Module	*CASA007 – Quantitative Methods*
Assignment	*Written Investigation*
Word Count	*1845*

Choosing optimal station locations for the Bakerloo Line Extension with Linear Programming

According to Transport for London (TfL's 'Strategic Case for Metroisation in South and Southeast London' (TfL, 2019), residents of the area are missing out on opportunities because of inadequate public transport 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. The report claimed that there are four times as many jobs within 45 minutes of Harrow (North London) compared to Sutton (South London). Due to poor public transport connectivity, residents switch to more environmentally unsustainable modes such as personal automobiles.

With an expected population of more than 10 million by 2030, this unequal access to good public transport 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 other rail services together with two brand new stations along Old Kent Road (**Error! Reference source not found.**).

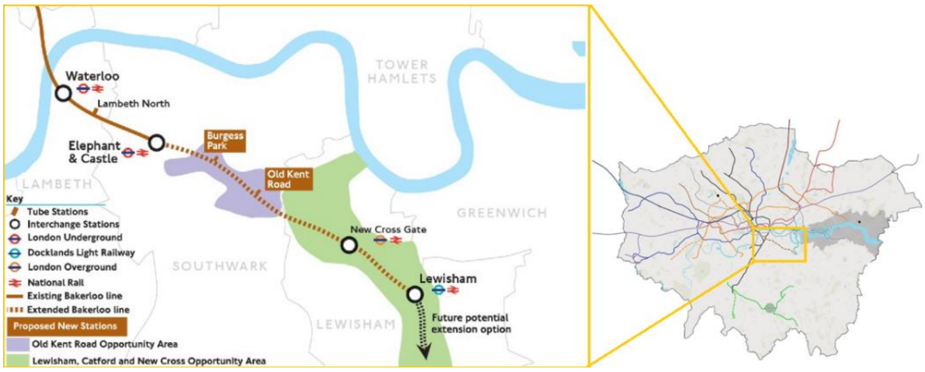

Figure 1 - Proposed Bakerloo Line Extension (Source: TfL)

The Bakerloo Line Extension (BLE) is one of the major projects as part of the larger strategy to revitalise this region within Greater London and improve its residents' lives. As a result, it is essential to ensure that the line, once completed, will make taking the Tube the preferred alternative to the current transport mode.

This research will use different Linear Programming models to determine the number and location of new stations along the new rail corridor needed to ease pedestrian access, i.e., to minimise the walking time to a station, which is one of many factors that can motivate individuals to change their commute behaviour (Anwar, 2012). 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 staple in many aspects of transport planning, from service and employee scheduling (Gavish and Shlifer, 1979) to location planning (Jafari and Yaghini, 2019). When it comes to location planning, the main objective is to ensure that where the stations are located provides a net positive effect on the travel experience compared to the status quo within the financial, technical, and operational constraints that all infrastructure projects need to contend with. Within the scope of this study, the travel experience of a potential transit user to reach the station will be the goal, while their experience in the station, on the train, or with the overall service will not be considered.

The theoretical framework for locating stations has been well-outlined by Hamacher et al. (2001), and it encompasses two sub-models:

1. Accessibility Model: Chosen stations must cover all or a predetermined share of the demand. In the case of a single line, the problem will resemble a classic Location Set Covering Problem.
2. Travel Time Model: The station locations must seek to minimise time spent to reach them and minimise the incremental delays to passengers onboard with each additional stop.

To solve these with Linear Programming, we formulated these priorities as linear programming problems. First, to minimise stations built to cover all demand, the classic Location Set Covering Problem formulation was used (Church and Murray, 2018). Then, to minimise the walking time (representing cost), an adapted P-Median Problem formulation (Hakimi, 1965) was used. The P-median Problem was chosen over the Maximum Coverage Location Problem for its focus on explicitly minimising population-weighted cost (walking time) rather than maximising population-weighted coverage (Karatas, Razi and Tozan, 2016). Note that, for this paper, time delay will not be considered, and all focus will be on walking time.

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)

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

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

such that

$$\sum_{j \in J} Y_j \leq k \quad (5)$$

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

$$X_{ij} \leq Y_i \qquad \forall i \in I \ \forall j \in J \quad (7)$$

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

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

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 of the problems and respective constraints 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.

As with most real-life problems, we would like to find a solution that can optimise for these two objectives concurrently rather than in isolation, which would necessitate the formulation of a Multi-objective Mathematical Programming (MMP) problem, commonly seen in research on transit planning using linear programming. In these instances, 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. (Chen and Zhou, 2022).

Due to the limited scope of this paper, we instead explored the universe of optimal solutions by varying one input parameter per problem within manually set ranges. However, note that this brute-force approach would not be suitable for complex problems with more than two objectives.

Data preparation

The candidate location and the neighbourhood sets, J and I, respectively, were acquired from the following workflow[1]:

1. Create a linestring vector for the corridor connecting Elephant & Castle with Lewisham, mainly following Old Kent Road (total length 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 (i.e., a_i)
5. Calculate the walking distance all i and j pairs ($i.e., d_{ij}$), using the publicly available OpenStreetMaps pedestrian routing server.

For this specific case, we also designated candidates that must be included in the solution, because they 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) to the two problems:

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

The resulting sets are visualised in Figure 2.

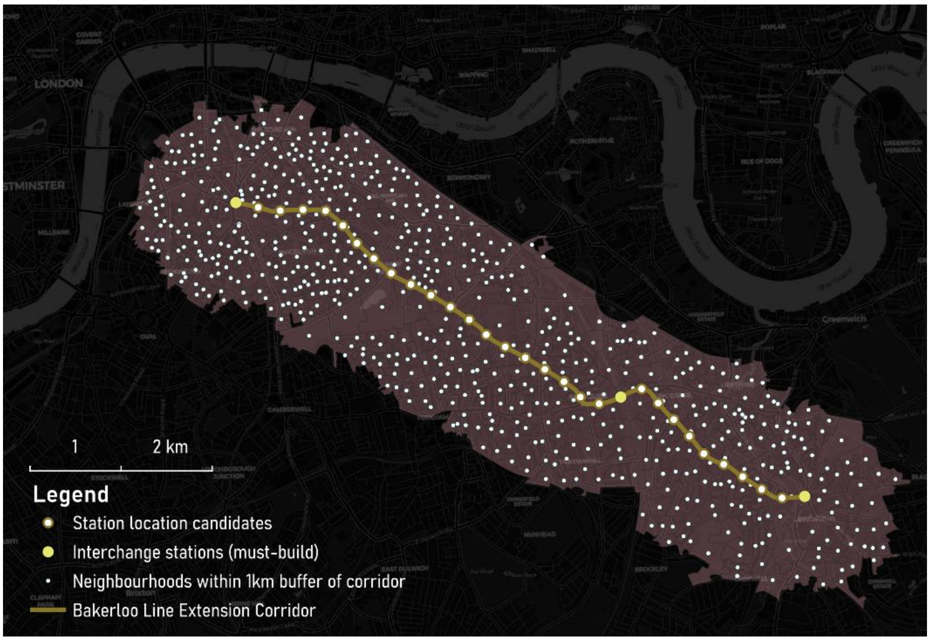

Figure 2 - Station candidate and neighbourhood sets

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

- The formation of the neighbourhood location set I does not consider the propensity to use public transport, future transit demand, or local politics that might stipulate

[1] Data source: Office of National Statistics

certain neighbourhoods to be included in (or excluded from) the set, such as the designated Opportunity Areas in Southeast London. (*London City Hall*, 2023)

- The formation of the station candidate location set J does not consider which site is technologically feasible to sustain the construction of an underground train station.
- Different routing services, such as that offered by Google or Mapbox, may yield a different cost matrix (of walking distance) and, thus, different solutions.
- We assume that the BLE is a standalone line segment with no interactions with other current and future transit lines, whose stations might also cover the demand points in set I. Without this assumption, the optimisation problem would be computationally difficult (Hamacher *et al.*, 2001).

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)[2]

Results

Error! Reference source not found. shows all optimal solutions for ***Problem 1*** *(Minimise station)* at different T values (i.e., max walking time). From here we can see that, if T is set below 1800 seconds (30 minutes), the problem is unsolvable. Since urban residents are only willing to walk up to around 10 minutes to reach a rapid transit station (Sarker, Mailer and Sikder, 2019), we can conclude that there are no feasible solutions to this problem at T = 600 seconds (10 minutes)

We now look at the solutions to ***Problem 2*** *(Minimise walking time).* **Error! Reference source not found.** exhibits all optimal solutions at different k values (i.e., maximum number of stations). Here, the problem is solvable at all k values between 3 and 31, with a 'knee' at $k = 7$, at which the optimal solution is 586 seconds (~10 minutes) of average walking time[3].

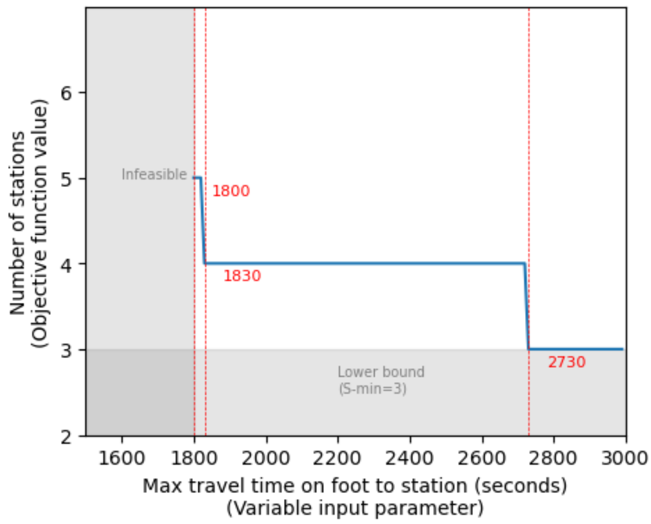

Figure 3 - Minimum stations needed

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

[3] Avg. walking time is equal to total walking time (objective function) divided by total population

at varying max walking distance (Problem 1)

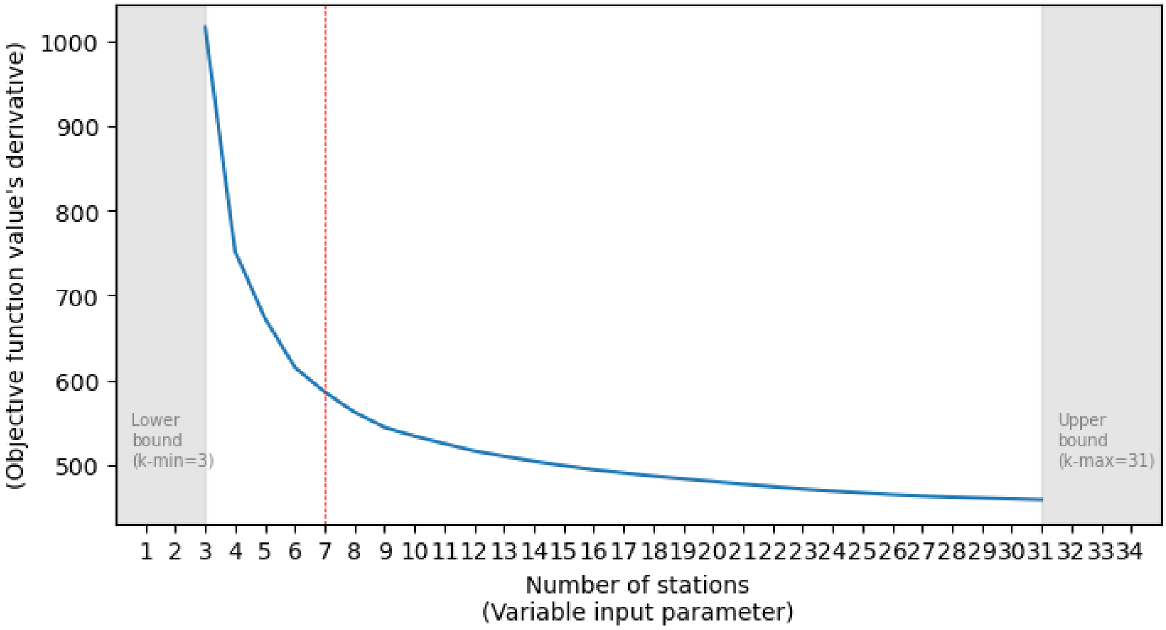

Figure 3 - Minimum average walking time
at varying numbers of stations given (Problem 2)

Discussion

When the chosen locations (k=7) are juxtaposed with the official proposal for the BEL, we can observe several mismatches:

- The segments between Elephant & Castle and Burgess Park, and between New Cross Gate and Lewisham have no intermediary stations planned. As the solution suggests, adding infill stations here will benefit residents in these densely populated areas.
- The optimal solution also suggests that two stations 250m apart are needed between New Cross Gate and Lewisham, possibly due to this area's density or poor pedestrian connection. Since such an alignment is unlikely, stakeholders could consider adding only one station but with better = accessibility to minimise the walk.
- The planned Old Kent Road Station does not correspond to any chosen candidate, possibly because this area is sparsely populated but well connected on foot. This also reveals a shortcoming of the problem: We did not account for the area's designation as an Opportunity Area (*London City Hall*, 2023), which is bound to see growth stemming from increased investments.

There are various unexplored possibilities in formulating the objective functions and their constraints. For *Problem 1*, the construction cost of each station candidate can be added to serve as the weight parameter for the decision variable. (Church and Murray, 2018). For *Problem 2,* time delay to passengers on the trains as a function of additional stations could be added to the objective minimisation function. (Hamacher *et al.*, 2001).

Beyond the objective functions, several realistic operational constraints should be considered, such as station capacity, project budget, minimum distances between two stations based on contemporary rail technology, etc. However, one must be aware that enforcing more constraints may make the problems unsolvable or computationally difficult.

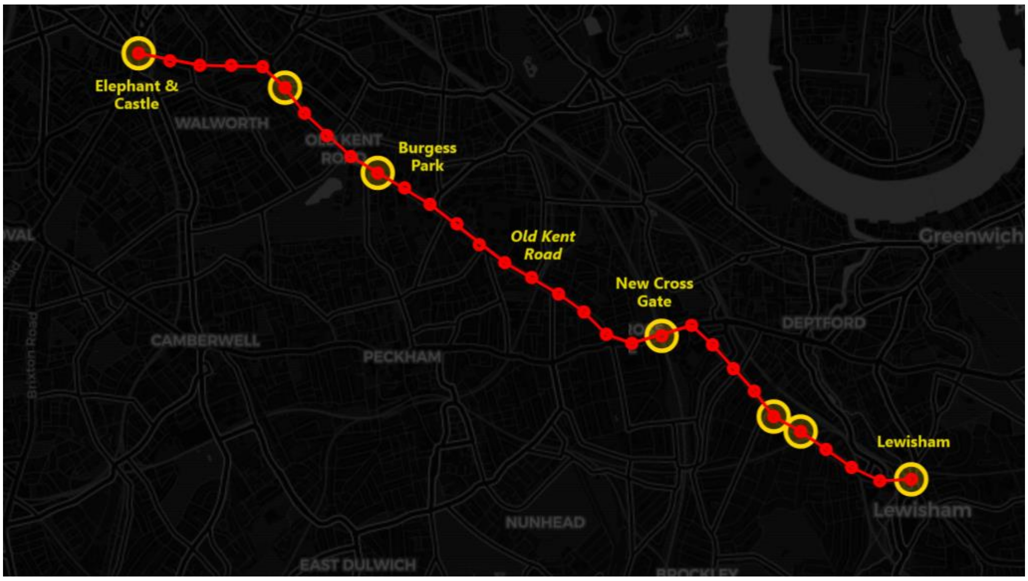

Figure 4 – Optimal solution with seven stations built (gold circles) vs. TfL proposal (station names)

Conclusion

This has been an attempt to apply Linear Programming in two ways to determine where to build new stations for the Bakerloo Line Extension that can cover all neighbourhoods in a certain area and minimise walking time to the station.

Our findings suggest that a simple formulation of the Location Set Covering Problem (#1) yielded unsatisfactory results if a maximum walking time constraint of 10 minutes was added. On the other hand, an adapted P-Median Problem (#2) seeking to minimise walking time as the primary objective returned a feasible solution. Contrasting the solution with the official proposal for the BEL reveals potential new station candidates and the limitations of our formulation in factoring in temporal changes.

The formulation explored in this paper can be generalised for use by future research on station location planning. Expanding the P-Median Problem (#2) into a Multi-objective Mathematical Problem (MMP) with other secondary objectives is recommended to efficiently derive a more insightful set of Pareto-efficient solutions that can inform decision-making more effectively.

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.

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’.