns of Urban Linear Ferry Passengers: Analysis of Smart Card Fare Data for Brisbane, Australia*. Presented at the 94th Annual Meeting of the Transportation Research Board, Washington D.C., 2015.
13. East River Ferry. Route Map. <http://www.eastriverferry.com/RouteMap.aspx>. Accessed July 31, 2015.
14. Ben-Akiva, M., Macke, P. and Hsu, P. Alternative methods to estimate route-level trip tables and expand on-board surveys. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1037, Transportation Research Board of the National Academies, Washington, D.C., 1985, pp. 1-11.

15. Mishalani, R. et al. *Iteratively Improving the Base Matrix of the IPF Method for Estimating Transit Route- Level OD Flows from APC Data*. Presented at the 13th World Conference on Transportation Research, 2013.
16. Ji, Y., Mishalani, R. and McCord, M. *Estimating transit route OD flow matrices from APC data on multiple bus trips using the IPF method with an iteratively improved base: method and empirical evaluation*. *Journal of Transportation Engineering* 140.5, 2014.
17. Mishalani, R., Ji, Y., and McCord, M. Effect of Onboard Survey Sample Size on Estimation of Transit Bus Route Passenger Origin-Destination Flow Matrix Using Automatic Passenger Count Data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2246, Transportation Research Board of the National Academies, Washington, D.C., 2011.
18. Wong, David, W.S. *The Reliability of Using the Iterative Proportional Fitting Procedure*. *The Professional Geographer*, Volume 44.3, 1999, pp. 340-348.
19. Adbi, H. *Distance*. *Encyclopedia of Measurement and Statistics*, Sage. Thousand Oaks, CA, <https://www.utdallas.edu/~herve/Abdi-Distance2007-pretty.pdf>, 2007. Accessed August 1, 2015.
20. Beckman, R., Baggerly, K. and McKay, M. Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*. Volume 30.6, 1996, pp. 415-429.
21. Müller, K. and Axhausen, K. *Population synthesis for microsimulation: State of the art*. ETH Zürich, <http://www.strc.ch/conferences/2010/Mueller.pdf>, 2010. Accessed August 1, 2015.

List of Tables and Figures

FIGURE 1 East River Ferry Route Map (13)

FIGURE 2 Demonstration Screenshots of Mobile Ticketing Application

TABLE 1 Summary of Survey and Mobile Ticketing Data for the Four Time Periods

FIGURE 3 Methodology for the OD Estimation

TABLE 2 AM Peak Period Seed (left) and Adjusted (right) OD Matrices for Survey (top) and Mobile Ticketing (bottom) Data

FIGURE 4 Euclidean Distance Calculation for Four Time Periods

TABLE 3 East River Ferry Survey Travel Behavior Questions

TABLE 4 Mobile Ticketing Fare Types

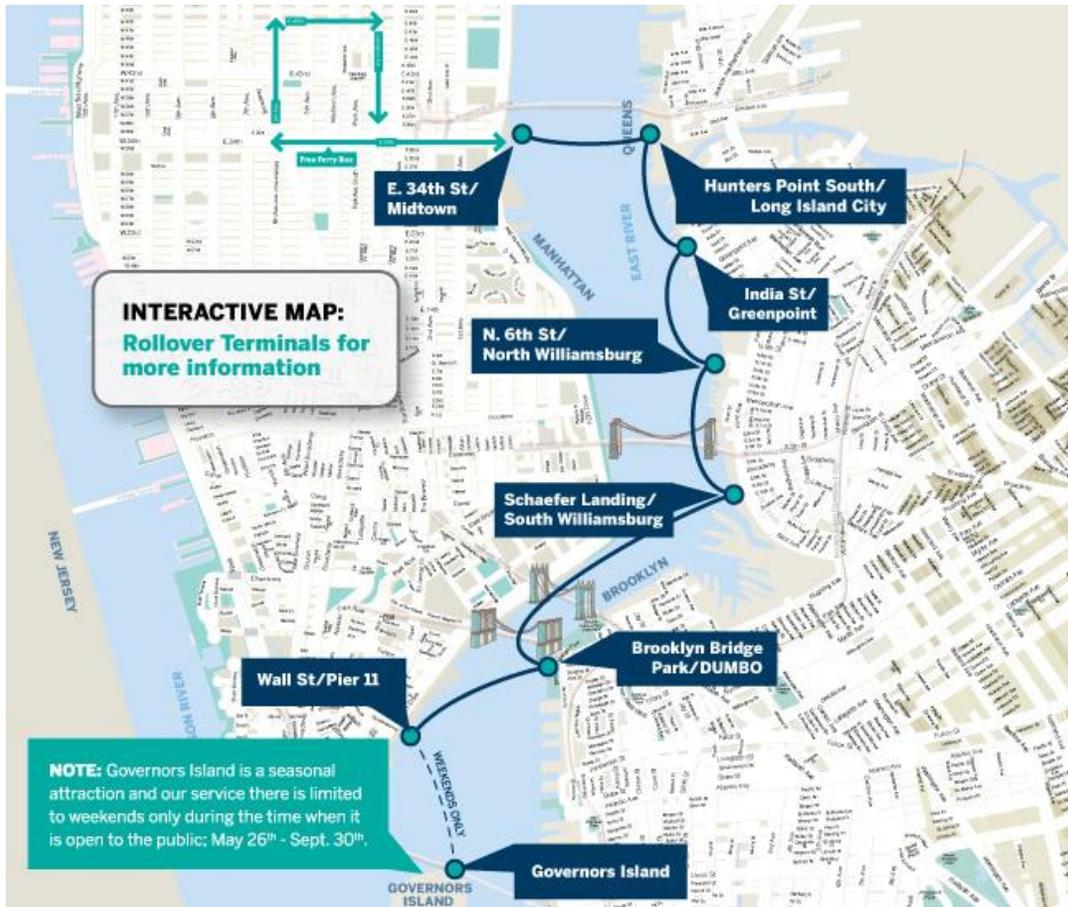


FIGURE 1 East River Ferry Route Map (13).

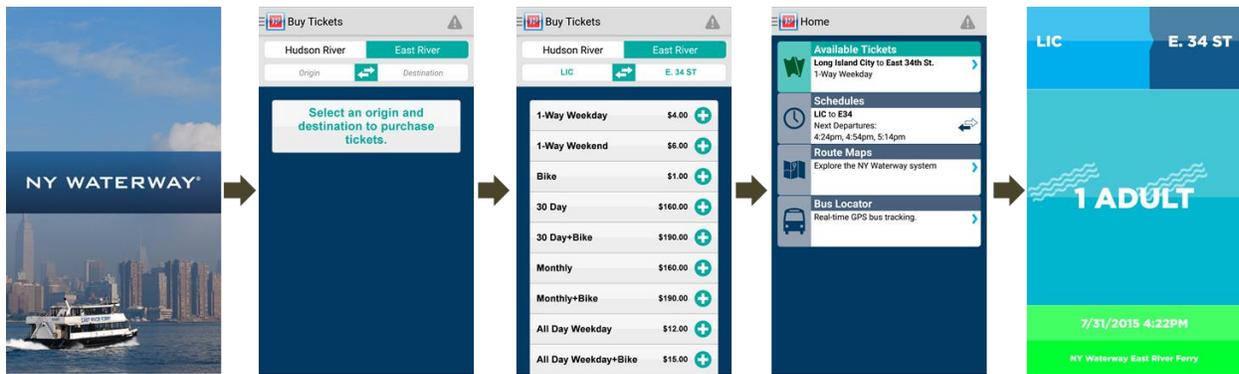


FIGURE 2 Demonstration Screenshots of Mobile Ticketing Application.

TABLE 1 Summary of Survey and Mobile Ticketing Data for the Four Time Periods

		AM Peak	PM Peak	Midday	Weekend	Total
Onboard Survey Data	Count	370	338	329	330	1,367
	%	27%	25%	24%	24%	100%
Mobile Ticketing Data	Count*	1,476	785	967	316	3,544
	%	42%	22%	27%	9%	100%
Dates included		Oct. 7 & 8	Oct. 8 & 22	Oct. 8, 9 & 30	Oct. 18	-
<i>*All days when the survey was conducted in each period are included in the mobile ticketing data.</i>						

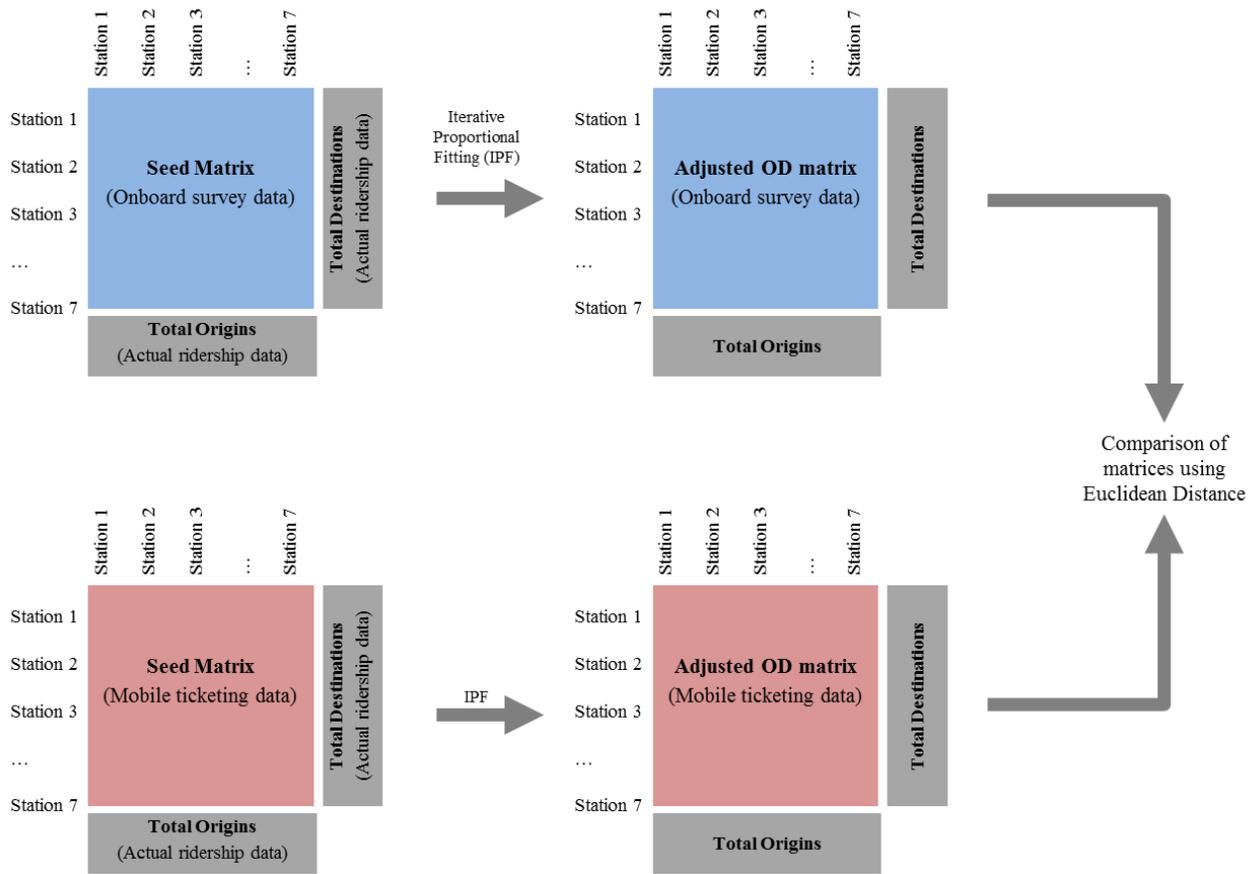


FIGURE 3 Methodology for the OD Estimation.

TABLE 2 AM Peak Period Seed (left) and Adjusted (right) OD Matrices for Survey (top) and Mobile Ticketing (bottom) Data

		Seed Matrix (Onboard survey data)							Adjusted OD Matrix (Onboard survey data)									
Destinations ▶																		
Origins ▼		Pier 11	DUMBO	S. Williamsburg	N. Williamsburg	Green point	Long Island City	E. 34th street	Total		Pier 11	DUMBO	S. Williamsburg	N. Williamsburg	Green point	Long Island City	E. 34th street	Total
	Actual Ridership ▼▶	524	100	12	23	17	18	651	1345	Actual Ridership ▼▶	524	100	12	23	17	18	651	1345
Pier 11	38	0%	1%	0%	1%	1%	0%	0%	2%	38	0%	1%	0%	1%	1%	0%	0%	3%
DUMBO	104	7%	0%	0%	1%	0%	0%	1%	8%	104	6%	0%	0%	0%	0%	0%	1%	8%
S. Williamsburg	140	3%	2%	0%	0%	0%	0%	6%	11%	140	3%	1%	0%	0%	0%	0%	6%	10%
N. Williamsburg	530	14%	3%	0%	0%	0%	0%	21%	38%	530	13%	2%	0%	0%	0%	0%	24%	39%
Greenpoint	190	6%	2%	0%	0%	0%	0%	6%	15%	190	5%	1%	0%	0%	0%	0%	7%	14%
Long Island City	259	11%	1%	0%	0%	0%	0%	9%	22%	259	9%	1%	0%	0%	0%	0%	9%	19%
E 34th St	84	1%	1%	0%	1%	1%	1%	0%	4%	84	2%	1%	1%	1%	1%	1%	0%	6%
Total	1345	42%	10%	1%	2%	1%	1%	43%	100%	1345	39%	7%	1%	2%	1%	1%	48%	100%

IPF Method →

		Seed Matrix (Mobile ticketing data)							Adjusted OD Matrix (Mobile ticketing data)									
Destinations ▶																		
Origins ▼		Pier 11	DUMBO	S. Williamsburg	N. Williamsburg	Green point	Long Island City	E. 34th street	Total		Pier 11	DUMBO	S. Williamsburg	N. Williamsburg	Green point	Long Island City	E. 34th street	Total
	Actual Ridership ▼▶	524	100	12	23	17	18	651	1345	Actual Ridership ▼▶	524	100	12	23	17	18	651	1345
Pier 11	38	0%	2%	2%	5%	1%	3%	1%	15%	38	0%	1%	0%	0%	0%	0%	1%	3%
DUMBO	104	3%	0%	0%	1%	1%	0%	2%	7%	104	5%	0%	0%	0%	0%	0%	3%	8%
S. Williamsburg	140	3%	1%	0%	0%	0%	0%	4%	8%	140	3%	1%	0%	0%	0%	0%	6%	10%
N. Williamsburg	530	14%	1%	0%	0%	0%	0%	17%	32%	530	15%	2%	0%	0%	0%	0%	22%	39%
Greenpoint	190	6%	1%	0%	0%	0%	0%	7%	14%	190	5%	1%	0%	0%	0%	0%	8%	14%
Long Island City	259	6%	1%	0%	0%	0%	0%	5%	12%	259	9%	1%	0%	0%	0%	0%	9%	19%
E 34th St	84	1%	0%	1%	6%	2%	2%	0%	12%	84	1%	1%	1%	1%	1%	1%	0%	6%
Total	1345	32%	7%	4%	13%	4%	5%	36%	100%	1345	39%	7%	1%	2%	1%	1%	48%	100%

Source: Ridership Counts (gray), Onboard Survey (blue) and Mobile Ticketing Data (red) from October 2014

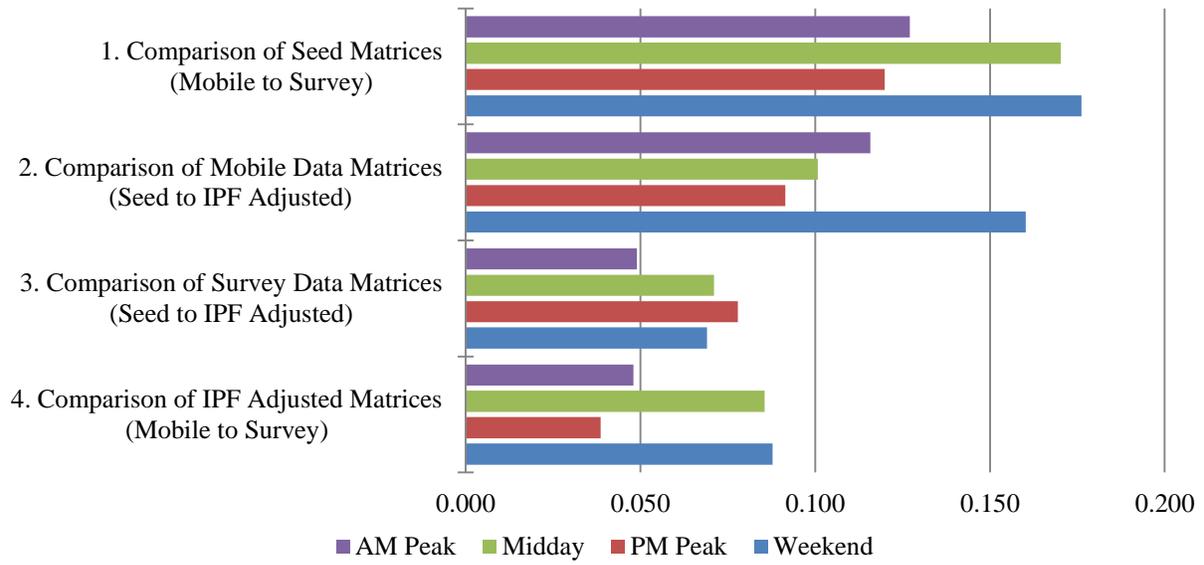


FIGURE 4 Euclidean Distance Calculation for Four Time Periods.

TABLE 3 East River Ferry Survey Travel Behavior Questions

Questions ⁱ	Answers	AM Peak ⁱⁱ	Midday ⁱⁱ	PM Peak ⁱⁱ	Weekend ⁱⁱ
What is the primary purpose of your trip today?	1. Commuting	92%	40%	83%	13%
	2. Leisure /Fun	3%	44%	12%	69%
	3. No Response	5%	14%	5%	16%
How many trips do you typically take on the East River Ferry in a week? (count each direction as one trip)	1. 11 or more	11%	4%	10%	1%
	2. 4 to 10	71%	18%	57%	4%
	3. 2 or 3	7%	15%	9%	8%
	4. 0 or 1	4%	14%	9%	27%
	5. First time rider	2%	35%	8%	45%
	6. No Response	5%	14%	5%	15%
Total Respondents		370	329	338	330
<i>i. Question wording is exactly as it appeared on the questionnaire</i>					
<i>ii. Percentages are rounded to the nearest whole number</i>					

TABLE 4 Mobile Ticketing Fare Types

	AM Peak	Midday	PM Peak	Weekend	Total Count
1-Way	0%	0%	0%	1%	8
1-Way Weekday	87%	84%	87%	4%	2,788
1-Way Weekend	0%	0%	0%	83%	266
30 Day	5%	5%	5%	4%	178
30 Day + Bike	0%	1%	0%	0%	11
All Day	0%	0%	0%	0%	2
All Day Weekday	0%	2%	0%	3%	28
All Day Weekday +Bike	0%	0%	0%	1%	2
Bike	5%	6%	3%	4%	159
Monthly	3%	1%	5%	1%	101
Monthly +Bike	0%	0%	0%	0%	1
Percent Total	100%	100%	100%	100%	
Total Count	1,476	967	785	316	3,544

Note: Percentages shown may not add up to 100% due to rounding.