# Public Transport Stop Selection to Reduce the Inequality in the Distribution of Accessibility

Duo Wang*[1], Andrea Araldo[2], and Mounim El Yacoubi[3]

[1]PhD student, SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France

[2]Assoc. Prof., SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France

[3]Prof., SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, France

February 2, 2024

## SHORT SUMMARY

What-to methods for the design of Public Transport (PT) traditionally maximize overall efficiency. They do not generally embed the inequality of the distribution of accessibility into the optimization objective. However, such inequality is crucial, as it contributes to the car-dependency of areas underserved by PT. In fact, while inequality is generally considered in what-if methods, embedding it directly into a PT design optimization algorithm is challenging. We show here that this can be achieved by setting selected bottom quantiles of the accessibility distribution as objective function. For simplicity, we focus on the PT stop selection problem. With a heuristic algorithm, we compare the resulting equality-maximizing PT with conventional efficiency-maximizing PT. Our numerical results show that, already with the sole stop selection, inequality of accessibility is significantly reduced. This shows potential for even greater inequality reduction by applying our approach to other PT planning decisions.
**Keywords**: Accessibility, Equality, Public Transport Design, Heuristic Algorithm

## 1 INTRODUCTION

The main indicator we consider in this paper is *accessibility*, i.e., the ease of reaching the *opportunities*, such as schools, jobs, other people, from a location, via PT (Miller (2020)). A significant inequality characterizes the geographical distribution of accessibility in real cities (Badeanlou et al. (2022)). Recent work attempts to include in PT network design indicators of equality of PT deployment across a territory.

Camporeale et al. (2019) proposes a classic cost-minimizing PT design problem, with the addition of a constraint preventing inequality of PT supply to exceed a certain threshold. PT supply is a *local measure*, accounting for number of buses passing by or stops in a location: improving supply in a location can be directly achieved by modifying PT production therein. We are instead interested in accessibility, which is not local, and is thus more complex: the accessibility of a location may depend on the PT topology or opportunities in locations that may be further away but strongly connected. Therefore, improving accessibility in a location cannot be achieved by simply changing PT production in the very same location, but always require to consider the territory and the PT topology in their entirety. Kim et al. (2019) design PT lines to reduce the difference in travel times between cars and PT. They do not consider accessibility either.

More similar to us, Dai et al. (2022) optimize headway and stop spacing of a single line for equality of accessibility distribution. We however tackle a more realistic network with multiple lines, which prevent the use of their method. The objective function of Yoo & Lee (2023) is the sum of accessibility of all areas, weighted by "Transit Need" (TN) coefficients, so as to prioritize accessibility improvement in high-TN areas. This submits PT design to political will, via appropriately adjusting TN computation. We propose instead a method as objective as possible, based on geographical horizontal equality, which is technically more difficult: we cannot rely on a simple weighted sum, but always deal with the entire distribution of accessibilities.

Finding an effective way of expressing our inequality-reduction objective into an optimization algorithm is challenging. Our preliminary attempts of using inequality indicators (Costa et al. (2019)) directly into the objective function resulted in poor performance. The classic max-min optimization, which in our case would mean to maximize the lowest accessibility, also was ineffective, as it tends to improve accessibility of

too few locations, often remote or isolated, which then results in a weak improvement of global inequality indices and an excessive loss in overall efficiency.
We propose here a novel approach for equality-maximization PT design, consisting in maximizing some bottom quantiles of the accessibility distribution. We apply our approach to the PT stop selection problem, which we solve with a heuristic algorithm. Numerical results show significant improvement in equality, with a negligible loss of overall efficiency. Via statistically analyzing the results of the optimization algorithms, we infer some simple guidelines for equality-based PT design. Our code is available as open source (Wang (2023)).

## 2 METHODOLOGY

### *Model of the Public Transport Network*

A PT network is modeled as a graph $\mathcal{G}$ as (Figure 1). A PT line (train, metro or bus) is a sequence of stops, linked by corresponding edges. On each PT line $l$ a number $N_l$ of vehicles operate, traveling at speed $v_{\text{veh}}$. At each stop, a vehicle stays for a dwell time $t_{\text{dwell}}$ to allow boarding and alighting. We include in $t_{\text{dwell}}$ the time lost in acceleration and deceleration. Assume a line with stop sequence $s_1,\dots,s_{K_l}$ and let us denote with $d(s_j,s_{j+1})$ the Euclidean distance between $s_j$ and $s_{j+1}$. The time a vehicle takes to visit the entire line is $T_l = \sum_{j=1}^{K_l-1}\left(\frac{d(s_j,s_{j+1})}{v_{\text{veh}}}+t_{\text{dwell}}\right)$. As in Desaulniers & Hickman (2007) and Pinto et al. (2020), the headway and the average waiting time of line $l$ are $H_l = 2\cdot T_l/N_l$ and $w_l = H_l/2$, respectively.
Graph $\mathcal{G}$ represents PT network in a certain timeslot, within which we assume that line routes and headway values do not change. In further work, the operation of PT during a day could be represented as a sequence of graphs, one per each timeslot.
For simplicity, we consider a square study area, partitioned with a regular tessellation (no matter the shape of the tiles). The center of each tile is called *centroid*. We denote with $\mathcal{V}$ the set of centroids. Each tile contains a certain amount of opportunities, that we associate to the respective centroid. Users start at any centroid and are willing to reach opportunities around other centroids.In the numerical results, we will consider a particular form of accessibility, called *sociality score* (Biazzo et al. (2019)), where people are considered as opportunities.
A trip can be performed entirely by walking, or combining walking and PT. Within the PT network, a user can change from line $l_1$ to line $l_2$, if the egress stop $s$ in $l_1$ and the ingress stop $s'$ in $l_2$ are within distance $d_{\text{exchange}}$. Multiple changes are allowed. When entering line $l$ (directly after walking or during a change of line), a user suffers waiting time $w_l$. User always travels along shortest time path. More advanced PT assignments (Spiess & Florian (1989)) could be considered without changing the spirit of this work.

### *Accessibility*

The *gravity-based* definition of accessibility (Miller (2020)) of centroid $v \in \mathcal{V}$ is

$$acc(v) = \sum_{u\in\mathcal{V}} X(u)\cdot f(T_{v,u}) \tag{1}$$

where $T_{v,u}$ is the shortest time to go from $v$ to $u$,$f(\cdot)$ is the impedance function and $X(u)$ is the amount of opportunities of the tile having centroid $u$. Choosing impedance function $f(T) = T^{-1}$ has the advantage of expressing accessibility in intuitive units of measurement, i.e. number of opportunities that can be reached per hour. When optimizing for equality, we will focus in particular on set $\mathcal{V}^{m\%}$ of the *worst m% centroids*, i.e., those with the lowest accessibility. We define the accessibility of graph $\mathcal{G}$ as:

$$Acc(\mathcal{G};m) = \frac{1}{|\mathcal{V}^{m\%}|}\sum_{v\in\mathcal{V}^{m\%}} acc(v) \tag{2}$$

which, with $m = 100$, corresponds to average accessibility $\overline{acc}(\mathcal{G}) = Acc(\mathcal{G},100)$.

### *Inequality index*

Denoting the accessibility values of centroids $v_1,\dots,v_K$ by $y_1,\dots,y_K$ respectively, the Atkison inequality index is (from Costa et al. (2019), setting $\varepsilon = 2$):

$$Atk(\mathcal{G}) = 1 - \frac{1}{\overline{acc}(\mathcal{G})}\cdot\left(\frac{1}{K}\sum_{i=1}^{k} y_i^{-1}\right)^{-1} \tag{3}$$

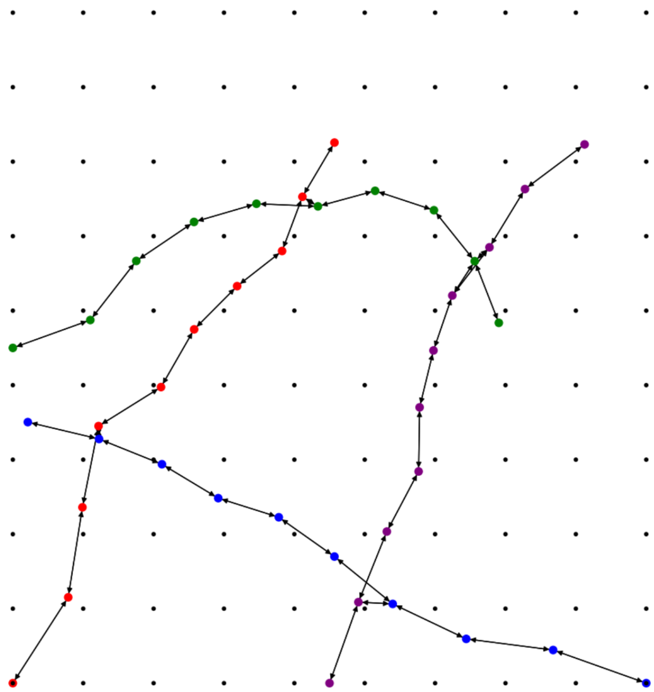

Figure 1: An example of PT network: 4 PT lines containing 40 stops and 100 centroids

The Atkinson index goes from 0 (perfect equality) to 1 (maximum inequality). When we tried to embed the inequality indices directly into the optimization, we found poor results, as such indices are poorly sensitive to PT modifications. We therefore only use (2) during optimization and compute the inequality index a posteriori, to check the quality of the solution. If there is data about population age, income distribution, etc., we can also calculate the vertical equity resulting from our optimization. Results did not changes when we used other indicators,e.g., Theil, Gini, Pietra.

***Heuristic algorithm***

As shown in Algorithm 1, for a given PT graph $\mathcal{G}$, we run $n$ independent instances of the heuristic algorithm. Each instance randomly deactivate stops. Each time a stop is deactivated, the corresponding line is consolidated: the edges incident to that stop are removed and a new edge is added between the stop before and the stop after the removed one, thus reducing travel and dwell times. We consider in the numerical results a bus network, where edge consolidation also reduces the traveled distance. The method can be generalized to rail networks, where distance would not be reduced. The algorithm might also remove, in some iterations, stops that are connected to multiple lines. If such connections were used by many shortest paths, this may generally decrease accessibility. However, the algorithm selects the best solution (Lines 9,11), which is likely to filer out solutions where "important" connections have been cut. Figure 6 will confirm this. Note that deactivating a stop does not necessarily mean to physically removing it, but to simply skip it during the considered timeslot.

**Efficiency-based optimization** (in short **Ef-Opt**) is obtained by running Algorithm 1 with $m = 100$, so that the search algorithm will tend to maximize the average accessibility. **Equality-based optimization** (in short **Eq-Opt**) is instead obtained with $m = 5$, to preferentially improve the accessibility of the worst 5% centroids. Given initial graph $\mathcal{G}$, we denote with $\mathcal{G}^*_{\text{ef}}$ and $\mathcal{G}^*_{\text{eq}}$ the respective results. We found that for $m$ smaller than 5, we tend to improve accessibility of only few remote centroids, without achieving good overall inequality (poor Atkison index) and excessively degrading overall accessibility, similar to maxmin optimization. On the other hand, higher values of $m$ would excessively penalize equality in favor of overall accessibility.

**Algorithm 1:** Heuristic algorithm

**Input** Public transport graph $\mathcal{G}$ with stops $\mathcal{S}$.
    Parameter $m$ of the accessibility formula (2).
**For** search instance $i \leftarrow 1$ to $n$ :
    Initialize $\mathcal{G}_0 \leftarrow \mathcal{G}$ and $\mathcal{S}_0 \leftarrow \mathcal{S}$.
    **For** step $t \leftarrow 0$ to $\infty$ until **less than** $n_{end}$ **stops remain active**:
        Select a random stop $s_t \in \mathcal{S}_t$ and deactivate it.
        Set $\mathcal{S}_{t+1} \leftarrow \mathcal{S}_t \setminus \{s_t\}$ and let $\mathcal{G}_{t+1}$ the resulting PT graph.
        Compute the new accessibility: $Acc(\mathcal{G}_{t+1};m)$
    **EndFor**
    Record $\mathcal{G}^i = \arg\max_{\tau=0}^{t+1} Acc(\mathcal{G}_\tau, m)$.
**EndFor**
**Return** PT graph $\mathcal{G}^* = \arg\max_{i=1}^{n} Acc(\mathcal{G}^i, m)$.

## 3 Numerical evaluation

### *Considered scenarios*

Due to city peculiarities, morphological constraints, pre-existent infrastructure and even political influences real PT networks are extremely heterogeneous and often characterized by irrationality and inefficiency. We could have considered few cities, model their PT graph and apply our optimization. However, we would have ended up in a few case studies, without assessing the generality of our approach faced with PT heterogeneity. Since the focus of this paper is on the method, rather than specific case studies, we choose instead to synthetically generate 50 PT graphs and verify the performance is satisfying, no matter the graph at hand. We study the graphs resulting from *Ef-Opt* (baseline) and from *Eq-Opt*, which we advocate.
A synthetic graph includes multiple lines, generated via Algorithm 2, which captures the properties of real PT lines. Stop spacing increases with the distance from the center, as in Furth et al. (2000) (Lines (4)-(5) of Algorithm 2). To achieve the clear directionality, we allow limited direction change from a line segment to the next (Lines (6)-(8) of Algorithm 2). If a stop falls outside the study area, the algorithm restarts. Figure 1 gives an example of graph. We only present experiments with 8 lines, as with fewer lines trends were similar after adjusting number $n$ of search instances of Algorithm 1: with 4 lines increasing $n > 5$ did not bring improvement in the performance, while with 8 lines improvmement stops after $n = 20$ instances. We visualize a large set of synthetic PT graphs and verify they reasonably represent realistic PT networks.

**Algorithm 2:** Generation of a synthetic PT line.

1: **Input**: Range $\Theta$ of $\Delta\theta$; Maximum number $S$ of stops;
2: **Initialization**: Random initial bus stop location: $s_0$; Set $\mathcal{S} = \{s_0\}$ of bus stops; Random initial angle $\theta$ (with respect to the horizontal axis); Set the initial graph $\mathcal{G}$ as only composed by node $s_0$ with no edges.
3: **repeat**
4:     Calculate the distance $x$ from stop $s_{i-1}$ to the center
5:     Calculate the distance between $s_{i-1}$ and $s_i$ as $1/\beta(x)$ (see Table 1)
6:     Choose an angle $\Delta\theta$ uniformly at random from set $\Theta$
7:     Calculate $\theta = \theta + \Delta\theta$
8:     Calculate the location of the new bus stop, at distance $1/\beta(x)$ and angle $\theta + \Delta\theta$
9:     Add the new stop $s_i$ to $\mathcal{S}$ and an edge from $s_{i-1}$ to $s_i$
10: **until** The number of stops in $\mathcal{S}$ reaches $S$
11: **Return** PT graph $\mathcal{G}$ made of stops $\mathcal{S}$ and the added edges

We set amount $X(u)$ of opportunities in the tile around centroid $u$. Since the considered accessibility is the sociality score, opportunities correspond to residents:

$$X(u) = C \cdot \rho(u) \tag{4}$$

where $C = 0.25\,km^2$ is the area of a tile, $\rho(u)$ is the population density, uniformly distributed inside the tile around $u$ and constant within the considered timeslot. Population density follows the classic pattern (Muth

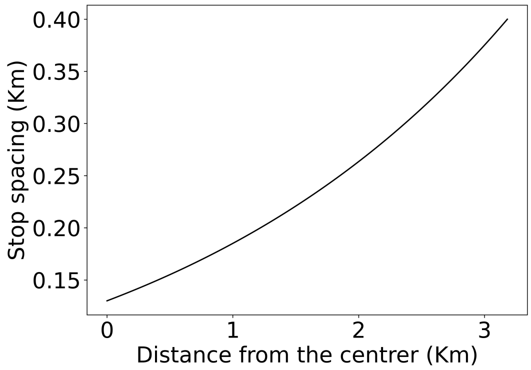


Figure 2: Stop spacing.

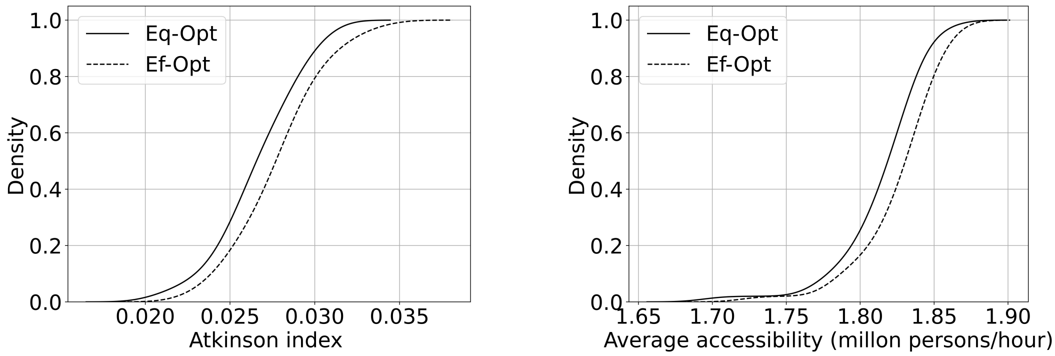


Figure 3: Cumulative Distribution Functions (CDFs) of Atkinson index and average accessibility after *Eq-Opt* and *Ef-Opt*

(1967)):

$$\rho(u) = \rho_0 \cdot e^{-\gamma \cdot d(u)} \tag{5}$$

where $d(u)$ is the distance between centroid $u$ and the city center, $\rho_0$ and $\gamma$ are hyperparameters. Stop spacing at distance $x$ from the center is $1/\beta(x)$, where $\beta(x)$ is the stop density and, similar to population density, follows an exponential distribution Furth et al. (2000) (Figure 2). Table 1 reports the adopted parameter values.

***Results***

For all generated graphs $\mathcal{G}_i, i = 1..50$, we compute average accessibilities $a_i = Acc(\mathcal{G}^*_{i,\text{eq}}, 100)$ of graph $\mathcal{G}^*_{i,\text{eq}}$ optimized via *Eq-Opt*. Subsequently, we perform the same calculation using *Ef-Opt* optimization, resulting in values $b_i, i = 1..50$. The $t$-test on these two lists yelds $p$-value = 0.0377, indicating a significant difference in average accessibility between *Eq-Opt* and *Ef-Opt*. A similar $t$-test on the lists of Atkinson inequality indices yelds $p$-value = 0.0005, indicating that the reduction of inequality obtained via *Eq-Opt* is very significant. For each graph, we compute the two indices above and we report their Cumulative Distribution Function (CDF) over the 50 graphs in Figure 3. Atkinson's inequality is consistently lower with *Eq-Opt* than with *Ef-Opt*, at the cost of a slight decrease in average accessibility. This is a trade-off between efficiency and equality. If we place excessive emphasis on efficiency, the resulting design may be very unequal.
Figure 4 focus on one exemplary graph $\mathcal{G}_i$, chosen randomly. Each dot corresponds to a centroid $u$. The $x$-coordinate indicates its accessibility $acc(u)$ (before optimization), while the $y$-coordinate indicates the improvement in accessibility obtained in $\mathcal{G}^*_{i,\text{eq}}$ and in $\mathcal{G}^*_{i,\text{ef}}$, i.e., the number of additional persons per hour that is possible to reach with respect to the initial graph $\mathcal{G}_i$. Quadratic regression lines and 95% confidence intervals are also shown. Observe that *Eq-Opt* concentrates the improvement within disadvantaged centroids, i.e., those with low accessibility in $\mathcal{G}_i$, at the cost of a slight decrease in accessibility for the advantaged centroids.

Table 1: Scenario parameters

| Parameter | Value |
|---|---|
| **PT networks** | 4500 m × 4500 m |
| Number of synthetic graphs | 50 |
| Number of PT lines per graph | 8 |
| Number of PT stops per line | 10 |
| Number of centroids | 100 |
| Distance | |
| between two nearest centroids | 500 m |
| maximum distance $d_{\text{exchange}}$ that user can walk to change line | 300 m |
| Average speed | |
| walking (Ali et al. (2018)) | $60\,\text{m}\,\text{min}^{-1}$ |
| PT | $300\,\text{m}\,\text{min}^{-1}$ |
| Dwell time | 1 min |
| Stop density at distance $x$ from the center ($\beta_0$ and $\gamma'$ are such that spacing is $1/130m$ in the center and $1/400m$ at the extremum of the study area, in line with Furth et al. (2000)) | $\beta(x) = \beta_0 \cdot e^{-\gamma' \cdot d(x)}$ |
| **Parameters of Algorithm 1** | |
| $m$ | 5 or 100 |
| $n$ | 20 |
| $M_{end}$ | 10 |
| number of PT overall stops for termination $n_{end}$ | 10 |
| **Parameters of Algorithm 2** | |
| $\Theta$ | $[-\pi/8, \pi/8]$ |
| $S$ | 80 |
| **Parameters of Formula** (5)(Bai et al. (2015)) | |
| $\rho_0$ | 36 000 people/km$^2$ |
| $\gamma$ | $0.01\,\text{km}^{-1}$ |

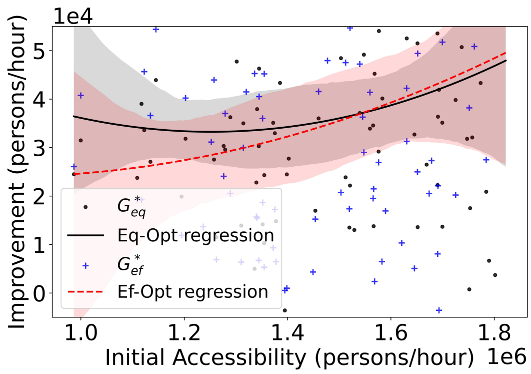


Figure 4: Change in accessibility on one exemplary graph.

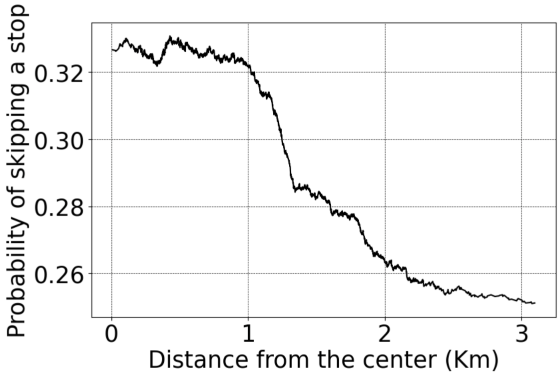


Figure 5: Relation between the probability of skipping a stop and distance from the center by *Eq-Opt*

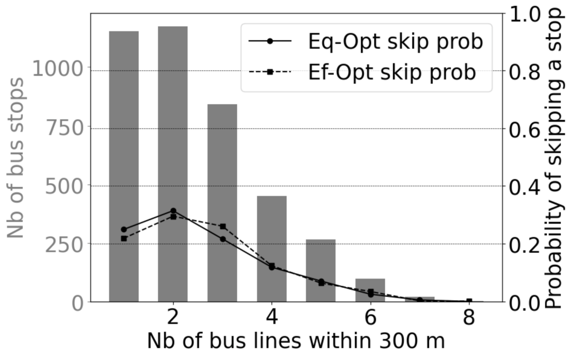


Figure 6: Relation between the probability of skipping a stop and the number of bus lines nearby *(Note $d_{exchange}$ = 300 m is the maximum distance users walk to change line, see Table 1)*

Next, we consider two features of each bus stop: distance $D_s$ from the center and number $nbl_s$ of bus lines recheable within 300 m. For each value of $D$, we take the 100 stops having $D_s$ closest to $D$ and calculate what proportion of them were skipped. Figure 5 shows that the closer the distance it from the center, the higher the probability that *Eq-Opt* skips bus stops. The significant higher skipping probability within 1Km from the city center means that in the initial graphs, where spacing is generated as in real deployments (Furth et al. (2000)), stop density is excessive and is not useful to accessibility. This confirms that need of de-densifying stop distribution, already raised by Furth et al. (2000). Figure 6 plots feature $nbl_s$ (gray bars) per each stop $s$, and the probability of skipping a stop is calculated in the same way as in Figure 5. Figure 6 shows *Eq-Opt* and *Ef-Opt* both tend not to skip stops that are close to other bus lines. Indeed, skipping such stops would strongly reduce the possibility of line change, thus cutting out paths that may greatly contribute to accessibility. Figure 7 shows that it is generally preferable, for both efficiency and equality, to maintain higher stop density close to the center.

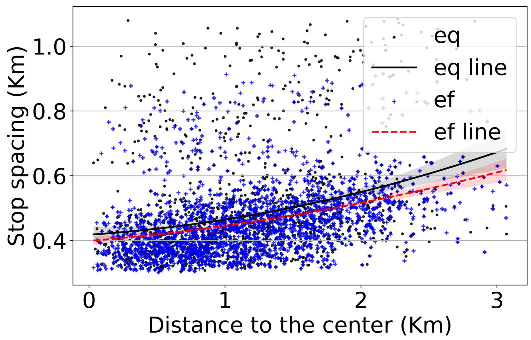


Figure 7: Relation between the distance to the nearest stop and the distance to the center

## 4 Conclusions

We proposed a method to embed the inequality of the distribution of accessibility into the decisions within a what-to PT planning strategy. We applied it the stop selection problem. Numerical results show that significant reduction of inequality of accessibility is achievable, which encourages to pursue future work, applying such a method on other dimensions of PT planning.

## References


Ali, M., et al. (2018). A case study on the walking speed of pedestrian at the bus terminal area. *E3S Web of Conferences*.

Badeanlou, et al. (2022). Assessing transportation accessibility equity via open data. In *heart*.

Bai, Z., et al. (2015). Characterizing spatial patterns of population distribution at township level across the 25 provinces in china. Retrieved from http://www.geog.com.cn/CN/10.11821/dlxb201508004

Biazzo, I., Monechi, B., & Loreto, V. (2019). General scores for accessibility and inequality measures in urban areas. *Royal Society open science*.

Camporeale, R., Caggiani, L., & Ottomanelli, M. (2019). Modeling horizontal and vertical equity in the public transport design problem: A case study. *Transportation Research Part A*, *125*, 184–206.

Costa, R. N., et al. (2019). *Not all inequality measures were created equal: The measurement of wealth inequality, its decompositions, and an application to european household wealth* (ECB Statistics Paper No. 31). Frankfurt a. M..

Dai, et al. (2022). Accessibility-based ethics-aware transit design.

Desaulniers, G., & Hickman, M. D. (2007). Chapter 2 public transit. In C. Barnhart et al. (Eds.), *Transportation*. Elsevier.

Furth, et al. (2000). Optimal bus stop spacing through dynamic programming and geographic modeling. *Transportation Research Record*.

Kim, M., Kho, S.-Y., & Kim, D.-K. (2019). A transit route network design problem considering equity. *Sustainability*.

Miller, E. (2020). Measuring accessibility. *OECD*.

Muth, R. F. (1967). The distribution of population within urban areas. In *Determinants of investment behavior.* NBER.

Pinto, H., et al. (2020). Joint design of multimodal transit networks and shared autonomous mobility fleets. *Transp. Res. Part C*.

Spiess, H., & Florian, M. (1989). Optimal strategies: a new assignment model for transit networks. *Transportation Research Part B: methodological*.

Wang, D. (2023). *Code.* https://github.com/Duowang99/PT-stop-selection-for-equality-of-accessibility.

Yoo, S., & Lee, J. B. (2023). Revising bus routes to improve access for the transport disadvantaged: A reinforcement learning approach. *Journal of Public Transportation*.