`

Optimal Tube station locations with Linear Programming: The Bakerloo Line Extension case study

According to Transport for London’s report ‘Strategic Case for Metroisation in South and Southeast London’ (TfL, 2019), South London residents are missing out on opportunities because of inadequate public transit connectivity, with access to employment drops significantly in gaps between the few rapid transit lines that serve it such as the Northern Line, The Docklands Light Rail (DLR) and the London Overground – East London Line. In fact, there are four times as many jobs within 45 minutes of Harrow (North London) compared to Sutton (South London). Consequently, residents are more likely to switch from public transit to personal automobiles for their commute.

This necessitated the proposal by the Mayor of London to extend the Bakerloo Line from its current terminus at Elephant & Castle to Lewisham, connecting it to Tube and other rail services together with two new stations along Old Kent Road (Figure 1). The Bakerloo Line Extension (BLE) is, therefore, one major component of a larger strategy to revitalise South London and improve its residents' lives.

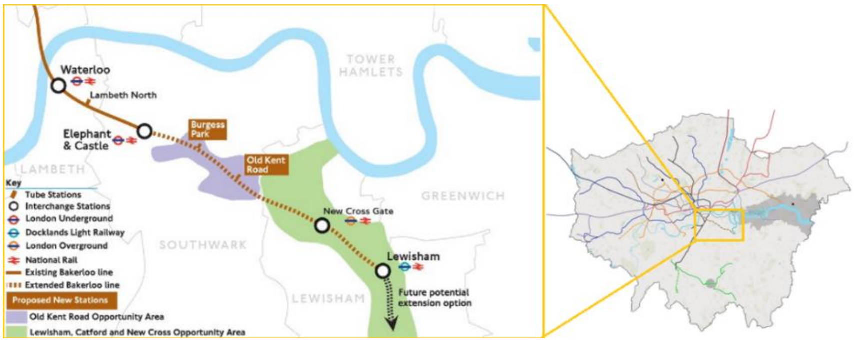

Figure 1 - Proposed Bakerloo Line Extension and station alignment (Source: TfL)

With the BLE as the subject, this paper will apply different Linear Programming models to determine the number and location of new stations along the new rail corridor needed to facilitate pedestrian access by reducing the time required to reach stations. Finally, if found, the optimal solution for chosen station locations will be compared to TfL’s proposal to surface potential new location candidates.

Methodology

Linear programming is a well-established methodology in transport planning, from service and employee scheduling (Gavish and Shlifer, 1979) to location planning (Jafari and Yaghini, 2019). In location planning, one of the main objectives is to ensure that where the stations are located positively affects the travel experience compared to the status quo within the financial, technical, and operational constraints that all infrastructure projects must contend with. In this study, the travel experience we seek to optimise will be how quick and easy it is for residents to access the new stations, one of many factors that motivate individuals to change their commute behaviour (Anwar, 2012). Empirical research has shown that urban residents are only willing to walk up to 10 to reach a rapid transit station (Sarker, Mailer and Sikder, 2019); therefore, this value will be used as our benchmark for later evaluation of the model

A theoretical framework for determining station sites was outlined in detail by Hamacher et al. (2001), who defined two main objectives of this endeavour:

1. *Accessibility:* All or a predetermined share of the transit demand must be covered by at least a station. In the case of a single line segment, the problem will resemble the classic Location Set Covering Problem.
2. *Travel Time:* The station locations must also minimise both the time spent reaching them on foot and the incremental delays to passengers onboard with each additional stop added.[1]

This framework forms the foundation for our methodology, which involves formulating these two objectives as linear programming problems: First, to minimise stations built to cover all demand, the Location Set Covering Problem was used (Church and Murray, 2018). Secondly, an adapted P-Median Problem formulation (Hakimi, 1965) was used to minimise the walking time. This was chosen over the Maximum Coverage Location Problem for its focus on explicitly minimising population-weighted cost (or, in this context, walking time) rather than maximising population-weighted coverage (Karatas, Razi and Tozan, 2016).

Problem 1 – *Minimise number of stations built (LSCP)*

$$Min\ S = \sum_{j \in J} Y_j \tag{1}$$

such that

$$\sum_{j \in N_i} Yj \geq 1 \qquad \forall i \in I \tag{2}$$

$$Y_j \in \{0,1\} \quad \forall j \in J \tag{3}$$

where:

$i \in I$: index of demand points (neighbourhoods)

$j \in J$: index of candidate station locations

$N_i = \{j | d_{ij} \leq T\} \subseteq J$: set of stations within J within a travel time T of neighbourhood i, with:

- d_{ij} : shortest travel time from each i to each j
- T : maximum time needed to reach a station (service radius)

$Y_j \in \{0, 1\}$: binary, 1 if station j built, and 0 otherwise (decision variable)

[1] Note that, for the sake of simplicity, in our model, we will assume potential transit riders will only walk to stations instead of driving or taking the bus. The passenger time delay component will also be omitted.

Problem 2 – *Minimise total walking time to station (P-Median problem)*

$$Min\ W = \sum_{i \in I, j \in J} a_i d_{ij} X_j \tag{4}$$

such that

$$\sum_{j \in J} Y_j \le k \tag{5}$$

$$\sum_{j \in J} X_{ij} = 1 \qquad \forall i \in I \tag{6}$$

$$X_{ij} \le Y_i \qquad \forall i \in I\ \forall j \in J \tag{7}$$

$$X_{ij} \in \{0,1\} \quad \forall i \in I\ \ \forall j \in J \tag{8}$$

$$Y_j \in \{0,1\} \quad \forall j \in J \tag{9}$$

where:

$i \in I$: index of demand points (neighbourhoods)
$j \in J$: index of candidate station locations
d_{ij} : shortest travel time from each i to each j
a_i : population at i
k : predefined number of stations to be located
$X_{ij} \in \{0, 1\}$: binary, 1 if assign demand i to station j, and 0 otherwise (decision variable)
$Y_j \in \{0, 1\}$: binary, 1 if station j built, and 0 otherwise (decision variable)

The mathematical formulation above may be verbally interpreted as follows:

(1)(4) Objective functions for the two problems.
(3)(8)(9) Binary constraints for the decision variables.
(2) Every neighbourhood i is within a max walking time T of min 1 station.
(5) Maximum k stations can be assigned.
(6) Each neighbourhood i is assigned only one station.
(7) Each neighbourhood i is assigned to station j only if it's built.

A note on our approach: As with most real-life problems having interlocking objectives, it is often a recommended approach to formulate a Multi-objective Mathematical Programming (MMP) problem, commonly seen in transit planning research using linear programming (Benli and Akgün, 2023). With MMP, the objective functions are solved in interaction with the others, producing a final set of Pareto-efficient solutions (i.e., 'Pareto front'), from which decision-makers can weigh different options, each with its trade-off on one objective or another. (Chen and Zhou, 2022).

Due to our limited scope, we will instead explore a snapshot of the optimal solution universe by varying a key input parameter per problem within a manually set range. Such a brute-force approach would otherwise not be suitable for complex problems with more objective functions.

For our model, we will explore all optimal solutions to **Problem 1** by varying T (max walking time) from 0 to 3600 seconds (0 to 60 minutes), and all optimal solutions to **Problem 2** by varying k (max number of stations that can be built) from the minimum numbers of must-built stations to the total number of location candidates.

Data preparation

The neighbourhood and candidate location sets, I and J, respectively, were acquired from the following workflow[2]:

1. Create a linestring for the corridor connecting Elephant & Castle with Lewisham, mainly following Old Kent Road (7.5 km)
2. Create a set of points 250m apart along the linestring as candidate stations. $|J| = 31$
3. Create the set of neighbourhoods (demand points) from the set of the Output Area centroids that intersect within a 1 km buffer area of the corridor. $|I| = 687$
4. Extract the population of all points in I to population a_i
5. Calculate the walking time of all i and j pairs, or d_{ij}, using the OpenStreetMaps pedestrian routing server.

For BLE, we also designated candidate locations that must be included in the solution to allow connection with other lines at Elephant & Castle, New Cross Gate, and Lewisham stations. The candidates closest to the three locations above are $j \in \{1, 21, 31\}$. Therefore, we added the ad-hoc constraint (10) :

$$Y_{1,21,31} = 1 \qquad (10)$$

The resulting sets are visualised in Figure 2.

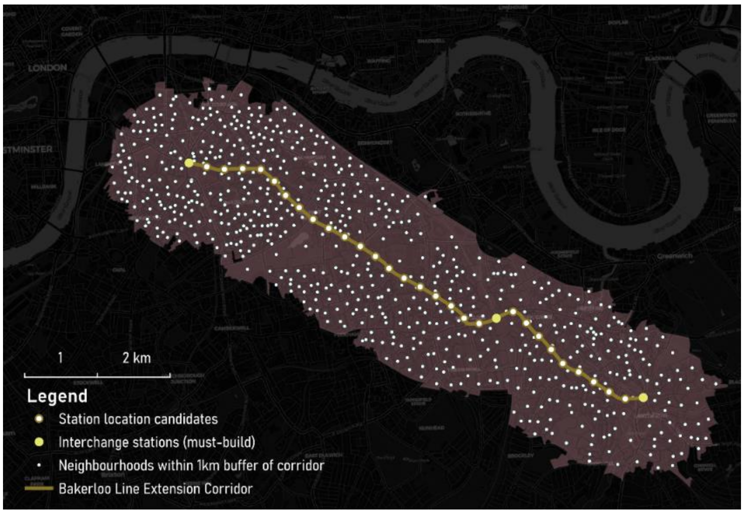

Figure 2 - Station candidate and neighbourhood sets

Finally, we need to address some oversimplifications made so far:

- The neighbourhood location set I does not consider the propensity to use public transit, future transit demand, or local politics that might stipulate certain neighbourhoods to be included in (or excluded from) the catchment area, such as the designated Opportunity Areas in Southeast London. (*London City Hall*, 2023)
- The station candidate location set J does not consider engineering feasibility, assuming all sites can accommodate an underground train station.

[2] Data source: Office of National Statistics, Bing Maps

- Different routing services, such as that offered by Google or Mapbox, may yield a different cost matrix and, thus, different solutions to the problems.

To solve the two optimisation problems, we used the COIN-OR Linear Program solver deployed with `pulp` and `spopt` (specialised Python libraries for linear program and spatial optimisation, respectively)[3]

Results

Figure 3 shows all optimal solutions for ***Problem 1*** *(Minimise station)* at different T values (i.e., max walking time). We can see that, if T < 1800 seconds (30 minutes), the problem is unsolvable. Therefore, if an average walking time of 10 minutes is used as a benchmark, there are no feasible solutions.

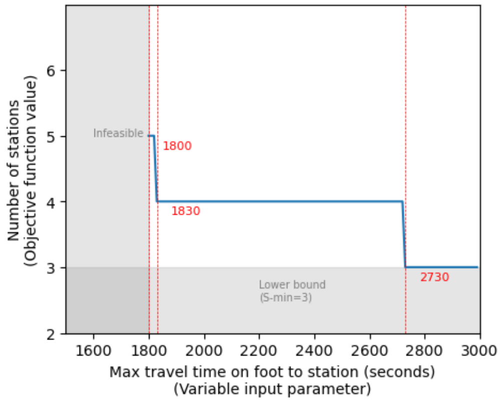

Figure 3 - Minimum stations needed at varying max walking time (Problem 1)

Figure 4, on the other hand, shows all optimal solutions to ***Problem 2*** *(Minimise walking time)* at different k values (i.e., number of stations). The problem is solvable at all k values between 3 and 31, with a 'knee' at $k = 7$, where the optimal solution yielded 586 seconds (~10 minutes) of average walking time[4]. Based on the external benchmark of 10 minutes, we can declare this solution where $k = 7$ to be selected.

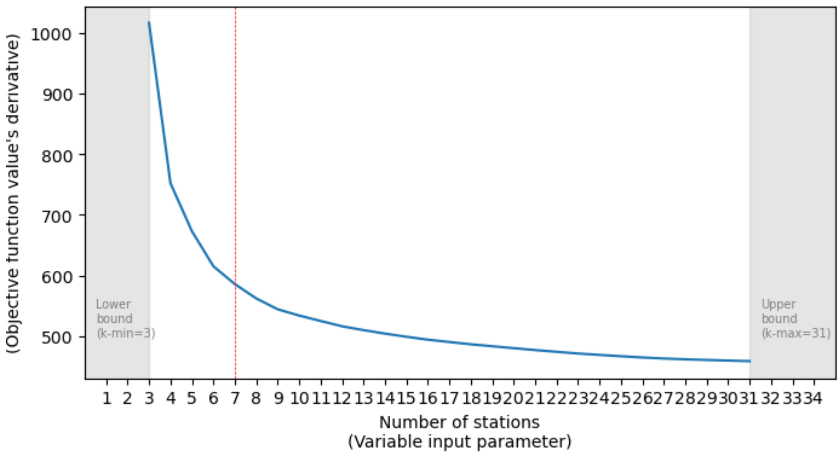

Figure 4 - Minimum average walking time at varying numbers of stations given (Problem 2)

---

[3] The codes used were based on PySAL library's tutorials: https://pysal.org/spopt/tutorials.html and can be found in the GitHub repository: https://github.com/hanukikanker/bakerloo-ext-lp

[4] Avg. walking time equals total walking time (objective function) divided by total population.

Discussion

When the chosen locations (k=7) are juxtaposed with the official proposal for the BEL, as shown in Figure 5, we can make several observations and recommendations:

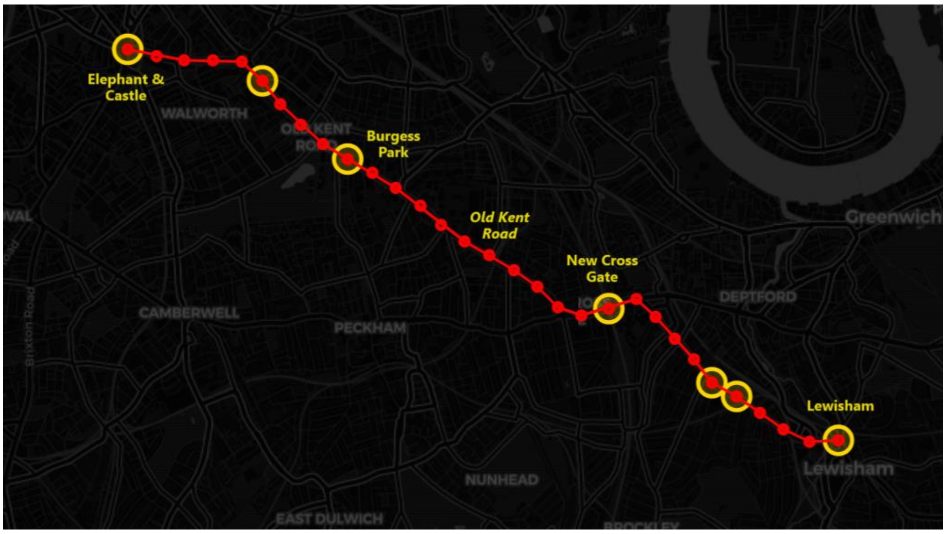

Figure 5 – Optimal solution with seven stations built (gold circles) vs. TfL proposal (station names and approximate locations)

- Besides the three must-build stations, TfL's proposal and our optimal solution both suggest a station be built around Burgess Park.
- The segment between Elephant & Castle and Burgess Park has no stations planned in between, while the solution suggests adding an infill station here to benefit residents in the populated area around Bricklayer's Arms / Ring Road Square.
- The solution also suggests adding two stations 250m apart between New Cross Gate and Lewisham (near St John's railway station). This is possibly due to this area's density or poor pedestrian connection. Since such an alignment is unlikely to pass, stakeholders could consider adding only one station here and improving pedestrian accessibility around the station's catchment area.
- The Old Kent Road Station in the TfL proposal does not correspond to any chosen candidate in our solution, possibly because this area is sparsely populated with good connection on foot to other stations. However, it is worth noting that this area will see more development in the future as a designated Opportunity Area (*London City Hall*, 2023) and will anticipate convenient access to a Tube station.

To summarise the findings: With the help of the model, we can put forth a recommendation to add two stations to the TfL proposal, one between Elephant & Castle and Burgess Park, and another between New Cross Gate and Lewisham, for a total of *six* new Tube stations for the BLE (excluding Elephant & Castle)

Lastly, it is worth acknowledging some limitations in formulating the objectives and constraints. For *Problem 1,* the construction cost of each station candidate should be added to serve as the weight parameter, combined with a project budget constraint. (Church and Murray, 2018). For *Problem 2,* time delay for passengers on the trains with each additional station can also be added as an objective to be minimised (Hamacher *et al.*, 2001). Moreover, it can be beneficial to include operational constraints such as station capacity to ensure no one station gets assigned an outsized demand share, as well as a minimum distance between any two stations to ensure engineering feasibility. Moreover, our model could not accommodate a future population growth parameter for each neighbourhood. To remedy this, the future projected population could be used to populate the input parameter a. sub i. (demand weight) instead of the current population.

Conclusion

This has been an attempt to apply Linear Programming to determine where to build new stations for the Bakerloo Line Extension that can cover all neighbourhoods in a certain area and minimise time to reach the stations, to improve transit accessibility in South London. Our findings suggest that a classic Location Set Covering Problem formulation (#1) yielded unsatisfactory results in terms of max walking time. On the other hand, an adapted P-Median Problem (#2) to minimise walking time as the primary objective returned a more workable solution. Contrasting the solution with the official proposal for the BEL led us to recommend *six new Tube stations for the BLE* instead of 4 in the current proposal.

The formulation explored in this paper can be further generalised for use by future research on station location planning. More specifically, there are merits in expanding the P-Median Problem (#2) into a Multi-objective Mathematical Programming (MMP) problem by combining with other secondary objectives to derive a more insightful set of Pareto-efficient solutions, and to make decision-making on complex problems more effective.

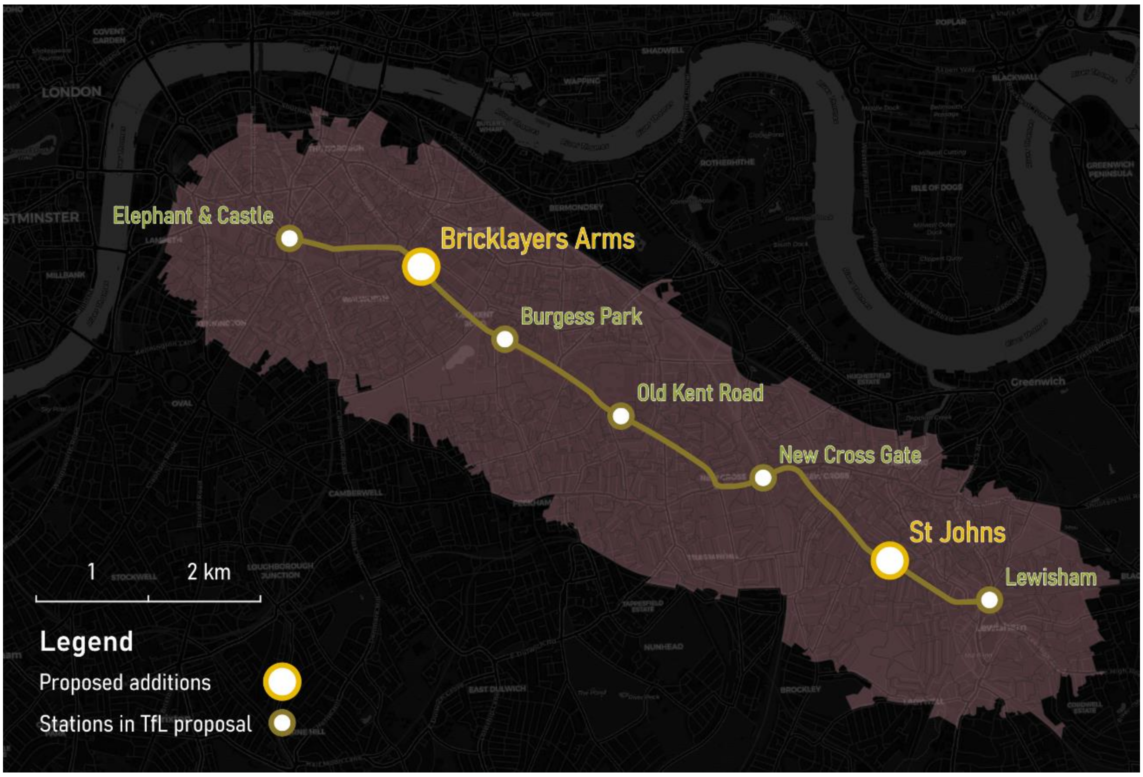

Figure 6 - Proposed additional BLE station locations (provisional names)

References

Anwar, A. M. (2012). ‘Paradox between Public Transport and Private Car as a Modal Choice in Policy Formulation’. *Journal of Bangladesh Institute of Planners*, 2. doi: 10.3329/jbip.v2i0.9568.

Benli, A. and Akgün, İ. (2023). ‘A Multi-Objective Mathematical Programming Model for Transit Network Design and Frequency Setting Problem’. *Mathematics*, 11, p. 4488. doi: 10.3390/math11214488.

Church, R. L. and Murray, A. (2018). ‘Location Modeling and Covering Metrics’. in Church, R. L. and Murray, A. (eds) *Location Covering Models: History, Applications and Advancements*. Cham: Springer International Publishing (Advances in Spatial Science), pp. 1–22. doi: 10.1007/978-3-319-99846-6_1.

Gavish, B. and Shlifer, E. (1979). ‘An approach for solving a class of transportation scheduling problems’. *European Journal of Operational Research*, 3 (2), pp. 122–134. doi: 10.1016/0377-2217(79)90098-5.

Hakimi, S. L. (1965). ‘Optimum Distribution of Switching Centers in a Communication Network and Some Related Graph Theoretic Problems’. *Operations Research*. INFORMS, 13 (3), pp. 462–475. doi: 10.1287/opre.13.3.462.

Hamacher, H. W., Liebers, A., Schöbel, A., Wagner, D. and Wagner, F. (2001). ‘Locating New Stops in a Railway Network1 1This work was partially supported by the Human Potential Programme of the European Union under contract no. HPRN-CT-1999-00104 (AMORE).’ *Electronic Notes in Theoretical Computer Science*. (ATMOS 2001, Algorithmic MeThods and Models for Optimization of RailwayS (Satellite Workshop of ICALP 2001)), 50 (1), pp. 13–23. doi: 10.1016/S1571-0661(04)00162-8.

Jafari, H. and Yaghini, M. (2019). ‘Optimal Location of Subway Stations: A Case Study on Tehran Subway’. *International Journal of Railway Research*, 6 (2), pp. 123–131. doi: 10.22068/IJRARE.6.2.123.

Karatas, M., Razi, N. and Tozan, H. (2016). ‘A Comparison of p-median and Maximal Coverage Location Models with Q–coverage Requirement’. *Procedia Engineering*. (International Conference on Manufacturing Engineering and Materials, ICMEM 2016, 6-10 June 2016, Nový Smokovec, Slovakia), 149, pp. 169–176. doi: 10.1016/j.proeng.2016.06.652.

London City Hall. (2023). Available at: https://www.london.gov.uk/programmes-strategies/planning/implementing-london-plan/londons-opportunity-areas/oa-locations (Accessed: 17 January 2024).

Sarker, R. I., Mailer, M. and Sikder, S. K. (2019). ‘Walking to a public transport station: Empirical evidence on willingness and acceptance in Munich, Germany’. *Smart and Sustainable Built Environment*. Emerald Publishing Limited, 9 (1), pp. 38–53. doi: 10.1108/SASBE-07-2017-0031.

TfL. (2019). ‘Strategic case for metroisation in south and southeast London’.