APPLYING SPATIAL AGGREGATION METHODS TO IDENTIFY OPPORTUNITIES FOR NEW BUS SERVICES IN LONDON

Cecilia Viggiano, Corresponding Author
Affiliation at time of research: Massachusetts Institute of Technology
Current Affiliation: Economic Development Research Group
155 Federal St. 6th Floor, Boston, MA 02110 USA
Tel: 617-338-6775, x223; Email: cviggiano@edrgroup.com

Haris N. Koutsopoulos
Northeastern University
403 SN, 360 Huntington Avenue, Boston, MA 02115
Tel: 617-373-6263; Email: h.koutsopoulos@neu.edu

Nigel H.M. Wilson
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge MA 02139
Tel: 617-253-5046; Email: nhmw@mit.edu

John Attanucci
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge MA 02139
Tel: 617-253-7022; Email: jattan@mit.edu

Word count: 6,412 words text + 6 tables/figures x 250 words (each) = 7,912

Submission Date: November 15, 2017

ABSTRACT

Innovative analyses of origin-destination (OD) data derived from automatic fare collection and automatic vehicle location systems in public transport networks enable planners to gain new insights into how passengers travel in the network and the quality of service provided, and can even inform decisions about network improvements. Particularly in large, complex networks, systematic, data-driven approaches to network evaluation and planning are essential. New methodologies are needed to transform OD data into informative metrics and planning recommendations. This paper proposes a framework for this process and applies it to London's public transport network. While there are many ways to improve public transport networks, this paper focuses on the addition of new bus routes to reduce circuity. The proposed framework includes three steps that combine OD level analysis with spatial aggregation methodologies for the identification of corridors for new bus services. First, bus stops and rail stations are clustered into geographic zones. Second, a subset of zonal OD pairs with circuitous service are identified as candidates for improvement through new bus routes, based on performance standards established with user-defined parameters. Third, an algorithm that clusters OD pairs into corridors is applied to identify promising corridors for new bus service. This paper discusses corridors identified for new service in the London case study.

Keywords: Spatial, data, planning, bus, network, circuity

INTRODUCTION

Innovative approaches, including quantitative and spatial aggregation techniques, can produce new insights for transportation evaluation and planning. In tandem with these techniques, increased availability of origin-destination (OD) data from new sources presents opportunities for improved understanding of behavior and better evaluation of mobility and accessibility, which can lead to better planning.

The framework for data-driven sketch planning applied in this paper represents an alternative to existing approaches to public transport planning. One challenge, adding new bus routes to reduce circuity and transfers, is explored here, with an application to London's public transport network. In this paper, bus routes are defined as circuitous when the distance traveled by bus significantly exceeds the shortest path distance. Using geographic zones that reflect the structure of the existing public transport network, OD level analysis is conducted to determine which OD pairs can be improved by new bus services that reduce circuity or eliminate transfers. Improvable OD pairs are spatially aggregated into corridors that are appropriate for new bus services. This paper includes a review of related literature, an overview of the framework applied, and a discussion of the implementation and results of the framework for the London network.

Opportunity for a Data-Driven Approach

The increasing availability of travel information at the OD level presents an opportunity for improved bus planning. Automatic fare collection (AFC) and automatic vehicle location (AVL) data are routinely collected in many public transport networks. Recently, methodologies have been developed to infer complete journey information from this data (*1*,*2*). These methodologies infer destination and interchange information for a large share of boardings that use fare cards, often upwards of 70% (*2*). Compared to OD data collected from surveys, which is generally collected for a small number of passengers on a few selected days, this vastly increases the sample size, improving the accuracy and precision of estimates of OD-level demand and travel time. Making use of emerging information for planning requires new metrics and methods for evaluation and aggregation.

Traditional Network Planning Approaches

A methodology that can be applied to an existing public transport network to identify opportunities for improvement fills a gap in the public transport planning literature. Traditional approaches to public transport network design use partial optimization methods to identify the set of routes or lines which best serve a specified demand matrix. These approaches may produce a network that bears little resemblance to the existing network. As a result, within established networks, planners generally have little use for such results. Instead, planners typically make incremental changes to the existing network over time. Because changes to bus services are significantly less costly (and less permanent) than in rail networks, planners may use a trial-and-error approach to planning, piloting routes to assess the impacts. While this can be effective in small networks, in large, complex networks, it is inefficient. This leaves an unaddressed need to systematically identify opportunities for incremental improvement.

The Role of Circuity

This paper focuses on the introduction of new routes to reduce circuity and transfers in a public transport network. Huang and Levinson found evidence that commuters live in places where their commute trip will have low circuity, with circuity measured as the ratio of public transport distance to straight line distance (*3*). They also found that public transport accessibility measures for 36 metropolitan areas were negatively correlated with circuity. This suggests that circuity is not only undesirable for specific trips, it may also degrade the overall quality of the public transport system. In support of this, many graph theoretic metrics, designed to assess public transportation networks, characterize them in terms of circuity or directness (*4*,*5*).

Transferring is widely regarded as detracting from public transport attractiveness and quality of service, with a large body of research dedicated to understanding the transfer penalty. While the penalty can be reduced by coordination of service or high frequency of service, eliminating transfers can be beneficial.

Introducing new services to eliminate circuity and transfers requires investment of money and resources. A methodology that systematically assesses the benefits from reduced circuity can help planners make resource allocation decisions.

London's Public Transport Network

The London example contextualizes the need for a bus planning framework that can be applied to an existing network to identify opportunities for new services. Like many urban areas, London has experienced growth in recent years. From 2001 to 2011, the population grew at an average rate of 87,000 people per year. The number of jobs in London is projected to increase from 4.9 million in 2011 to 5.8 million in 2036 (*6*).

London has an extensive public transport system that plays a critical role in mobility. Almost half of all Londoners use the bus at least two days per week (*7*). Demand for bus services grew by 64% from 1999 to 2013. However, from 2014 to 2017 bus journeys decreased by 5.4%, correlating with decreases in bus speeds (*8*).

Given the vast size of the London bus network, trial-and-error approaches are inefficient. Instead, a data-driven approach can guide investment. The Oyster card, a smart payment card that was first issued in 2003, was used by approximately 80% of rail passengers and 90% of bus passengers in 2012, generating large quantities of data (*2*).

LITERATURE REVIEW

This paper demonstrates one application using OD data. Other applications are summarized in the following section. Next, a review of the use spatial aggregation in transport analysis is provided, and a summary of traditional network planning and sketch planning techniques.

Use of Automatically Collected Origin-Destination Data

Automated data collection systems in public transport networks have been lauded as a source for better estimates of existing measures and the development of new measures (*9*). Trépanier et al. demonstrated the estimation of a variety of supply and demand-oriented measures from AFC systems (*10*). Uniman et al. developed new metrics for service reliability based on AFC data (*11*). Zhu used AFC data and train tracking data to assign passengers to specific trains (*12*).

Individual itineraries can be inferred from AFC and AVL data (*1*,*2).* In networks where smart card usage is prevalent, this provides a comprehensive picture of public transport travel. Vanderwaart et al. developed a framework for using complete OD-level journey data for service planning (*13*). The framework includes the selection of target locations, evaluation of network and demand characteristics, and proposed service changes. Compared to Vanderwaart's research, the proposed framework generates service alternatives directly from data, rather than relying on interpretation of statistics and maps. This makes the proposed framework more appropriate for large and complex networks. To do so, it focuses on the addition of new bus routes, rather than the diverse service changes considered in Vanderwaart's framework.

Spatial Aggregation

Much of the early work on network structure modeled transport networks as graphs composed of nodes and edges. Garrison and Marble defined summary metrics to assess connectivity of transport networks (*4*). More recently, Derrible and Kennedy used concepts from graph theory to develop indicators of the maturity of a network, the relationship of the network to the built environment, and the network directness (*5*).

For spatial clustering of OD data, methods have been developed based on trajectory clustering. Lee et al. developed a method that clusters similar parts of trajectories (*14*). This methodology was adapted for desire line clustering by Bahbouh and Morency (*15*), and further by Bahbouh et al. (*16*). Their methodology, applied to identify opportunities for walkability improvements, decomposes desire lines into line segments, and imposes maximum width and maximum angle requirements to ensure that clusters form a reasonable travel shed.

The research presented here builds on the methodology developed by Bahbouh et al. However, corridors must be appropriate for bus service, and are prioritized by expected benefits. This requires several adaptations, including minimum and maximum length constraints, the requirement that full desire lines are clustered, and a new prioritization method.

Network Planning and Sketch Planning

This paper demonstrates a network planning approach that differs significantly from traditional approaches. The Transit Network Design Problem is usually formulated as an optimization problem, with methodologies tailored for bus (*17*, *18*), rail (*19*), and for general application to public transport networks (*20*). Most methodologies can be applied to existing or new networks, but are intended for full network redesign and usually recommend drastic changes. As an example, when Cipriani et al. applied their framework to Rome, only 20% of existing routes were retained (*21*).

In existing networks, planners are more likely to employ sketch planning methods, such as the one presented in this paper. Other examples of sketch planning methodologies in the literature include the use of EU standards to evaluate public transport proposals (*22*), and a methodology for demand and cost estimation for proposed extensions of the Bay Area Rapid Transit system (*23*).

The approach in this paper uses a network evaluation process to identify opportunities for new bus routes. One example of the use of performance indicators to locate improvement opportunities comes from highway infrastructure planning in Minnesota, where corridors were prioritized based on demand, growth trends, and connections to regional trade centers, and evaluated based on how closely actual speeds matched target speeds along the corridor (*24*).

Other approaches make recommendations about network design more generally. Daganzo proposed a methodology to determine the appropriate network shape (grid or hub-and-spoke) and modes, given network size and demographics (*25*). Badia et al. compared the performance and cost of different network structures (*26*). Compared to the proposed framework, the recommendations from these methodologies require more interpretation to be translated into planning decisions, particularly in existing networks.

METHODOLOGY

This paper applies a method developed and described in detail in (*27*). Here, an overview of the approach is presented. The framework consists of three steps: zone definition, OD analysis, and corridor identification.

Step 1

The first step in the framework is the definition of geographic zones, which create a set of zonal pairs to which all public transport journeys are assigned, based on their origin and destination bus stop or rail station. This serves to group journeys on multiple paths (including parallel and non-parallel bus and rail lines) that connect similar origins and destinations.

While many cities and regions have existing zonal schemes, such as postcodes or census tracts, these schemes usually use roads as boundaries, resulting in a large share of bus stops along zonal boundaries. Spatially clustering bus stops and rail stations based on their locations results in zones that reflect the structure of the existing public transport system and avoid the location of bus stops and rail stations close to the boundaries of zones. Viggiano et al. (*28*) explains the zone-definition methodology in detail.

Step 2

In an existing network, not all OD pairs can be improved by new service. Some OD pairs are already well-served by the existing network and/or have no anticipated travel time benefits from new service. The first part of the OD-level analysis filters out these pairs. Planners may also manually eliminate OD pairs they know are not good candidates for bus service due to other constraints, such as road network limitations.

The framework uses specific definitions of well-served and improvable OD pairs. OD pairs are defined as well-served if they are served by direct, single-stage service. In this framework, direct and circuitous are used to refer to how closely a path through the road network resembles the shortest possible path. Single-stage always refers to a journey requiring a single vehicle. It can refer to one bus stage (one boarding, one alighting) or one rail stage (one entry and one exit without any in-station or out-of-station interchange). Of those OD pairs that are not well-served, improvable OD pairs are defined as those that are expected to see travel time improvements from new service.

Following the filtering step, the potential benefits of new service are quantified for improvable OD pairs. The benefits of new service are twofold. There are travel time savings accumulated by existing public transport passengers who shift to the new service, and there are new passengers attracted from outside the public transport system. This includes passengers who shift from car or other modes and new journeys that previously were not made. Individuals may choose to make more trips due to improved service, and over longer time periods, individuals may change their home and work locations in response to new services. Expected new demand depends on many factors including the change in bus travel time, the demand and performance of alternative modes, and socio-demographic and land use factors. Predicting demand changes is challenging, both because reliable data on these factors is not always available and because of complexity and variation in how these factors influence individual behavior. The best estimates available for the planning time horizon desired will produce the most insightful results.

Step 3

The final step in the framework is to spatially aggregate improvable OD pairs into corridors. New services will not influence OD pairs in isolation. Rather, they will provide benefits along a linear (or pseudo-linear) corridor. Step 3 clusters OD pairs into corridors that can be served by a single bus route and have demand characteristics that can support new service. For each corridor, expected benefits are defined as the weighted sum of travel time savings for existing public transport passengers and the total journey-minutes of new journeys attracted from outside the public transport network. Depending on the planer perspective, different definitions of benefits could be applied. For example, economic benefit analysis often uses the "rule of half" to quantify the benefits from behavioral responses to cost or travel time changes, and a similar approach could be adopted. Here, benefits are used only to prioritize corridors within the corridor identification algorithm. Therefore, the travel time savings for existing public transport passengers and the total journey-minutes of new journeys attracted are simply summed.

Within the framework, corridor definitions can incorporate varying degrees of spatial specificity and complexity. For instance, corridors can be represented simply as straight lines, they can bend to account for barriers such as rivers and parks, or they can reflect actual paths through the road network. Similarly, the algorithm used to identify corridors can incorporate road network distance and barriers when assessing the similarity of OD pairs, or can use Euclidean distance as a simple proxy.

LONDON IMPLEMENTATION AND RESULTS

This section summarizes the process and results of applying this three-step framework to the London public transport network.

Data

The inputs to the framework are AFC and AVL data, which provide OD level demand, distance, and travel time estimates within the public transport network. Also required are OD level shortest path distance and travel time estimates through the road network, which are used to estimate potential travel time and distance for new bus routes. The fare card data used is from 10 weekday AM peak periods from October 2012 and February 2013. The AFC and AVL data were combined to infer origin bus stops, alighting bus stops, and link stages of multi-stage journeys using a methodology developed by Gordon et al. (*2*).

Estimates of shortest path distances and travel times for OD pairs were queried using the Google Maps Distance Matrix API. The API provides estimates of point-to-point car distance and travel time for a specified time of day, based on historical data, thus accounting for expected traffic conditions. A study comparing Google Maps estimates of travel times to estimates from GIS networks found the estimates to be reasonably consistent (*29*).

Step 1: Zones

The process of developing appropriately-sized zones that reflect the structure of the existing public transport system in London is described in detail in Viggiano et al. (*28*). The result is 1,000 zones (shown in Figure 1) with an average radius of approximately 0.7 kilometers, which corresponds to the average access distance to public transport (bus and rail) stops and stations, and to the 75th percentile of access distances specifically to bus stops in London (*30*). By spatially clustering bus stop and rail station locations, the resulting zones have fewer bus stops and rail stations along zonal boundaries, compared to existing London zonal schemes.

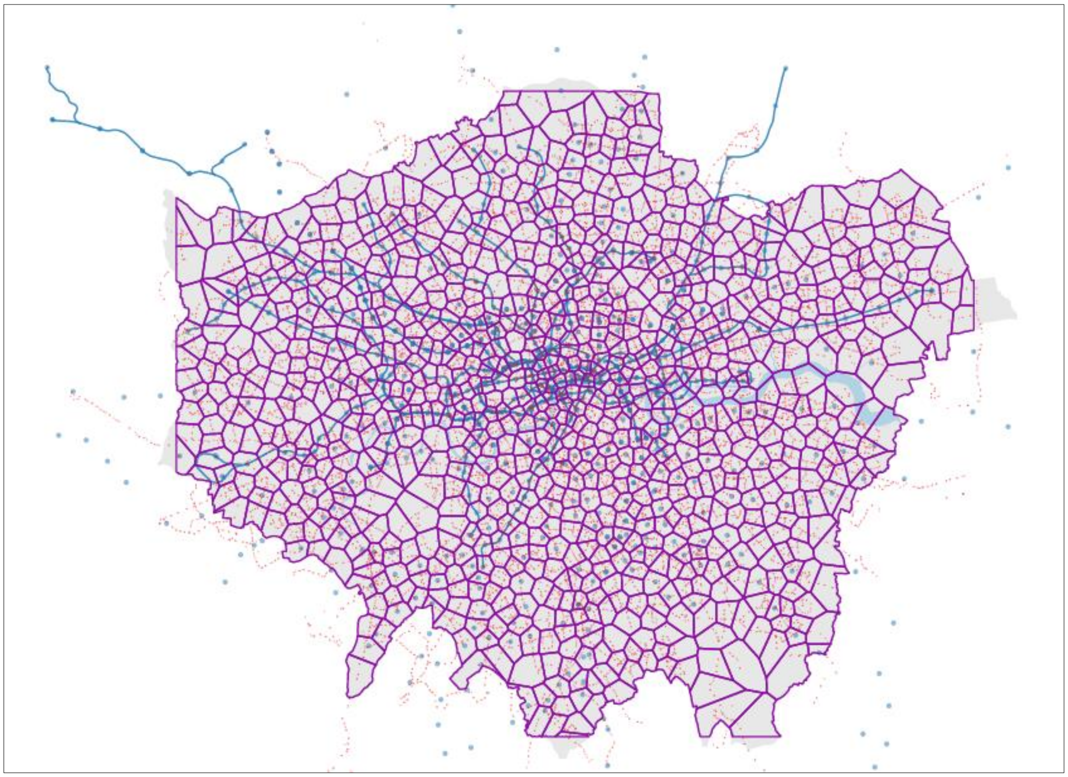

FIGURE 1 London zones.

Step 2: Improvable OD Pairs

In London, the peripheral region is more likely to have opportunities for new services. OD pairs that originate or end in Fare Zone 1 (Central London) are excluded from the OD analysis. This leaves 950 zones and 902,500 zonal OD pairs. 243,076 OD pairs have at least one journey in the 10 days of AM peak data analyzed. A minimum of 23 journeys over the 10 AM peaks analyzed, or just under 1 journey per AM peak hour is required to ensure travel times can be reliably estimated. Only 45,916 OD pairs have at least 23 journeys, demonstrating the sparsity of the OD matrix. The impact of this constraint on the percent of journeys included in the analysis is small, removing 16% of journeys.

As described in the methodology section, OD pairs are considered well-served by the existing network if they are served by direct, single-stage service. For OD pairs served by bus, this requires comparing the bus distance traveled in the current network to the shortest path distance through the road network.

Shortest paths were estimated using the full road network except for highways. The actual path of a potential new bus route is not expected to follow the shortest path exactly. Paths that pass important locations such as schools or hospitals may be desirable, and some roads and junctions cannot accommodate buses. Therefore, OD pairs are considered to have direct service if the current bus distance is less than or equal to the shortest path distance, increased by a multiplicative factor, m, and an additive factor, a. m reflects additional distance that is proportional to the shortest path distance for the OD pair. a reflects fixed additional distance that is independent of shortest path distance. Planners may use these parameters to specify the standard for directness they wish new bus routes to meet.

For the London application, relatively large values of m and a were used: 1.1 and 1.4

kilometers (0.9 miles), respectively. The choice of *m* and *a* was based on analysis of the directness of existing bus routes, with larger values selected to focus the analysis on OD pairs with significant potential for improvement in terms of directness. Of those OD pairs served by single-stage bus, 91% are considered well-served, given these parameters. The remaining 9% (1,737 OD pairs) are not well-served in terms of directness. In addition, there are 17,942 OD pairs that are served only by multi-stage service.

The next filtering step determines which of these OD pairs are improvable in terms of travel time. Here, the median travel time in the existing public transport network is compared to the expected travel time. The expected travel time is based on the shortest path travel time, the expected number of stops, and the estimated time required to make a stop. Again, this estimate can be modified with multiplicative or additive factors to reflect acceptable deviations from the shortest path.

Transport for London planners aim for a stop spacing of approximately 400 meters (0.25 miles) (*31*). Dwell time varies based on the number of boardings and alighting as well as characteristics of the bus and stop. Using the middle of the ranges estimated by York (*32*), and the average acceleration/deceleration time found by Robinson (*33*), 33 seconds are added for each stop. In addition, a multiplicative factor of 1.1 was applied.

Given these parameters, of the 1,737 OD pairs served by single-stage bus and identified as not well-served, 1,621 have travel times greater than the expected travel time for the OD pair. This is reasonable; most OD pairs served by circuitous bus service are improvable in terms of travel time. For OD pairs served by multi-stage service, the travel time filter has a much greater impact. Of 17,942 OD pairs served only by multi-stage service, 5,856 OD pairs have travel times that exceed the pair’s expected travel time. This is also unsurprising; many of these OD pairs are served by multi-stage rail, which is often faster than bus.

For the improvable OD pairs, the final part of the OD analysis estimates the demand a new bus service is expected to attract. There are many methods for predicting demand, including four-step models, direct demand models, discrete choice models, and elasticity-based methods (*34*). Elasticity methods require only two inputs, which are already used in the analysis: existing demand and anticipated travel time savings. Therefore, the elasticity method was used, eliminating additional data requirements. Based on a review of studies in the UK, elasticities of demand with respect to bus in-vehicle time of -0.4, -0.5, and -0.6 were applied for short trips (less than 20 minutes), medium trips (20 to 40 minutes) and long trips (more than 40 minutes), respectively (*35*). The varied elasticities account for the fact that for shorter trips, in-vehicle time generally makes up a smaller percentage of the total journey time. This method does not account for demand changes due to contemporaneous population or land use changes, and it may underestimate demand on OD pairs where current public transport demand is low due to poor existing service.

Step 3: Corridors

The corridor identification methodology applied was developed by Viggiano (*27*). It uses simplified spatial representations. Corridors are defined as a straight line with a surrounding buffer. Consideration of barriers and constraints on termini location are left to post-analysis. The effectiveness of this approach is discussed with respect to the example corridor below.

The methodology requires the specification of a minimum and maximum route length, a maximum distance, a maximum angle, and a minimum flow. The length parameters restrict the length of the corridors identified to ensure they are reasonable for bus service. For this application, based on the lengths of existing bus routes in London, corridors 5 to 9 miles in (straight-line) length are considered. Actual bus routes are expected to be significantly longer than this straight-line distance.

The maximum distance parameter constrains the Euclidean distance from zonal OD pair centroids to the straight line that represents the center of the corridor. Because zone size was set based on public transport access distances, ideal corridors should be one zone wide, i.e. with all centroids of assigned zonal OD pairs aligned along the linear corridor. In fact, zone centroids do not always align, so choosing a maximum distance that is too small eliminates many potential corridors. Through trial-and-error, it was determined that specifying a maximum distance of 0.6 kilometers (0.4 miles) ensures that corridors are mostly only one zone in width, without eliminating too many potential corridors.

Some short OD pairs may have origins and destination zone centroids within the maximum distance, but if the direction of travel on the OD pair and the route is not similar, the potential route would not serve the OD pair. To eliminate these OD pairs, angular similarity is defined as in Figure 2, by computing the minimum angle formed by the connected centroids of the OD pair and the potential linear corridor. If this angle is within the maximum angle specified, OD pairs can be assigned to a potential corridor. A maximum angle of 22.5 degrees was applied.

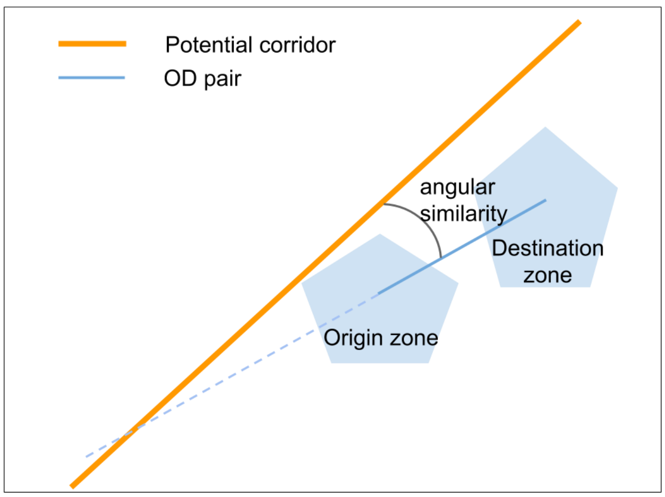

FIGURE 2 Angular similarity

Corridors identified should have sufficient expected demand to justify adding bus service. This is enforced through a minimum flow constraint. Flow on a potential corridor can be estimated as the expected hourly flow averaged along the corridor. This measure does not fully capture the shape of demand, such as spatial peaking, but it does account for flow relative to the length of the route, and it is easy to understand. For this implementation, a moderate minimum flow of 50 passengers per hour is used. This accounts for the fact that the methodology likely underestimates demand. More detailed demand estimation can be conducted at the corridor level in post-analysis.

Using these parameters, 11 corridors are identified (See Figure 3). Each corridor is prioritized by the sum of potential journeys-minutes saved (for existing passengers) and new journey-minutes induced; both benefits are estimated per AM peak hour. Only 5% of improvable OD pairs (9% of journeys on improvable OD pairs) are assigned to a corridor. This reflects the fact that the improvable OD pairs are spatially distributed across the network. The parameters selected also influence the number of corridors identified and the percent of OD pairs and journeys assigned to corridors. Reducing the minimum flow identifies more corridors and assigns more OD pairs and journeys to corridors. However, these corridors may not have sufficient demand to support bus services. Increasing the maximum distance or angle also results in more corridors being identified, as loosening these shape requirements allows more corridors to meet the minimum flow requirement. However, with less stringent shape requirements, it may be difficult to serve the entire corridor with one bus route. Overall, sensitivity analysis of the resulting corridors for varied shape and flow parameters revealed that the highest priority corridors were identified consistently across parameter choices.

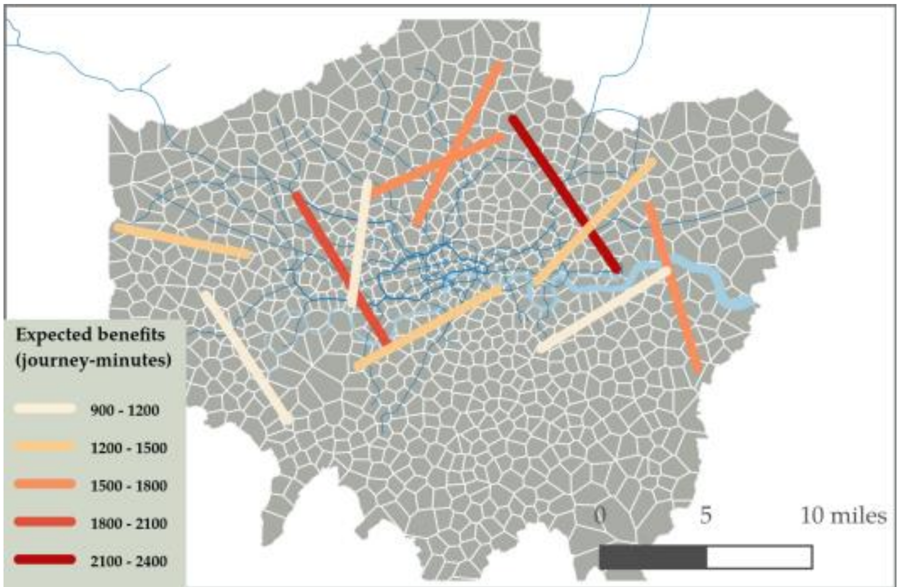

FIGURE 3 Corridors identified.

The results are also influenced by parameter choices in Step 2 of the framework. Reducing the multiplicative and additive factors, and thus decreasing expected distance and travel times, identifies more improvable OD pairs. This increases the spatial density of OD pairs to be clustered into corridors and results in more corridors being identified.

Example Corridor

The corridor discussed here is the second highest priority corridor, extending from Wembley Park to Fulham. In addition to having significant expected benefits, this corridor has the greatest expected demand of the corridors identified. Almost 60% of the current journeys on the improvable OD pairs assigned to the corridor are multi-stage bus journeys. Approximately 7% of the journeys are multi-stage rail journeys (defined as rail journeys in which the origin and destination stations are not served by the same line or rail journeys with out-of-station transfers) and 15% are combined bus and rail journeys. The remaining journeys are served by single stage bus that did not meet the directness threshold.

The expected benefits on this corridor, consisting of potential travel time savings for current passengers and anticipated new journey-minutes, are distributed over 41 OD pairs. Approximately 75% of the benefits, however, are concentrated on just 15 OD pairs shown in Figure 4. In the direction of Fulham Broadway, demand is distributed along the corridor and several long OD pairs are expected to receive significant benefits. The largest flows are in the opposite direction in the northern part of the corridor. On the three OD pairs ending in Wembley Park, there are 59 journeys expected per AM peak hour, and the expected travel time savings relative to current service are between 6 and 10 minutes.

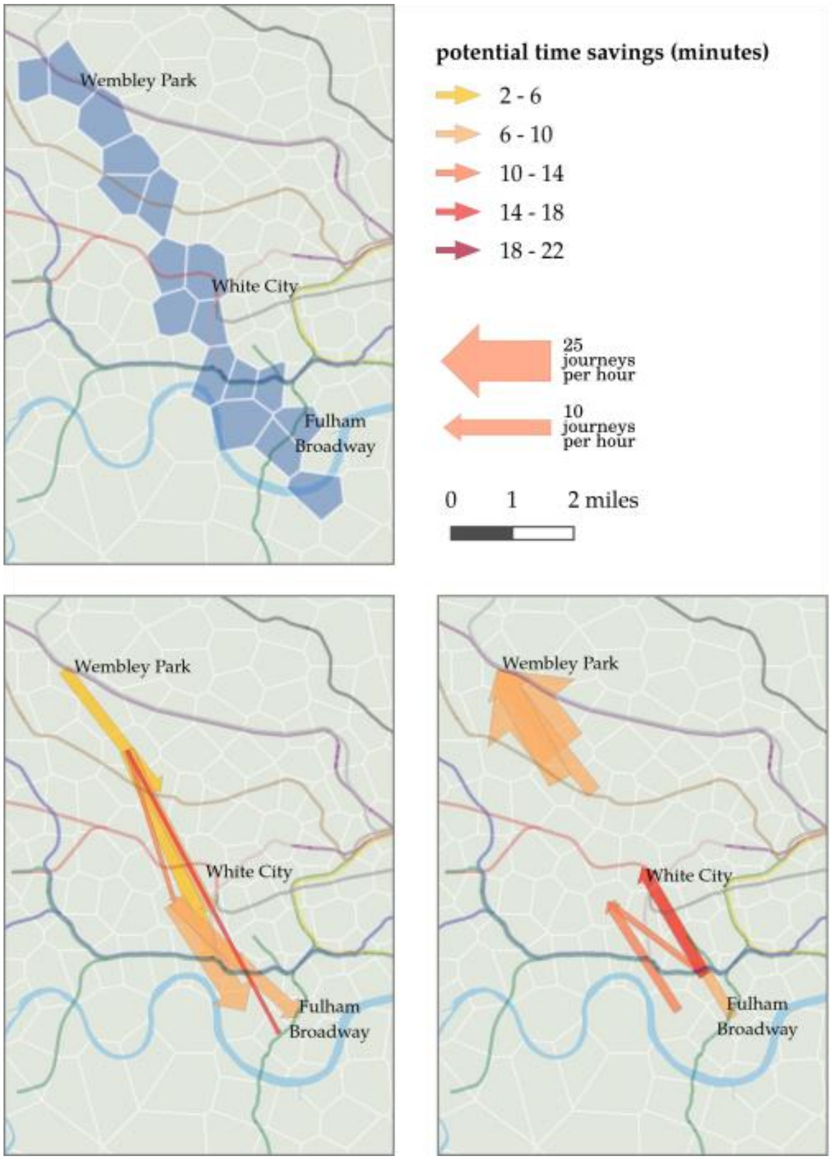

FIGURE 4 Demand and potential travel time savings in example corridor.

Figure 5 shows the main bus routes serving the corridor, which serve 80% of the demand on the OD pairs shown in Figure 4. Route 206 traverses the northern part of the corridor, but is circuitous. Route 220 serves the middle part of the corridor. However, several of the OD pairs included in the corridor begin north of the terminus of Route 220 and extend south, requiring either multi-stage journeys or circuitous travel on Route 266 or Route 260. Extending a single route the length of the corridor would serve these journeys better.

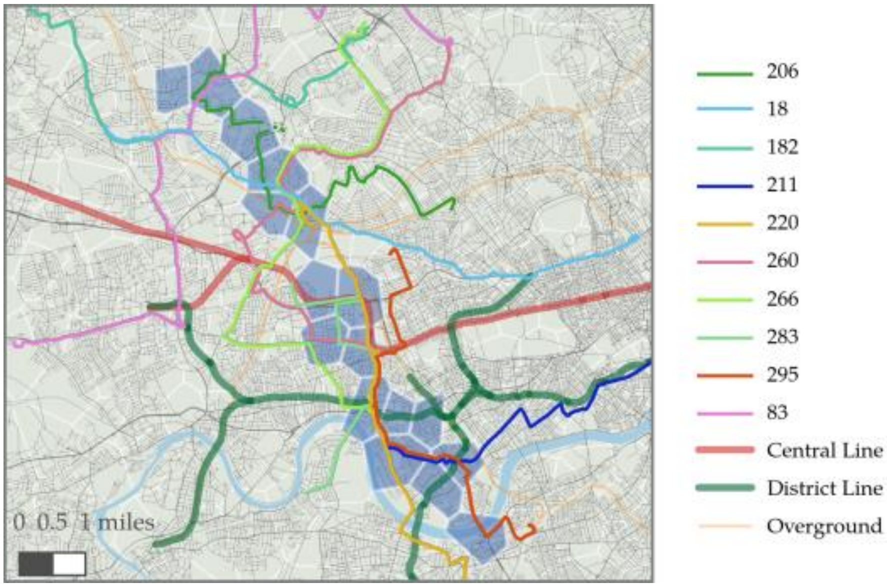

FIGURE 5 Existing bus routes and rail services in example corridor.

The southern half of the corridor is two zones wide in many places. As described in Step 3: Corridors, this is because corridors are linear, but zone centroids are not always aligned. The corridor width poses challenges to designing a route that serves all OD pairs. Figure 6 shows a potential design for a route. In the southern part of the corridor it serves the western side of the corridor more so than the eastern, as the eastern side has more poorly-served demand.

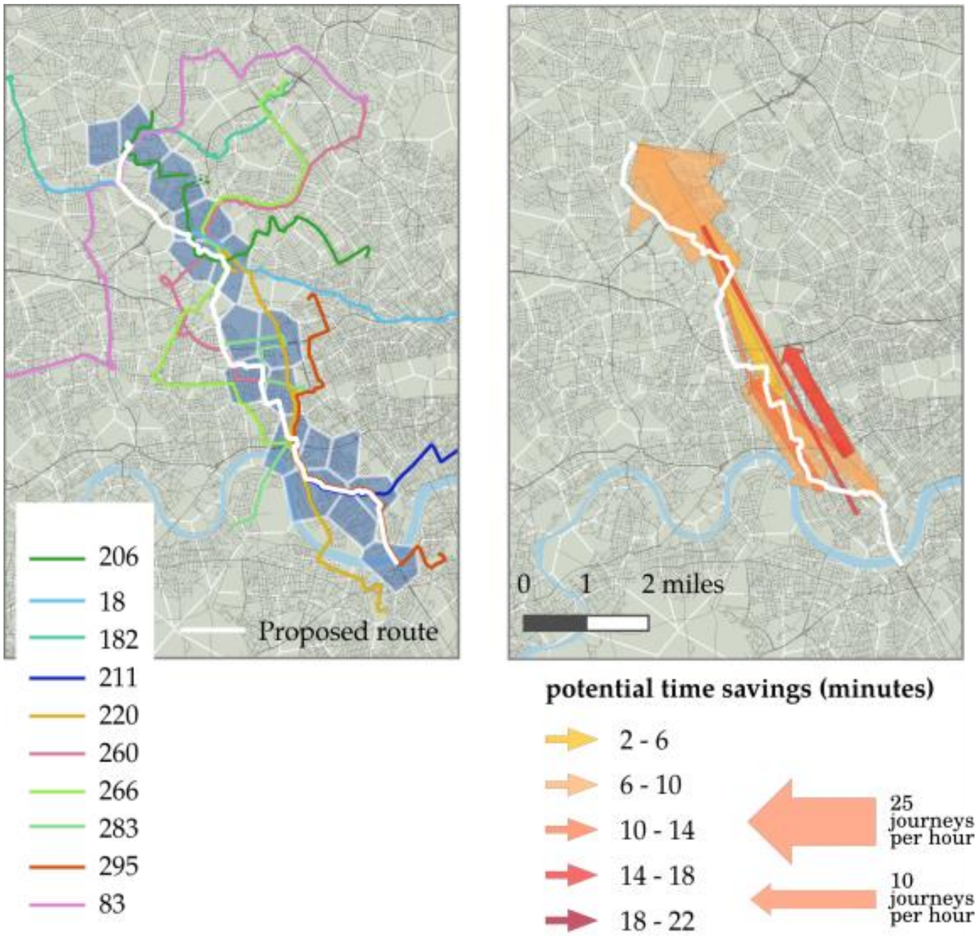

FIGURE 6 Potential new route design for example corridor.

DISCUSSION

The implementation of the framework reveals two possible areas for improvement. The first is the choice of spatial representation. More complex representation of corridors as actual paths through the road network, and the use of road network distance in place of Euclidean distance in the corridor identification algorithm could limit cases as in the southern part of the example corridor, where the corridor width and characteristics of the road network make it difficult to serve all OD pairs with a single route. The second area for improvement is the incorporation of a more sophisticated demand model to better account for new demand on OD pairs that are currently not well-served.

A large percentage of improvable OD pairs were not assigned to any corridor. These OD pairs are not good candidates for improvement through new bus services because they are not located within a corridor with sufficient demand for new bus service. In networks where a small percentage of OD pairs identified for improvement are assigned to corridors, planners might consider services other than fixed route bus for these OD pairs.

Two extensions to the framework may also be useful to Transport for London and other agencies. The first is to allow for new routes to provide improved service while still allowing a transfer, which may be appropriate in areas of sparser demand. This would require changes to the corridor identification algorithm such that demand could be served by a combination of two routes.

The second extension would be to provide recommendations for existing routes that should be removed. Given good estimates of how journeys shift from existing to new services, an iterative approach could be taken. In each iteration, routes with insufficient demand would be eliminated and the analysis would be repeated for the new network.

CONCLUSION

This paper presents an innovative application of OD data for bus network evaluation and planning. Spatial aggregation approaches, such as those applied in this paper, can help planners understand characteristics of transport networks and the quality of service they provide. In this case, the analysis evaluates OD-level circuity and helps planners identify opportunities for new bus services.

The success of this approach suggests that the application of spatial and quantitative methods to OD data can be adapted to plan other modes, such as on-demand service. In addition, similar approaches can be applied to other planning challenges, such as decisions about where to implement bus priority measures, and where to add express bus services. For bus priority measures, metrics and estimates to assess potential benefits must be developed, and spatial methods are needed to aggregate potential benefits to the corridor level. For express service, the demand patterns that are generally associated with this type of service, such as long journeys and concentrated OD-level demand would need to be identified through OD analysis and spatial aggregation. The framework applied here can form the groundwork for the development of new metrics assessing transportation system performance, mobility, and accessibility, at the OD and corridor level.

More broadly, with the development of appropriate methods, OD data is likely to play a growing role in transportation policy and planning. OD data is becoming more plentiful and is already supporting an improved understanding of travel behavior and system performance. The strategic definition of zonal OD pairs, development of appropriate metrics, and spatial aggregation of disaggregate data, as demonstrated here, allow planners to use the improved understanding for better planning.

ACKNOWLEDGEMENTS

This research received funding from Transport for London and support from the staff of London Buses and TfL Analytics.

REFERENCES

1. Chu, A.K.K. and R. Chapleau. Enriching archived smart card transaction data for transit demand modeling. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2063, 2008, pp. 63–72.
2. Gordon, J. and H. Koutsopoulos, N. Wilson, and J. Attanucci. 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, 2013, pp. 17–24.
3. Huang, J. and D.M. Levinson. Circuity in urban transit networks. *Journal of Transport Geography*, Vol. 48, 2015, pp. 145–153.
4. Garrison, W.L. and D.F. Marble. Factor-analytic study of the connectivity of a transportation network. *Papers in Regional Science*, Vol. 12, No. 1, 1964, pp. 231–238.
5. Derrible, S. and C. Kennedy. Characterizing metro networks: state, form, and structure. *Transportation*, Vol. 37, No. 2, 2010, pp. 275–297.
6. Greater London Authority. The London plan: The spatial development strategy for London consolidated with alterations since 2011. Technical report, 2016.
7. Transport Committee of the London Assembly. *Bus services in London*. Technical report, Greater London Authority, 2013.
8. Transport for London. *Transport for London Budget 2017/18*. Technical report, 2017.
9. Wilson, N.H.M., J. Zhao, and A. Rahbee. The potential impact of automated data collection systems on urban public transport planning. In *Schedule-based modeling of transportation networks*, Springer, Boston, MA, 2009, pp. 1-25. https://doi.org/10.1007/978-0-387-84812-9
10. Trépanier, M., C. Morency, and B. Agard. Calculation of transit performance measures using smartcard data. *Journal of Public Transportation*, Vol. 12, No.1, 2009, pp. 79-96.
11. Uniman, D., J. Attanucci, R. Mishalani, and N. Wilson. Service reliability measurement using automated fare card data: Application to the London underground. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2143, 2010, pp. 92–99.
12. Zhu, Y., H.N. Koutsopoulos, and N.H.M. Wilson. A probabilistic Passenger-to-Train Assignment Model based on automated data. *Transportation Research Part B: Methodological* (2017). Forthcoming. https://doi.org/10.1016/j.trb.2017.04.012
13. Vanderwaart, C., J.P. Attanucci, and F.P. Salvucci. Applications of Inferred Origins, Destinations, and Interchanges in Bus Service Planning. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017.
14. Lee, J., J. Han, and K. Whang. Trajectory clustering: a partition-and-group framework. *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, 2007, pp. 593–604.
15. Bahbouh, K. and C. Morency. Encapsulating and visualizing disaggregated origin-destination desire lines to identify demand corridors. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2430, 2014, pp. 162–169.
16. Bahbouh, K., J.R. Wagner, C. Morency, and C. Berdier. Traclus-dl: A desire line clustering framework to identify demand corridors. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.
17. Ciaffi, F., E. Cipriani, and M. Petrelli. Feeder bus network design problem: a new metaheuristic procedure and real size applications. *Procedia-Social and Behavioral Sciences*, Vol. 54, 2012, pp. 798–807.

18. Bielli, M., M. Caramia, and P. Carotenuto. Genetic algorithms in bus network optimization. *Transportation Research Part C: Emerging Technologies*, Vol. 10, No. 1, 2002, pp. 19–34.
19. Guan, J.F., H. Yang, and S.C. Wirasinghe. Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B: Methodological*, Vol. 40, No. 10, 2006, pp. 885–902.
20. Bagloee, S.A. and A. Ceder. Transit-network design methodology for actual-size road networks. *Transportation Research Part B: Methodological*, Vol. 45, No. 10, 2011, pp. 1787–1804.
21. Cipriani, E., S. Gori, and M. Petrelli. Transit network design: A procedure and an application to a large urban area. *Transportation Research Part C: Emerging Technologies*, Vol. 20, No. 1, 2012, pp. 3–14.
22. Cascetta, E. and A. Cartenì. A quality-based approach to public transportation planning: theory and a case study. *International Journal of Sustainable Transportation*, Vol. 8, No. 1, 2014, pp. 84–106.
23. Deakin, E., M. Payne, and V. Menotti. Development of bay area rapid transit system expansion criteria and process. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1887, 2004, pp. 18–25.
24. Zemotel, L, and D. Montebello. Interregional corridors: Prioritizing and managing critical connections between Minnesota’s economic centers. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1817, 2002, pp. 79–87.
25. Daganzo, C.F. Structure of competitive transit networks. *Transportation Research Part B: Methodological*, Vol. 44, No. 4, 2010, pp. 434–446.
26. Badia, H., M. Estrada, and F. Robusté. Bus network structure and mobility pattern: A monocentric analytical approach on a grid street layout. *Transportation Research Part B: Methodological*, Vol. 93, Part A, 2016, pp. 37 – 56.
27. Viggiano, C. Bus Network Sketch Planning with Origin-Destination Travel Data. PhD Thesis, Massachusetts Institute of Technology, 2017.
28. Viggiano, C, H. N. Koutsopoulos, NHM Wilson, and J. Attanucci. Journey-based characterization of multi-modal public transportation networks. *Public Transport* Vol. 9, No. 1-2, 2017, pp. 437-461.
29. Wang, F. and Y. Xu. Estimating O–D travel time matrix by Google Maps API: implementation, advantages, and implications. *Annals of GIS* Vol. 17, No. 4, 2011, pp. 199-209.
30. Viggiano, C., H.N. Koutsopoulos, J. Attanucci, and N.H.M. Wilson. Inferring public transport access distance from smart card registration and transaction data. Transportation Research Record: Journal of the Transportation Research Board, No. 2544, 2016, pp. 55–62.
31. Bus Priority Team, Transport for London. *Accessible bus stop design guidance*, 2006.
32. York, I.O. Factors affecting bus-stop times. *TRRL Project Report 2*. Technical report, Transport and Road Research Laboratory, Crowthorne, 1993.
33. Robinson, S. Measuring bus stop dwell time and time lost serving stop with London iBus automatic vehicle location data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2352, 2013, pp. 68–75.
34. Ortuzar, J. and L.G. Willumsen. *Modelling transport*. John Wiley & Sons, 2011.
35. Wardman, M. "Review and meta-analysis of UK time elasticities of travel demand." *Transportation* Vol. 39, No. 3, 2012, pp. 465-490.