|  |  |
| --- | --- |
| ***Module*** | *CASA007 – Quantitative Methods* |
| ***Assignment*** | *Written Investigation* |
| ***Word Count*** | *1750* |

# [Title]

## Introduction

According to the Bakerloo Line Extension Consultation Report (TfL, 2019), Southeast London is poorly serviced by the Underground network compared to the rest of Greater London, although its population is set to increase by more than 10 million by 2030. In fact,…stat…( )

This necessitated the proposal by the Mayor of London to extend the Bakerloo Line from its current terminus at Elephant & Castle to Lewisham (Figure 1).

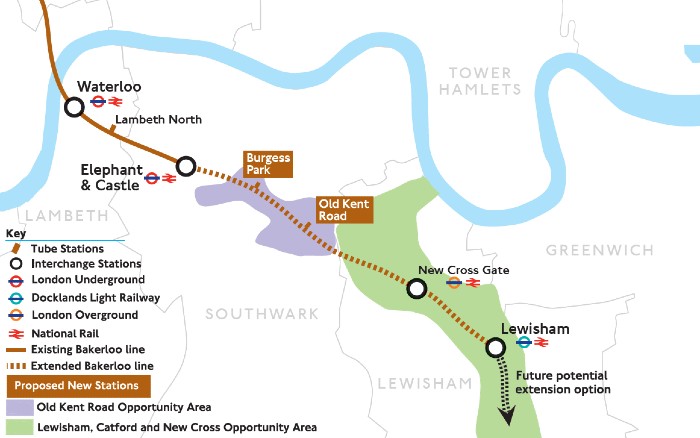


Figure 1 - Proposed Bakerloo line extension map (Source: TfL)

Howeverm

The goal of our case study will be to compare our solution for the locations of the new stations with the five proposed stations in the official proposal, namely *Elephant & Castle, Burgess Park, Old Kent Road, New Cross Gate, and Lewisham stations.*

## Literature Review and Methodology

[What are the things to optimise when considering building a train line]

[Why Linear Programming]

[Literature]

Finally, perhaps, as with all real-world problems, our ultimate goal is to optimise multiple objectives not in isolation but at the same time with a certain degree of trade-off, which in this context is between the number of stations to be built and walking time to the stations. Therefore, a more appropriate way to reformulate this is as a multi-objective mathematical programming (MMP) problem. One common approach is to apply an epsilon-constraint, whereby objective functions are to be solved in sequence, and the range of feasible solutions goes on to serve as the constraints to solve for the next objective function (Jafari and Yaghini, 2019). The final set of Pareto-efficient solutions can then form the Pareto front so that decision-makers can focus their attention on the set of efficient options, each with its trade-off. (Chen and Zhou, 2022)

We already performed a variation of this method in the BLE case study, albeit with brute force, by letting input parameters vary within a manually set range to form Figure … and …, based on which we can make the final decision. However, This manual approach is unsuitable for complex problems with more than two objectives.

## Problem Formulation

The desired objectives to be optimised are formulated as linear programming problems. To minimise stations built to cover all demand, the classic Location Set Covering Problem formulation is used (Church and Murray, 2018). On the other hand, to minimise the walking time (representing transport cost), an adapted P-Median Problem formulation is used (Hakimi, 1965)

The P-median Problem has been chosen over the Maximum Coverage Location Problem for specifically minimising population-weighted cost (walking time) rather than maximising population-weighted coverage (Karatas, Razi and Tozan, 2016)

The formulation of the problems and their respective constraints are interpreted as follows:

Objective functions for the two problems

Binary constraints for the decision variables

Every neighbourhood *i* is within a max walking time T of min. 1 station.

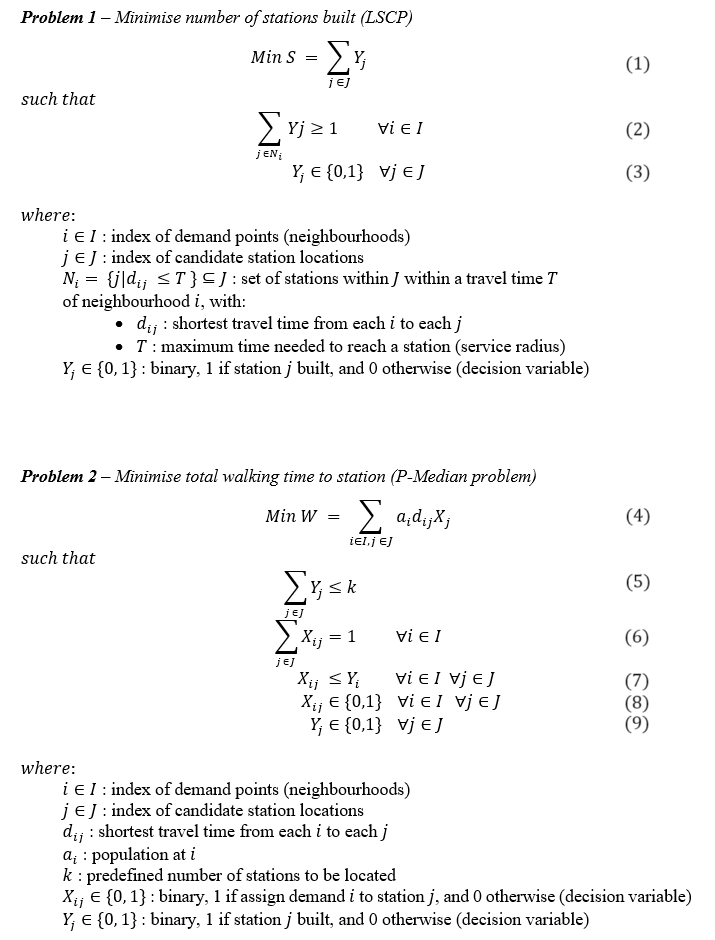
Maximum stations can be assigned.

Each neighbourhood is assigned only one station.

Each neighbourhood is assigned to station only if it’s built.

A paper with math equations

Description automatically generated



## Preparing the data for the optimisation problems

The candidate location and the neighbourhood sets, and respectively, were acquired from the following workflow[[1]](#footnote-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.
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.
4. Extract the population of all points in (i.e., )
5. Calculate the walking distance all and (), using the publicly available OpenStreetMaps pedestrian routing server[[2]](#footnote-2)

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 with the closest proximity to the three locations above are . Therefore, we added the ad-hoc constraint to the two problems:

(10)

The resulting sets are visualised in Figure 2.

A map of a city

Description automatically generated

Figure 2 - Station candidate and neighbourhood sets

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

* The formation of the neighbourhood location set does not consider the propensity to use public transport, future transit demand, or local politics that might stipulate 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 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 . Without this assumption, the optimisation problem would be computationally prohibitive (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)[[3]](#footnote-3)

## Optimisation results

Figure 3 shows all optimal solutions for ***Problem 1*** *(Minimise station)* at different values (i.e., max walking time). From here we can see that, if 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).* Figure 4 exhibits all optimal solutions at different values (i.e., maximum number of stations). Here, the problem is solvable at all values between 3 and 31, with a ‘knee’ at , at which the optimal solution is 586 seconds (~10 minutes) of average walking time[[4]](#footnote-4).

|  |  |
| --- | --- |
|  |  |
| Figure 3a - Minimum number of stations needed  at varying max walking time (Problem 1) | Figure 3b - 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 still sparsely populated but still 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) that is bound to see growth stemming from increased investments.

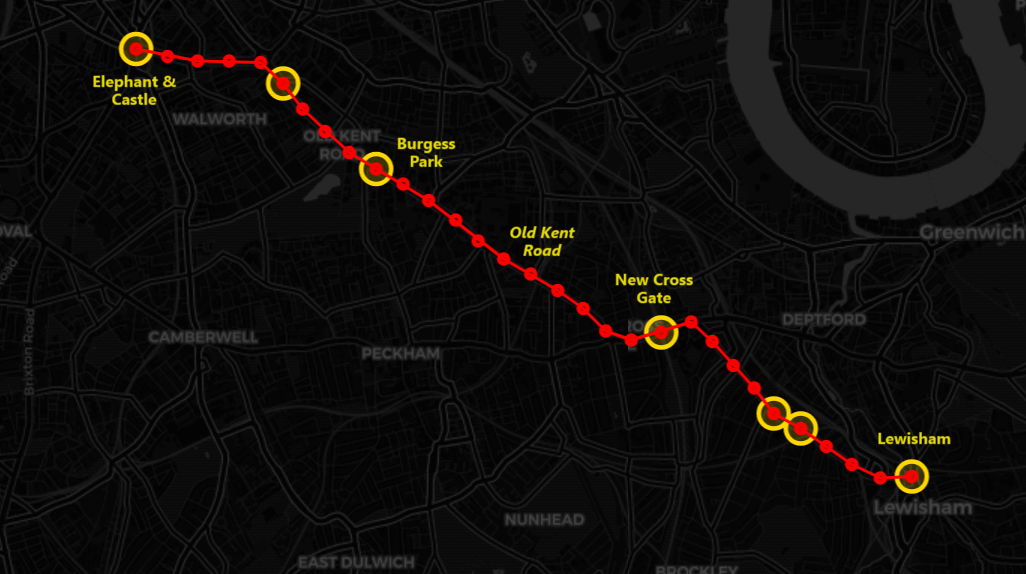


Figure 5 – Optimal solution with seven stations built (gold circles) vs. TfL proposal (station names)

There are various limitations 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 as a function of additional station may be added to the objective function to be minimised (Hamacher *et al.*, 2001).

Beyond the objective functions, there are several realistic constraints that should be accounted for, such as station capacity, project budget, minimum distances between two stations based on contemporary rail technology, etc. Lastly, due to the exploratory nature of this paper, there is no maximum walking time constraint, and is replaced with an input parameter variable within a manually set range to extract all possible optimal solutions. However, one must be aware that enforcing more constraints may make the problems unsolvable or computationally difficult.

## 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 station.

Our findings suggest that a simple formulation of the Location Set Covering Problem (#1) yielded unsatisfactory results if adding a maximum walking time constraint of 10 minutes. 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. More specifically, 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

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.

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.

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.

1. Data source: Office of National Statistics [↑](#footnote-ref-1)
2. Router server URL: http://routing.openstreetmap.de/routed-foot/ [↑](#footnote-ref-2)
3. 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… [↑](#footnote-ref-3)
4. Avg. walking time is equal to total walking time (objective function) divided by total population (constant) [↑](#footnote-ref-4)