

Highlights

Metro 3 in Brussels under uncertainty: scenario-based public transport accessibility analysis

Brecht Verbeken, Arne Vanhoyweghen, Vincent Ginis

- GTFS-based simulation of Brussels public transport accessibility.
- Three Metro 3 scenarios: status quo, half build, and full build.
- Accessibility measured between 647 evenly spaced 500 m points.
- Temporal robustness with multiple time slices and day types.
- Accessibility improvements cluster in a subset of more population dense and disadvantaged neighborhoods.

Metro 3 in Brussels under uncertainty: scenario-based public transport accessibility analysis

Brecht Verbeken^{a,c,*}, Arne Vanhoyweghen^{a,c}, Vincent Ginis^{a,b,c}

^a*Department Business Technology and Operations, Data Analytics Laboratory, Vrije Universiteit Brussel (VUB), Pleinlaan 2, Brussels, 1050, Belgium*

^b*School of Engineering and Applied Sciences, Harvard University, Cambridge, 02138, Massachusetts, USA*

^c*imec-SMIT, Vrije Universiteit Brussel, Pleinlaan 9, Brussels, 1050, Belgium*

Abstract

Metro Line 3 in Brussels is one of Europe’s most debated infrastructure projects, marked by escalating costs, delays, and uncertainty over completion. Yet no public accessibility appraisal exists to inform this policy debate. This paper provides a scenario-based assessment of the spatial and distributional accessibility impacts of Metro 3 using schedule-based public transport data. Official GTFS feeds from all regional operators were combined and adapted to represent three scenarios: (i) the current network, (ii) partial implementation of Metro 3 (conversion of the southern premetro section), and (iii) full implementation to Bordet. Accessibility was measured as public transport travel time between 647 evenly spaced 500 m points covering the Brussels-Capital Region. Simulations were conducted for morning, evening, and weekend midday periods, each with three departure times ($t - 10$, t , $t + 10$), capturing robustness to short-term timetable variation. Results show substantial but uneven accessibility gains, with the largest improvements occurring in neighborhoods with below-average incomes. Temporal robustness analysis highlights variability in accessibility outcomes, underscoring the need to account for uncertainty in departure timing. These findings suggest that Metro 3 has the potential to reduce socio-spatial inequalities in accessibility, providing transparent evidence for a project where public debate is dominated by cost concerns.

Keywords: Public transport accessibility, Scenario analysis, GTFS simulation, Metro 3, Brussels, Spatial inequality

*Corresponding author.

Email addresses: `brecht.verbeken@vub.be` (Brecht Verbeken), `arne.vanhoyweghen@vub.be` (Arne Vanhoyweghen), `vincent.ginis@vub.be` (Vincent Ginis)

1. Introduction

Metro Line 3 in Brussels, an envisioned north–south automated metro linking Forest (south) to Evere (north), has become one of the region’s most debated transport investments. Touted as a remedy for congestion and a boost to urban mobility, the project has been plagued by spiraling costs and delays. Initial cost estimates of €0.85 billion (2009) had risen to €2.3 billion by 2022; by 2023, projections stood at around €4.7 billion [1], while the most recent scenario analyses estimate the total could reach nearly €5 billion or even rise to €7 billion under a public–private–partnership model [2].

Construction setbacks (notably at Palais du Midi and Brussels-North) and permit complications have repeatedly pushed timelines, amplifying concerns about feasibility and phasing [2, 3]. As of 2025, both scope and financing remain unsettled, with scenarios on the table ranging from a full build-out to a partial conversion or a halt [4, 5].

In an internal report, the transit authority (STIB-MIVB) outlined options including a full build-out from Albert to Bordet (requiring an additional €3.6 billion investment) and a partial “slimmed-down” option converting only the existing pre-metro tram tunnel between Albert and Brussels-North (costing around €1 billion) [4]. Despite the magnitude of Metro 3’s impacts and costs, no public cost–benefit analysis or accessibility appraisal has been released to justify the project’s socio-economic merit. Indeed, some policymakers have argued that an updated assessment of benefits is urgently needed given that the line’s cost “is much more expensive than initially estimated” and circumstances such as teleworking patterns have changed [6]. In conclusion no published, reproducible analysis quantifies Metro 3’s accessibility impacts.

Within transport geography, this study is situated in the literature emphasizing accessibility as a key metric for evaluating transport and land-use projects. Accessibility, the ease of reaching valued opportunities, provides a people-centered perspective on infrastructure impacts, in contrast to narrow mobility indicators [7].

Following this paradigm, many scholars have called for major projects like new transit lines to be judged by their contributions to accessibility and spatial equity rather than just travel time savings or ridership. In the context of Brussels and Belgium, several studies have mapped and analyzed baseline accessibility patterns. For example, Vandenbulcke et al. [8] compared accessibility structures across Belgium and found significant regional disparities. Focusing on the capital, Lebrun [9] measured public transport travel times to key activity centers in Brussels and revealed pronounced intra-urban heterogeneity: roughly 30% of the Brussels population live

in neighborhoods with poor transit accessibility to jobs and services. In Brussels, where lower-income districts have lower car ownership and higher public-transport dependence, project appraisal should consider whether new lines deliver clear travel-time reductions in these areas, beyond improvements in citywide averages. However, to date no study has specifically examined how Metro 3 would reshape the accessibility landscape of Brussels. The present work aims to fill this gap by providing an accessibility-focused evaluation of Metro 3’s potential impacts, thereby contributing to the broader discourse on urban spatial inequality and transport justice [10, 11].

Methodologically, our study builds on recent advances in accessibility simulation using General Transit Feed Specification (GTFS) data. The proliferation of GTFS (standardized public transport schedules) and open-source routing engines now enables detailed, schedule-based accessibility analyses for both existing and planned transit networks. Researchers have demonstrated the value of GTFS-driven analysis for transit planning: for instance, Conway et al. [12] developed a public transport sketch planning tool that accounts for uncertainty in travel times, while Pereira et al. [13] employed GTFS-based cumulative opportunity measures to evaluate future Bus Rapid Transit scenarios in Rio de Janeiro. Such tools make it possible to construct “counterfactual” schedule scenarios and compare accessibility outcomes across different network configurations.

In this study, we leverage GTFS-based routing to simulate three scenarios for Metro 3 within the broader framework of existing public transport in the Brussels-Capital Region:

1. *status quo* scenario with no metro (the existing tram corridor retained)
2. *partial conversion* scenario (trams replaced by metro between Albert and Brussels-North, but no extension beyond Brussels-North)
3. *full metro* scenario (metro line running from Albert through the new tunnel to Bordet)

By generating GTFS feeds for each scenario and computing multi-modal travel times, we can quantify how Metro 3’s partial or full implementation would change access to opportunities across the Brussels metropolitan area. To our knowledge, this is one of the first applications of GTFS-driven accessibility modeling for an unreleased major transit project in Europe, and it demonstrates a novel approach to ex-ante transit appraisal.

A further innovation of our approach is the consideration of *temporal robustness* in accessibility outcomes, informed by the concept of ergodicity breaking. Typically, accessibility studies evaluate a single departure time (often the morning peak hour) as representative of network performance. In reality, however, travel times and transit

service levels vary across the day and week, meaning a one-shot temporal sample may not reflect a traveller’s everyday experience. Crucially, accessibility is generated by a non-ergodic process: the time-average along an individual’s trajectory generally diverges from the ensemble snapshot at an arbitrary instant. A single “representative” departure therefore estimates the wrong object for welfare and behavior; our multi-time, multi-day sampling instead approximates the relevant time-average and exposes downside risk masked by snapshots [14, 15].

In non-ergodic systems, the time-average experience of an individual can differ markedly from the ensemble average at an arbitrary moment. Concretely, we calculate accessibility at three departure times in the morning peak (the scheduled time t and $t \pm 10$ min) and across three representative day-types (weekday, evening, and Saturday) to capture variability in service schedules. By evaluating nine time/day combinations per scenario, we obtain a distribution of accessibility outcomes that reflects typical fluctuations.

By quantifying accessibility under three realistic scenarios, this study offers the first reproducible evidence of Metro 3’s impacts, translating technical planning assumptions into implications that matter for both decision-makers and citizens.

In summary, this paper addresses three key gaps in the literature. First, it presents the first accessibility analysis of the proposed Brussels Metro 3, offering evidence on the project’s spatial impacts in the absence of any official public appraisal. Second, it advances the use of GTFS-based scenario modeling for transport infrastructure evaluation, illustrating how future “what-if” transit networks can be rigorously assessed using open data and reproducible methods. Such counterfactual accessibility modeling remains relatively uncommon in transport geography, and our case study demonstrates its potential value for ex-ante policy analysis [13, 16]. Third, it introduces a temporal robustness perspective into accessibility analysis, addressing the limited attention of departure-time variability in most accessibility studies [12, 17, 18]. By doing so, we respond to recent calls for more resilient and temporally aware accessibility metrics that better reflect travelers’ lived experiences. Collectively, these contributions aim to enrich both the empirical understanding of Metro 3’s implications and the methodological toolbox for transport geographers and planners.

The remainder of this article is structured as follows. Section 2 describes the study context and data, including the construction of GTFS scenarios for the Metro 3 options. Section 3 details the methodology, including accessibility measures, the multi-temporal simulation procedure, and analytical metrics for comparing scenarios. Section 4 presents the results of the accessibility simulations, highlighting the spatial distribution of benefits under each Metro 3 scenario and the sensitivity to temporal

variability. Finally, Section 5 discusses the implications of these findings as well as the limitations of our approach, we conclude with a summary of key insights and recommendations for policy and future research.

2. Study area, data, and scenarios

The study area is the *Brussels–Capital Region (BCR)*. We compute public-transport travel times on the full multimodal network (*metro, tram, bus, and rail*) by integrating the official GTFS feeds for *STIB/MIVB*, *De Lijn*, *TEC*, and *SNCB/NMBS* [19, 20, 21, 22]. Rail is retained both because the suburban Brussels S Train provides dense intra-urban stopping patterns within the BCR and because commuting flows from the wider metropolitan area interact with intra-BCR accessibility.

Simulations are anchored to three *representative 2025 service days* and, for each, three departure instants to capture short-horizon timetable sensitivity:

- **Weekday AM peak:** Tue 10 June 2025 at 07:50, 08:00, 08:10.
- **Weekday PM peak:** Thu 12 June 2025 at 17:20, 17:30, 17:40.
- **Saturday mid-day:** Sat 12 July 2025 at 12:50, 13:00, 13:10.

These slices implement the canonical peak/off-peak bands used in schedule-based accessibility studies; the *partial* and *full* Metro 3 scenarios necessarily adapt the STIB/MIVB feed relative to these base calendars (details in §2.2). Evaluating $t-10$, t , and $t+10$ per band operationalises the paper’s temporal robustness principle (cf. §1).

2.1. Spatial sampling, OD construction and data sources

Origins and destinations are the centres of 647 tiles on a regular 500×500 m lattice clipped to the BCR. This resolution balances coverage and tractability. The expected straight-line distance from a random point in a 500 m square to its centre is ≈ 191 m; at typical urban walking speeds (≈ 1.3 m/s), this corresponds to ≈ 2.5 min. Hence the origin-distance (OD) paths between tile centres approximate real journeys with an *expected* access+egress overhead of ≈ 5 min, delivering citywide coverage while keeping the OD matrix computationally feasible.

We use four GTFS feeds (all accessed 2024-06-10): STIB/MIVB, De Lijn, TEC, and SNCB/NMBS [19, 20, 21, 22]. The pedestrian network for access/egress is derived from OpenStreetMap (Geofabrik extract `belgium-latest.osm.pbf`) [23, 24]. Feeds are loaded into a single multimodal graph with `r5py/R5` [25], enabling inter-agency transfers at co-located stops. Following established schedule-based accessibility practice [26, 12, 27], we evaluate *earliest arrival time*.

2.2. Scenario construction (*status quo*, *partial*, *full*)

This study evaluates three network configurations: (i) *status quo* (premetro retained), (ii) *partial conversion* (Albert–Brussels–North converted to metro), and (iii) *full Metro 3* (Albert–Bordet). Each scenario is implemented as a derived GTFS feed obtained by editing the base feeds for the 2025 slices above.

Our primary source for scenario design was the official Metro 3 project site [28]. Public communication is sparse and occasionally inconsistent, so we complemented the published material with limited engineering inference to produce internally coherent GTFS scenario feeds.

Full Metro 3. We encoded the line using the stations published by the project, placed at the coordinates communicated in official materials: Albert, Horta, Parvis de Saint-Gilles, Porte de Hal, Gare du Midi, Toots Thielemans, Anneessens, Bourse, De Brouckère, Rogier, Gare du Nord (Brussels–North), Liedts, Colignon, Verboekhoven, Riga, Tilleul, Paix, Bordet. Inter-station run times were derived from great-circle distances and calibrated against performance on existing Brussels metro lines, yielding an end-to-end runtime just under 20 minutes, in line with official statements [28]. Because the tram network is slated to be restructured if Metro 3 opens, we also reflected the operator’s patterns in this scenario: Tram 7 is extended from Vanderkindere to Albert; Tram 4 is curtailed to Stalle–Albert; and Tram 10 (successor to the former Tram 3) is adjusted to operate between Esplanade and Rogier along its published alignment. For trams, inter-station run times were taken from observed segment times on the corresponding existing lines. Service frequency (headways and daily runs) was replicated from current STIB/MIVB timetables for those lines.

Partial conversion (Albert–Brussels–North). We implemented metro service only between Albert and Brussels–North (Albert, Horta, Parvis de Saint-Gilles, Porte de Hal, Gare du Midi, Toots Thielemans, Anneessens, Bourse, De Brouckère, Rogier, Gare du Nord (Brussels–North)), using the same dwell and run-time assumptions as on the corresponding segment of the full line. In line with the operator’s public remarks about interim operational constraints, including the absence of a local depot on this line, we reduced metro supply by approximately 20% on the affected corridor in this scenario [4]. A key constraint is the absence of a dedicated depot. Operationally, rolling stock must be positioned from elsewhere each day, which limits staging and turnbacks and caps the maximum number of trainsets that can be in service simultaneously. Trams 4, 7, and 10 follow the same routing patterns as in the full-line scenario.

Computation OD travel times between all tile pairs are computed for every

scenario×time slice using batched `r5py`/R5 runs . Computations ran on the Vlaams Supercomputer Centrum (VSC); cumulative workload amounted to ≈ 4 CPU-years.

3. Methods: data analysis, robustness, and scenario comparison

3.1. Raw simulation outputs and panel structure

For each scenario $s \in \{\text{baseline}, \text{partial conversion}, \text{full metro 3}\}$, for each day-type $d \in \{\text{AM-peak}, \text{PM-peak}, \text{Sat}\}$, and for each departure instant $k \in \{-10, 0, +10\}$ within that day-type, the routing engine returns directed door-to-door public-transport travel times between spatial samples (tile centroids) $i, j \in \mathcal{S}$:

$$T_{s,d,k}(i, j) \in \mathbb{R}_{\geq 0} \cup \{\text{NA}\}, \quad i, j = 1, \dots, N.$$

Rows in the analysis table have the schema

$$(o_i, d_j, s, \text{snapshot_time} = k, \text{time_s} = T_{s,d,k}(i, j), \text{day}, \text{start_time}),$$

e.g. `origin=(4.2493,50.8195), destination=(4.2564,50.8195), scenario=baseline, snapshot_time=t-10, time_s=3119, day=0610, start_time=0750`. Eight spatial samples were structurally unreachable (six in Laeken Park, one within NATO grounds in Evere, and one between the tracks at Bruxelles-Midi), so we drop them as both origins and destinations. This yields $N = 639$ nodes and $M = N(N - 1) = 639 \times 638$ directed OD pairs per slice, for $K = 27$ slices overall, i.e. $M \times K$ rows in the long panel.

3.2. Distributional diagnostics: quantiles, shift functions, and $\Delta ECDF$

Rather than only reporting a single mean effect, we characterise how *the entire distribution* of OD times shifts under each scenario, using three complementary, nonparametric diagnostics:

(i) *Deciles of ΔT* . For each comparison s , we report the deciles of ΔT_s across OD pairs $(i, j) \in \mathcal{S}$:

$$Q_s^{(\Delta)}(p) = \text{Quantile}(\{\Delta T_s(i, j) : (i, j) \in \mathcal{S}\}, p), \quad p \in \{0, 0.1, \dots, 1\}.$$

This shows heterogeneity: e.g. whether the 90th percentile OD pair still improves. We also compute key shares (in %):

$$\begin{aligned} \text{share}_{\text{improved}} &= \frac{\#\{(i, j) \in \mathcal{S} : \Delta T_s(i, j) < 0\}}{|\mathcal{S}|}, \\ \text{share}_{\geq 5\% \text{ better}} &= \frac{\#\left\{(i, j) \in \mathcal{S} : \frac{\Delta T_s(i, j)}{\bar{T}_{\text{baseline}}(i, j)} \leq -0.05\right\}}{|\mathcal{S}|}. \end{aligned}$$

Here, $\#\{\cdot\}$ denotes the cardinality (number of elements) of the set.

comparison	mean Δt (s)	median Δt (s)	improved (%)	equal $\pm 1s$ (%)	$\geq 5\%$ better (%)	$\geq 10\%$ better (%)
partial conversion	-16.5	0	49.8	28.8	5.6	0.7
full metro 3	-36.1	-6.1	52.8	28.1	9.4	2.8

Table 1: Summary statistics of travel time changes relative to the status quo. Partial conversion yields modest mean savings and limited shares of substantial improvements, while the full metro 3 delivers larger average reductions and a greater proportion of OD pairs with improvements exceeding 5% and 10%.

(ii) *Shift function of aggregated times.* Let $A = \{\tilde{T}_{\text{baseline}}(i, j) : (i, j) \in \mathcal{S}\}$ and $B = \{\tilde{T}_s(i, j) : (i, j) \in \mathcal{S}\}$. The *shift function* compares quantiles of B and A across $p \in (0, 1)$:

$$\Delta Q_s(p) = Q_B(p) - Q_A(p), \quad Q_X(p) = \text{quantile}(X, p).$$

Unlike mean differences, $\Delta Q_s(p)$ exposes where in the distribution the scenario helps or harms (e.g. large gains for slow OD pairs but little change for already fast pairs). This is standard in robust distributional comparisons [29].

(iii) *Difference of empirical CDFs ($\Delta ECDF$).* With empirical CDFs

$$\hat{F}_A(t) = \frac{\#\{(i, j) \in \mathcal{S} : \tilde{T}_{\text{baseline}}(i, j) \leq t\}}{|\mathcal{S}|},$$

$$\hat{F}_B(t) = \frac{\#\{(i, j) \in \mathcal{S} : \tilde{T}_s(i, j) \leq t\}}{|\mathcal{S}|}.$$

we plot

$$\Delta \hat{F}_s(t) = \hat{F}_B(t) - \hat{F}_A(t).$$

Positive values at a given t mean a larger share of OD pairs can be completed in $\leq t$ seconds under the scenario. This is closely related to the Kolmogorov–Smirnov perspective on distributional change [30, 31].

3.3. Spatial visualisation of per-origin time benefits

From OD panels to per-origin outbound changes. For each scenario $s \in \{\text{partial}, \text{full}\}$ we summarise the *outbound* accessibility of an origin i by the average of OD times

comparison	p0 Δt (s)	p20 Δt (s)	p40 Δt (s)	p60 Δt (s)	p80 Δt (s)	p100 Δt (s)
partial conversion	-693	-76	-20	0	40	551
full metro 3	-1703	-93	-29	0	35	567

Table 2: Percentile distribution of travel time changes (Δt , in seconds) relative to the status quo. The full metro 3 scenario exhibits more extreme improvements in the lower tail while both scenarios show similar upper-tail increases; see also Figure 1c.

to all destinations:

$$\bar{T}_{i,s} = \frac{1}{M_i} \sum_{\substack{j=1 \\ j \neq i}}^N \bar{T}_s(i, j), \quad M_i = \#\{j \neq i : \bar{T}_s(i, j) \in \mathbb{R}\}.$$

Here $\bar{T}_s(i, j)$ is the *mean* over slices

$$\bar{T}_s(i, j) = \frac{1}{K} \sum_{k \in \mathcal{K}} T_{s,k}(i, j),$$

with $K = 3$ and $\mathcal{K} = \{-10, 0, +10\}$. We then define the *per-origin time change* (seconds) relative to the status quo:

$$\Delta_{i,s}^{\text{out}} = \bar{T}_{i,s} - \bar{T}_{i,\text{baseline}}, \quad \Delta_{i,s}^{\text{out}} < 0 \Rightarrow \text{faster (benefit)}.$$

An analogous day-specific version aggregates only within day-type/date d :

$$\bar{T}_{s,d}(i, j) = \frac{1}{K} \sum_{k \in \mathcal{K}} T_{s,d,k}(i, j), \tag{1}$$

$$\bar{T}_{i,s}^{(d)} = \frac{1}{M_i^{(d)}} \sum_{j \neq i} \bar{T}_{s,d}(i, j), \tag{2}$$

$$\Delta_{i,s}^{\text{out}}(d) = \bar{T}_{i,s}^{(d)} - \bar{T}_{i,\text{baseline}}^{(d)}. \tag{3}$$

Practically, the code groups the long table by (**scenario**, **origin**, **destination**), collapses to the mean over slices, then averages across destinations to get $\bar{T}_{i,s}$; deltas $\Delta_{i,s}^{\text{out}}$ are computed on the *common origin support* with the baseline (inner merge).

Per-origin outbound variability (Δ SD) maps. In addition to mean-time changes, we map changes in *temporal dispersion*. For each (i, s) we take the population variance across slices at the OD level and aggregate as

$$\text{Var}_{s,d}(i, j) = \text{Var}_{k \in \mathcal{K}}[T_{s,d,k}(i, j)] \quad (\text{OD pair variability, day } d), \quad (4)$$

$$V_{i,s}^{(d)} = \frac{1}{M_i^{(d)}} \sum_{j \neq i} \text{Var}_{s,d}(i, j) \quad (\text{avg. OD variability, day } d), \quad (5)$$

$$\bar{V}_{i,s} = \frac{1}{|D|} \sum_{d \in D} V_{i,s}^{(d)} \quad (\text{avg. OD variability across days}), \quad (6)$$

$$\sigma_{i,s} = \sqrt{\bar{V}_{i,s}} \quad (\text{st.dev. of avg. variability}). \quad (7)$$

$$\Delta_{i,s}^{\text{SD}} = \sigma_{i,s} - \sigma_{i,\text{baseline}} \quad (\text{seconds; } \Delta_{i,s}^{\text{SD}} < 0 \Rightarrow \text{less variability}). \quad (8)$$

Values are computed on the *common origin support*. We report the variance only at the scenario level, since each scenario provides nine data points, whereas day-wise comparisons rely on only three data points and thus offer limited statistical interpretability.

Grid geometry. Origins are the centroids of a 500×500 m regular lattice clipped to the BCR, yielding square tiles via

$$\mathcal{P}_i = \{(x, y) : |x - x_i| \leq 250 \wedge |y - y_i| \leq 250\}.$$

Expected centres without a nearby observed origin (within 35 m) are flagged as grid *holes*. These holes are rendered as black squares on maps to make data gaps legible (the eight unreachable samples: Laeken Park ×6, NATO Evere, track interstice at Bruxelles-Midi). We associate the corresponding $\Delta_{i,s}^{\text{out}}$ with each tile and color the tiles according to this value. Commune boundaries are overlaid for orientation only.

Map outputs.. We render (i) an *all-dates* map per scenario using $\Delta_{i,s}^{\text{out}}$, and (ii) per-daycode maps using $\Delta_{i,s}^{\text{out}}(d)$ for each snapshot date (labels derived from MMDD codes).

3.4. Directionality diagnostics and justification for outbound focus

Matrix setup and decomposition. For each scenario s , let $A_s \in \mathbb{R}^{N \times N}$ be the directed OD travel-time matrix returned by the simulator, with $(A_s)_{ij}$ the door-to-door public transport time (seconds) from origin i to destination j on the $N = 639$ grid nodes (eight unreachable samples dropped). We decompose

$$A_s = S_s + K_s, \tag{9}$$

$$S_s = \frac{1}{2}(A_s + A_s^\top), \quad (\text{symmetric}), \tag{10}$$

$$K_s = \frac{1}{2}(A_s - A_s^\top), \quad (\text{skew-symmetric}). \tag{11}$$

which are orthogonal under the Frobenius inner product (so $\|A_s\|_F^2 = \|S_s\|_F^2 + \|K_s\|_F^2$). S_s captures direction-free structure; K_s captures pure directionality ($i \rightarrow j$ vs $j \rightarrow i$).

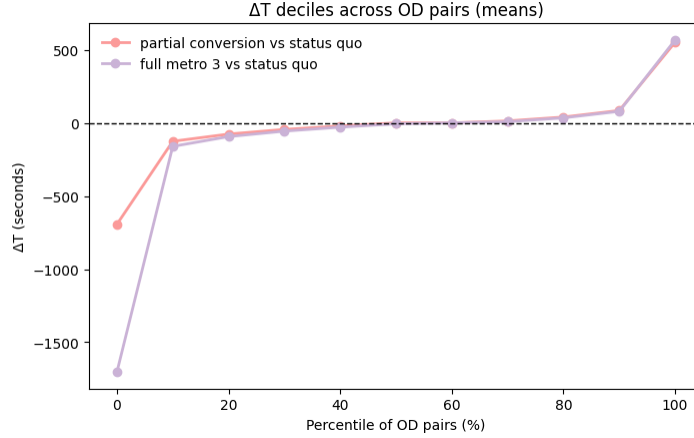
Directionality Index and pairwise thresholds.. We quantify system-wide asymmetry by the scale-free

$$\text{DI}(s) = \frac{\|K_s\|_F}{\|S_s\|_F},$$

We obtained a DI of around 0.035, which implies that directionality only explains

$$\frac{\|K_s\|_F^2}{\|A_s\|_F^2} = \frac{\text{DI}(s)^2}{1 + \text{DI}(s)^2} \approx 0.12\%$$

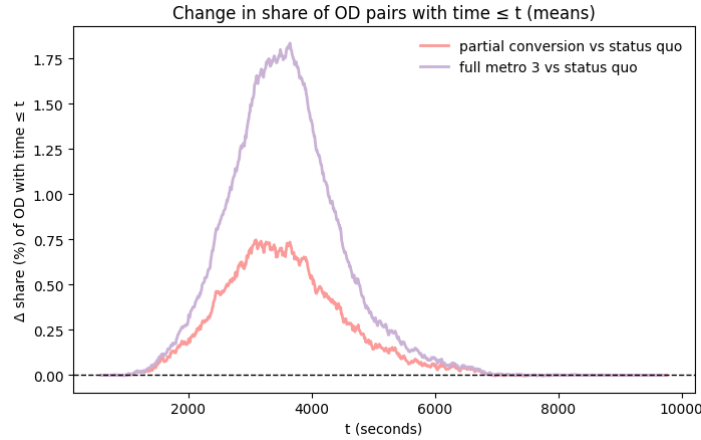
of OD variance. I.e., the OD matrix is *nearly symmetric*. Accordingly, we report outbound results only: with $\text{DI}(s) = 0.035$ (so directionality explains $\approx 0.12\%$ of $\|A_s\|_F^2$), inbound patterns are effectively identical up to the small skew component K_s .



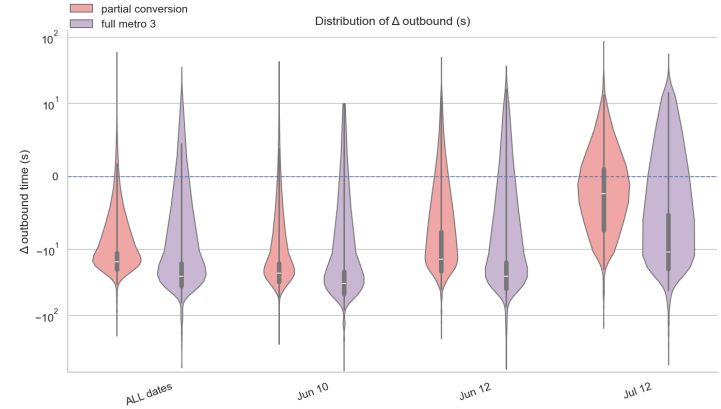
(a) **Deciles of OD-pair travel time differences.** For each OD pair we compute $\Delta T = T_{\text{scenario}} - T_{\text{baseline}}$, then summarize the distribution of these differences across all OD pairs by its deciles. Negative values denote time savings. This captures the distribution of *pairwise changes* in travel times.



(b) **Shift function of OD time distributions.** For the entire distribution of OD times under each scenario, we compute the quantile function $Q(p)$ of aggregated travel times and then plot $\Delta Q(p) = Q_{\text{scenario}}(p) - Q_{\text{baseline}}(p)$ over $p \in [1, 99]$. Negative values denote time savings.



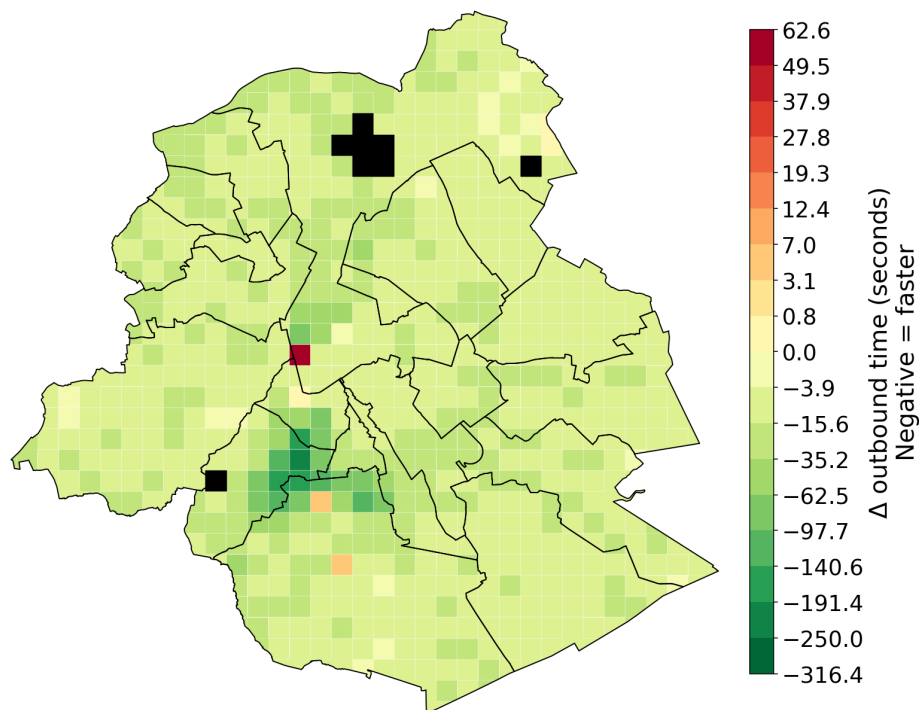
(c) **Change in share of OD pairs completed within t .** $\Delta \hat{F}(t) = \hat{F}_{\text{scenario}}(t) - \hat{F}_{\text{baseline}}(t)$, plotted in percentage points. Positive values mean more OD pairs are reachable within t .



(d) **Distribution of per-origin Δ outbound time.** Violin plots of $\Delta^{\text{out}} = \bar{T}_{\text{scenario}} - \bar{T}_{\text{baseline}}$. Symmetric log y -scale (linear within ± 10 s).

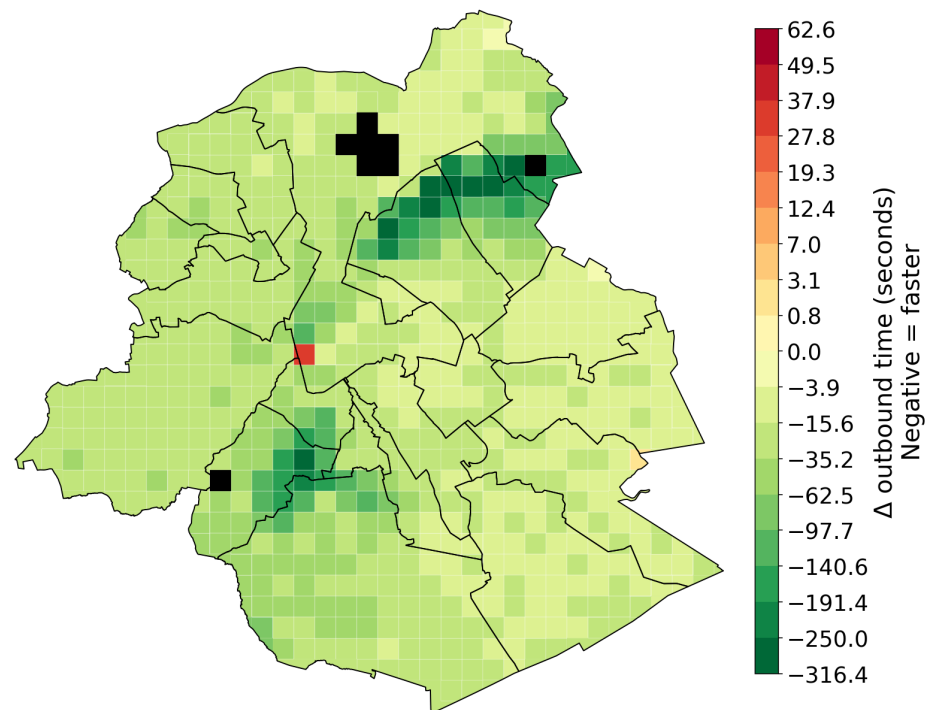
Figure 1: **Distributional diagnostics for time benefits.** Panels (a)–(b) use common-support OD pairs and scenario-aggregated OD times; negative values indicate faster travel. Together, they show broad, distribution-wide improvements for the full line and smaller gains for the partial conversion; panel (c) shows the difference in reachability within time t ; panel (d) summarises the same effects at the per-origin level.

Δ outbound (s) — partial conversion vs status quo (ALL)



(a) All-days composite — partial conversion

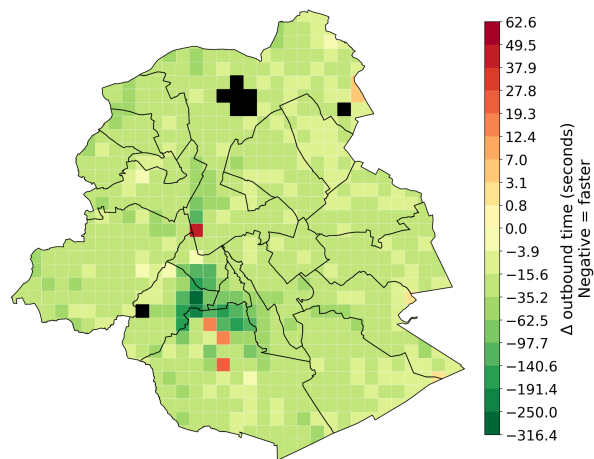
Δ outbound (s) — full metro 3 vs status quo (ALL)



(b) All-days composite — full line

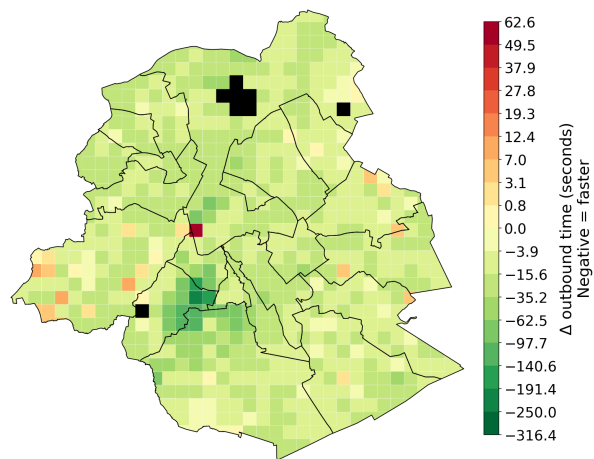
Figure 2: **Systemwide maps.** All-days composites of differences in means for partial conversion and full line.

Δ outbound (s) — partial conversion vs status quo (Jun 10)



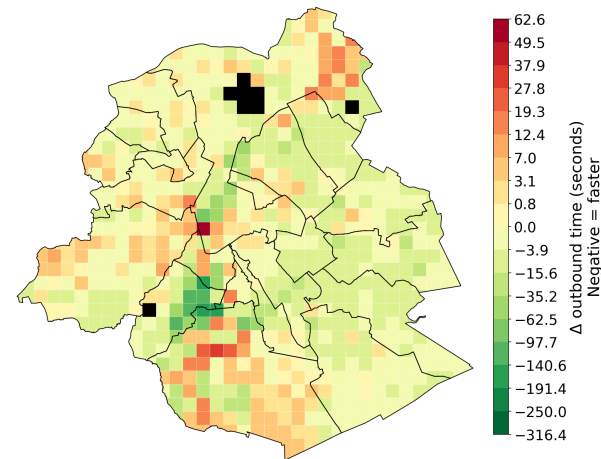
(a) 0610 — partial

Δ outbound (s) — partial conversion vs status quo (Jun 12)



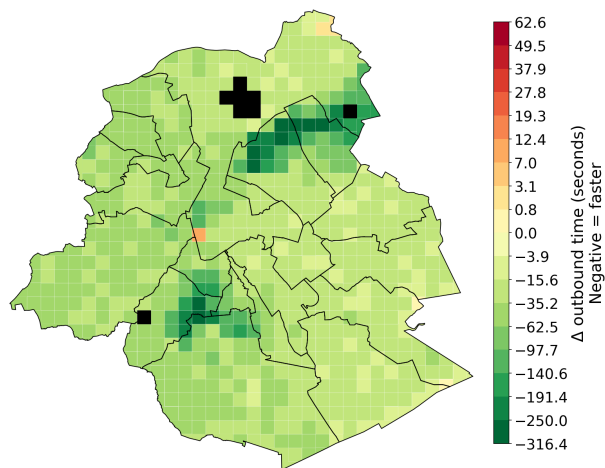
(b) 0612 — partial

Δ outbound (s) — partial conversion vs status quo (Jul 12)



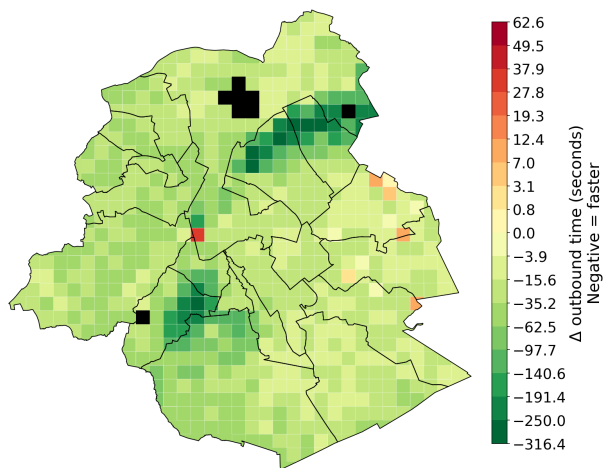
(c) 0712 — partial

Δ outbound (s) — full metro 3 vs status quo (Jun 10)



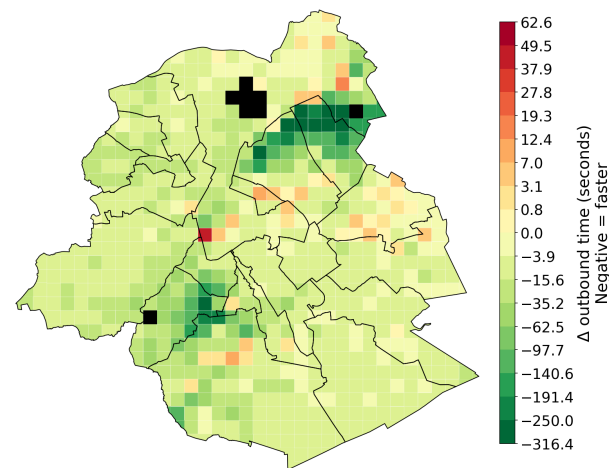
(d) 0610 — full

Δ outbound (s) — full metro 3 vs status quo (Jun 12)



(e) 0612 — full

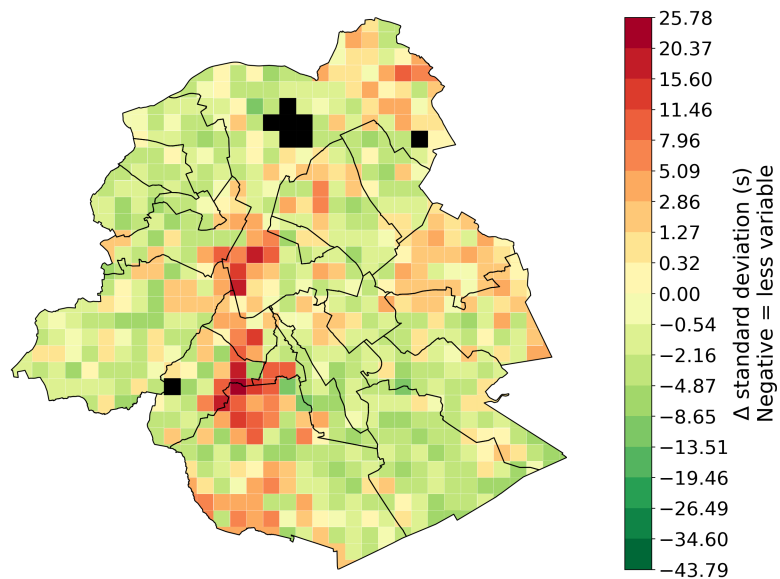
Δ outbound (s) — full metro 3 vs status quo (Jul 12)



(f) 0712 — full

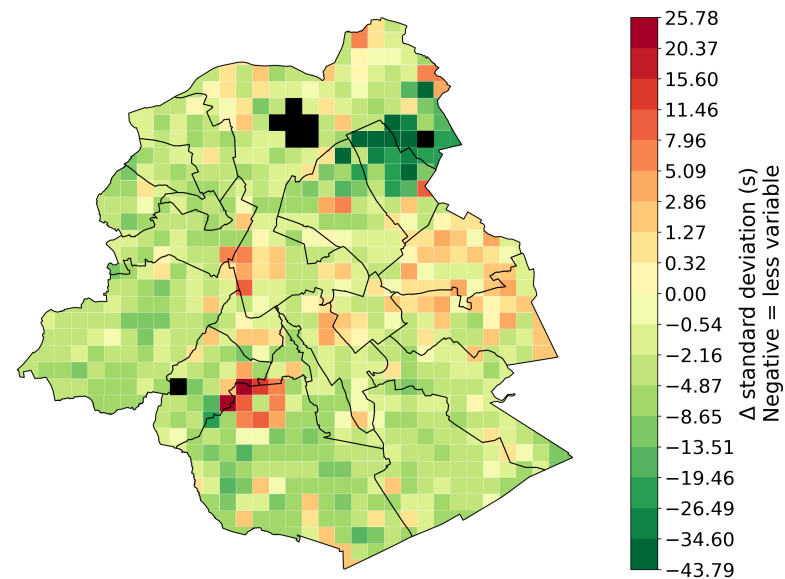
Figure 3: **Dated maps.** Daily snapshots by date (0610, 0612, 0712) and scenario (partial/full) of differences in means for partial conversion and full line.

Δ outbound time SD (s) — partial conversion vs status quo (ALL)



(a) All-days composite — partial conversion

Δ outbound time SD (s) — full metro 3 vs status quo (ALL)



(b) All-days composite — full conversion

Figure 4: **Systemwide maps.** All-days composites for differences in standard deviation for partial conversion and full line.

3.5. Reliability and risk-adjusted accessibility metrics

Per-OD, per-day reliability statistics. For each scenario s , OD pair (i, j) , and day d , let the three canonical departure slices be $\mathcal{K} = \{-10, 0, +10\}$ and denote by $T_{s,d,k}(i, j)$ the simulated travel time (s) at slice $k \in \mathcal{K}$. Define the empirical quantile

$$Q_{p,s,d}(i, j) = \text{Quantile}(\{T_{s,d,k}(i, j) : k \in \mathcal{K}\}, p).$$

We report

$$\text{median}_{s,d}(i, j) = Q_{0.50,s,d}(i, j), \quad (12)$$

$$\text{IQR}_{s,d}(i, j) = Q_{0.75,s,d}(i, j) - Q_{0.25,s,d}(i, j), \quad (13)$$

$$\text{RBI}_{s,d}^{\text{abs}}(i, j) = Q_{0.95,s,d}(i, j) - Q_{0.50,s,d}(i, j), \quad (14)$$

$$\text{RBI}_{s,d}^{\text{rel}}(i, j) = \frac{Q_{0.95,s,d}(i, j)}{Q_{0.50,s,d}(i, j)} - 1. \quad (15)$$

Here RBI^{abs} is an absolute “reliability buffer” (95th–50th) widely used in reliability assessment [32], while RBI^{rel} is the median-based Buffer Index (unitless; $\times 100$ for %), a robust variant analogous to the mean-based Buffer Index [33].

Risk-adjusted (entropic) certainty equivalent. Many OD pairs exhibit skew with occasional long delays; we therefore seek a single summary that rewards lower typical times while penalising tail risk, echoing the time-average perspective which highlights the structural importance of reliability alongside speed [15]. The entropic certainty equivalent answers: “What certain travel time would make a traveller, who values both travel speed and reliability, indifferent to the observed distribution?” Lower is better; the gap $\mathbb{E}[T] - \text{CE}_\rho[T]$ is the risk premium (in minutes) the traveller would give up to avoid variability.

To penalise tail risk we adopt the entropic certainty equivalent for travel time T :

$$\text{CE}_\rho[T] = -\frac{1}{\rho} \log \left(\mathbb{E}[e^{-\rho T}] \right),$$

with $\rho > 0$ calibrated by a “half-life” parameter. We fix the half-life at 20 min after testing multiple values ($\rho = \frac{\ln 2}{20 \times 60 \text{ s}} = \frac{0.693147}{1200 \text{ s}} \approx 5.78 \times 10^{-4} \text{ s}^{-1}$); results were qualitatively stable across choices, and 20 min offers a clear interpretation: an extra 20 min halves the effective weight.

A second-order expansion shows

$$\text{CE}_\rho[T] \approx \mathbb{E}[T] - \frac{\rho}{2} \text{Var}(T),$$

i.e. the mean minus a variance penalty that grows with ρ [34]. Numerically, we use a log-sum-exp stabilised estimator:

$$\widehat{\text{CE}}_\rho = -\frac{1}{\rho} \left(\log \frac{1}{L} \sum_{\ell=1}^L e^{-\rho T_{s,d,\ell}} \right).$$

Illustration (in minutes, with $\rho = \ln 2/20$): consider a travel time T that is 30 with probability 0.9 and 70 with probability 0.1. The mean is $\mathbb{E}[T] = 34$, while the certainty equivalent is

$$\text{CE}_\rho[T] \approx 32.25,$$

so the implied *risk premium* is $\mathbb{E}[T] - \text{CE}_\rho[T] \approx 1.75$ minutes. This reflects the fact that travellers prefer a sure 32.25 minutes to facing the 90%–10% gamble. By contrast, for a degenerate case $T \equiv 34$, the certainty equivalent coincides with the mean: $\text{CE}_\rho[T] = 34$.

Scenario-level reporting. Let $\mathcal{S}$ denote the OD pairs (as defined earlier) and $\mathcal{D}$ the service days. For $M \in \{p_{50}, \text{IQR}, \text{RBI}^{\text{abs}}, \text{RBI}^{\text{rel}}, \text{CE}_\rho\}$, define

$$\overline{M}_{s,d} = \frac{1}{|\mathcal{S}|} \sum_{(i,j) \in \mathcal{S}} M_{s,d}(i,j), \quad \Delta \overline{M}_{s,d} = \overline{M}_{s,d} - \overline{M}_{\text{base},d},$$

and report

$$\Delta \overline{M}_s = \frac{1}{|\mathcal{D}|} \sum_{d \in \mathcal{D}} \Delta \overline{M}_{s,d}.$$

Interpretation: $\Delta \overline{p}_{50s} < 0$ implies faster typical trips; $\Delta \overline{\text{RBI}}_s^{\text{abs}} < 0$ implies a smaller absolute buffer (seconds); $\Delta \overline{\text{RBI}}_s^{\text{rel}} < 0$ implies a smaller median-based Buffer Index (more reliable in relative terms); $\Delta \overline{\text{CE}}_s < 0$ implies a lower risk-adjusted travel time for penalty ρ .

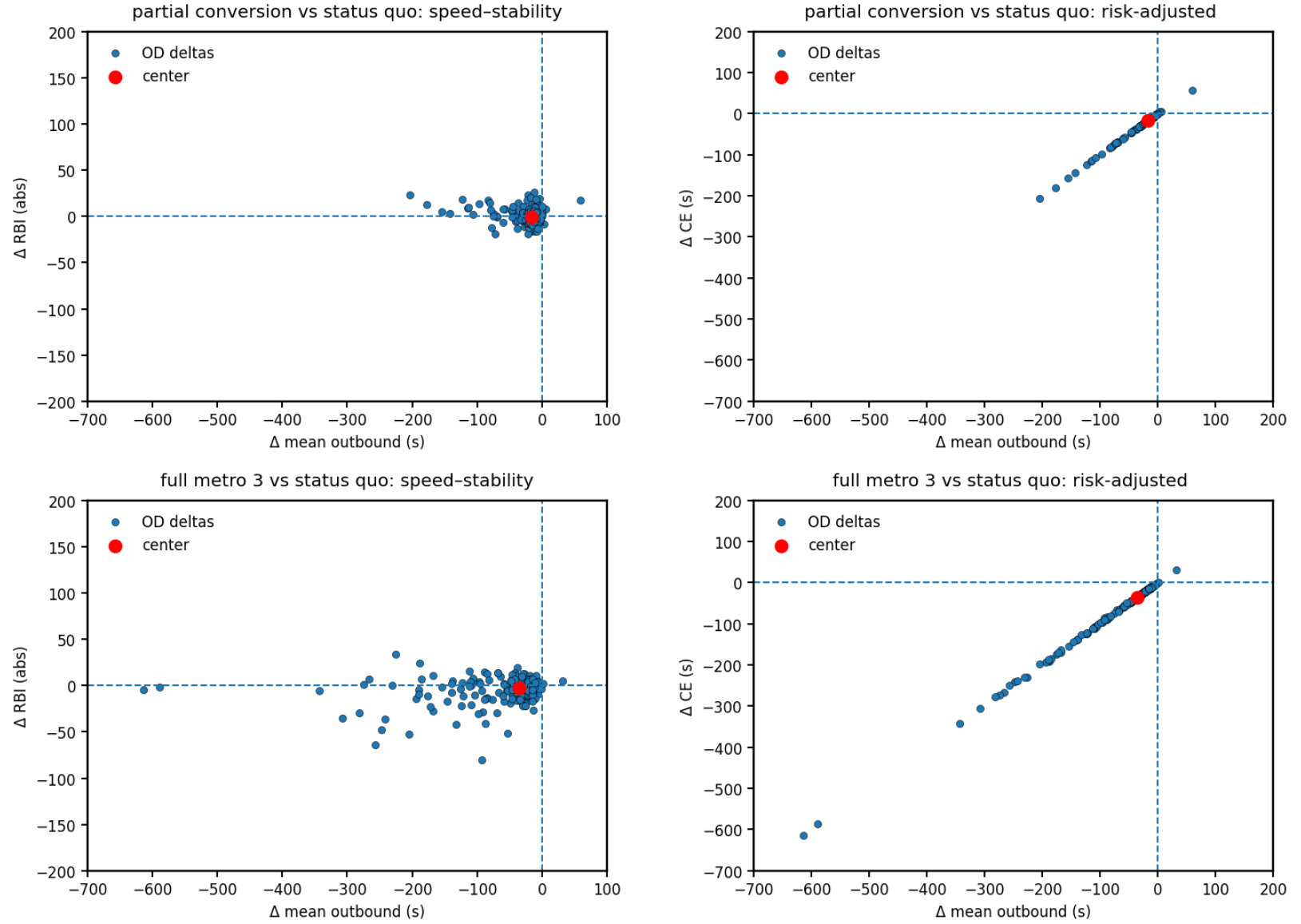


Figure 5: Mean effects across all days on OD pairs. Deltas are scenario – status quo; the red dot is the cloud mean. Left panels: Δ mean outbound (s) vs. $\Delta \text{RBI}_{\text{abs}}$ (s). Right panels: Δ mean outbound (s) vs. ΔCE (s), with CE the certainty-equivalent time under exponential utility, $\rho = \ln 2 / (20, \min)$. Dashed axes mark no change; left/down is better. Quadrants: $(-, -)$ win; $(-, +)$ faster/less stable; $(+, -)$ slower/more stable; $(+, +)$ dominated.

scenario	Δp_{50} (s)	ΔIQR (s)	ΔRBI_{abs} (s)	ΔRBI_{rel}	ΔCE (s)
partial conversion	-16.61	-0.61	-0.43	0.018%	-16.42
full metro 3	-36.34	-3.70	-3.01	-0.023%	-35.60

Table 3: Averages over the three service days of robustness-aggregated OD time-change metrics. Negative values indicate improvements.

4. Results

4.1. System-wide time benefits and their distribution

Table 1 summarises OD-level time changes relative to the status quo. The *partial conversion* reduces mean OD time by -16.5 s with a zero median; 49.8 % of OD pairs improve and 5.6 % improve by at least 5 %. The *full line* roughly doubles the mean effect, resulting in about half a minute gain for the average one-way commute (mean -36.1 s, median -6.1 s). The *full line* increases the share improved to 52.8 %, and lifts the share with ≥ 5 % gains to 9.4 % (and 2.8 % with ≥ 10 % gains).

Distributional diagnostics show that these averages hide considerable heterogeneity. Deciles of ΔT in Table 2 and Fig. 1a indicate broad left shifts up to about the 60th–70th percentile (improvements), with thin right tails (moderate slowdowns). Extreme improvements exist: the best 0.1 % of OD pairs save -700 s to -1700 s (partial vs. full). The *shift function* (Fig. 1b) makes this transparent: for the full line, the trough around $p \approx 0.4$ – 0.6 reaches ~ -50 s, versus ~ -20 s for the partial conversion. In other words, the bulk of trips get a small but systematic speed-up, larger under the full build.

The difference in empirical CDFs, $\Delta \hat{F}(t)$ (Fig. 1c), quantifies how many more OD pairs complete within a given time budget. Gains peak around $t \approx 3300$ s to 3800 s : roughly +1.8 pp for the full line and +0.75 pp for the partial conversion. These are distribution-wide effects, not driven by a handful of OD pairs.

4.2. Origin-level accessibility changes and robustness

Per-origin outbound changes, Δ^{out} (Fig. 1d), are centered below zero for all service days; the full line yields larger gains and slightly larger dispersion than the partial conversion. Robustness diagnostics using three departure instants per band (§3) show that the sign and magnitude of benefits are stable across the AM peak, PM peak, and Saturday; tails are heavier for the full line but remain predominantly on the benefit side.

Figures 5 show origin-averaged deltas (scenario – status quo) pooled over all service days. Left panels: Δ median vs. ΔRBI with $RBI = p_{95} - p_{50}$ (smaller =

more stable). Right panels: Δ median vs. Δ CE under exponential utility with a 20-min half-life (smaller = better risk-adjusted time). Dashed axes mark parity; left/down is improvement. Quadrants: lower-left = win; upper-left = faster/less stable (or worse CE); lower-right = slower/more stable (or better CE); upper-right = dominated. Other half-lives yield similar results as CE is dominated by the mean here.

Partial conversion yields modest average time savings (a few–tens of seconds) with near-zero change in stability; CE gains are small and track the mean closely. Full metro 3 shifts the cloud further left and delivers correspondingly larger CE improvements; stability is mostly neutral to slightly better, with some notable RBI drops. Overall the mass of points sits in the “win” quadrant with only a few mild regressions, and the near-linear CE–mean relationship implies distributional changes are limited relative to the central-tendency shift.

Partial conversion. Speed–stability: points concentrate tightly around the origin, with most lying slightly left but above the x -axis (Δ median $\approx -60 \dots +30$ s; Δ RBI $\approx +5 \dots +25$ s). Interpretation: small median gains are paired with modestly worse stability, and true reliability improvements are rare. Risk-adjusted: scatter straddles the horizontal axis with magnitudes generally small (Δ CE $\approx -50 \dots +20$ s). Net: effects are modest and mixed—minor speedups tend to be offset by degraded stability, leaving CE nearly unchanged.

Full conversion. Speed–stability: strong leftward shift with most points below zero on both axes (Δ median typically $-120 \dots -30$ s, several reaching -600 s; Δ RBI commonly $-10 \dots -60$ s). Interpretation: broad gains in both speed and stability; only a small minority trade stability for speed. Risk-adjusted: a pronounced down-left diagonal (Δ CE closely tracking Δ median at $\sim 1:1$), spanning $-50 \dots -600$ s. Nearly all origins improve CE substantially; dominated outcomes are rare.

We conclude that partial conversion yields modest gains in travel speed but risks deterioration in reliability, whereas full conversion delivers both substantial speed increases and consistent reliability improvements.

Aggregated reliability statistics (Table 3) confirm this: full-line averages are $\Delta p50 = -36.3$ s, Δ RBI = -3.0 s, Δ CE = -35.6 s; partial conversion effects are smaller in the same direction.

4.3. Where the gains are: spatial pattern

All-days composite maps (Fig. 2, Fig. 4) and day-specific mosaics (Fig. 3) show a consistent geography on the OD pairs:

- **Partial conversion.** Benefits concentrate along and south of the new line, with the strongest outbound improvements in *Forest* and *Saint-Gilles*, extend-

ing into northern *Uccle* (see Appendix B, Figure B.6 for the location of communes). Elsewhere, gains are light and scattered or near zero.

- **Full line.** The same southern core improves, and benefits propagate north–east into *northern Schaerbeek* and *Evere*, and north of *Saint-Josse-ten-Noode* and the *Ville de Bruxelles*, yielding a larger contiguous centre–north swath of improvements.

Per-day mosaics are more scattered than the composites, and the standard-deviation map (Fig. 4) makes the heterogeneity explicit: under partial conversion, pockets of deterioration appear in *Ixelles*, northern *Uccle*, and parts of *Forest*; under the full line, residual negatives persist north of *Uccle* and in parts of *Forest* but are smaller, while large positive effects concentrate north of the *Ville de Bruxelles* and across *Evere*. A structural downside of partial conversion might be the absence of a dedicated depot: rolling stock must be positioned from elsewhere each day, constraining staging and turnbacks and capping the number of trainsets in service.

Black tiles denote unreachable grid cells (six in Laeken Park, one within NATO grounds, one in the track interstice near Bruxelles-Midi) and have no analytical weight. The isolated *red tile near Lemonnier* should *not* be interpreted as a localised harm from the project: it is a data artefact of the 500 m centroid being unusually close to the current tram stops (Trams 4 and 10) and further from the planned metro entrance. Because the tram services are replaced by the metro in our scenarios, the nearest-stop access distance increases for that particular centroid.

4.4. What changes with a full build vs. a partial conversion

Three points emerge from the head-to-head comparison:

1. **Magnitude:** the full line roughly doubles the average benefit (mean -36 s vs. -17 s) and shifts the median below zero (Table 1); its shift function trough is more than twice as deep (Fig. 1b).
2. **Breadth:** more OD pairs meet common time budgets (Fig. 1c) and a larger share of origins experience improvements (Fig. 1d); spatially, benefits extend to the North-East (Fig. 2b).
3. **Stability:** the full line’s gains do not come at the cost of higher variability; if anything, RBI declines slightly (Fig. reffig:spesta-ramean, Table 3).

4.5. Key quantitative takeaways

- Network-wide averages are modest but systematic: -16.5 s (partial) and -36.1 s (full) mean OD savings; medians 0 and -6.1 s.

- Around half of OD pairs improve; 9.4 % (full) achieve ≥ 5 % savings (Table 1).
- The biggest gains occur in Forest and Saint-Gilles (both scenarios), extending to northern Schaerbeek and Evere under the full scenario (Figs. 2–3).
- Robustness checks across $t-10$, t , $t+10$ and three day types preserve all qualitative findings; risk-adjusted improvements track median improvements closely (Fig. 5).
- Apparent local slowdowns at the single red tile near Lemonnier are a *nearest-stop artefact* due to the grid centroid being nearer the existing tram stops than the future metro entrance; surrounding tiles follow the corridor-wide pattern of gains.

5. Implications and limitations

Time-savings ledger (lower bound, not a decision rule). Using the citywide mean gain of 16.5–36.1 s per trip and 400–500 million trips/year, even with VOT= €15/h the time-only benefit is just €60–€75 m/year. Capitalised at 4% over 40 years (CRF = 0.0505), this supports at most €1.5 bn before O&M and €0.3–€0.7 bn after plausible incremental O&M (€40–€60 m/year). Time savings alone therefore cannot carry a multi-billion project. We treat them as a lower bound and focus the appraisal on network option value, reliability and decarbonisation, and the distributional gains documented above.¹

Against multi-billion capital outlays, this suggests that short-term time savings alone are unlikely to justify Metro 3. Rather, the case must rest on wider system benefits such as network option value, reliability, decarbonisation, and the distributional gains we document. In this broader perspective, the project could also alleviate pressure on the *SNCB/NMBS* network by reducing reliance on the saturated North–South axis through Brussels, a persistent bottleneck in the national rail system [35]. At the same time, it creates the opportunity for a well-connected park-and-ride facility in the city’s northern periphery, an element currently missing from Brussels’ transport offer, which would further enhance multimodal integration and accessibility. Of course, additional network effects or alternative interventions may also be relevant, and a full appraisal should consider these alongside the impacts we highlight.

¹Formulae: $B = |\Delta t| Q \text{VOT}/3600$; $\text{CAPEX}_* = (B - \text{O\&M})/\text{CRF}$ with $\text{CRF} = r/(1 - (1 + r)^{-n}) = 0.0505$ for $r=0.04$, $n=40$. Full parameter ranges and sensitivity in App. Appendix A.

5.1. Policy interpretation: why a time-only CBA misses the point

Three considerations make a conventional short-horizon, time-only CBA uninformative in this case.

(i) *Network option value and path dependence.* Metro 3 is a *structural* intervention: it creates a high-capacity backbone that future surface restructurings and radial/diagonal links can hinge on. The option value of a grade-separated north–south trunk in Brussels (long identified as a binding network bottleneck) cannot be proxied by first-year time savings. It is an investment in *state space*: it expands what the network can become (and how it can be defended operationally) under growth, diversion, incident, and maintenance regimes.

(ii) *Externalities and non-marginal mode shift.* Our schedule-based, person-time results do not monetize *external* effects. At network scale, higher metro mode share plausibly reduces GHG/PM/NO_x, crash risk, and surface congestion, enables bus fleet repurposing, and supports station-area intensification; countervailing, place-based costs include construction-stage disruption, operational noise/vibration, platform/corridor crowding, risks of station-area social disorder (including visible homelessness), and local parking spillovers. Because the intervention induces *non-marginal* mode shift, externalities are non-linear: network benefits grow with long-run mode shares while certain disamenities scale with exposure at specific stations.

(iii) *Distribution and resilience as first-order policy goals.* The largest gains locate in *Forest* and *Saint-Gilles* for both scenarios, and extend into *northern Schaerbeek* and *Evere* for the full line (Figs. 2–3). These encompass densely populated, persistently disadvantaged neighbourhoods [36]. If the policy objective includes reducing *spatial inequality in access*, an equity-weighted welfare function would overweight these benefits and tilt appraisal in favour of the full build. Moreover, our robustness diagnostics show that improvements are not bought at the cost of instability; reliability metrics improve (Table 3), which matters disproportionately to lower-income travellers facing tighter temporal constraints.

5.2. Decision-relevant implications

- **Partial vs full.** The full line roughly doubles time benefits and extends them into the North-East, reaching additional vulnerable areas, without degrading reliability. If the decision shifts toward phasing, the analysis suggests that stopping at a partial conversion sacrifices a large share of distributionally valuable gains.

- **Transparency.** The absence of any official cost–benefit analysis underscores the need for open, reproducible evidence. Publishing “future GTFS” feeds and the scenario-building logic would enable independent replication and allow policy to evolve iteratively: tram and bus restructurings could be stress-tested transparently alongside the trunk alignment, rather than debated in the dark.

5.3. Limitations (and why they matter)

Coarse spatial sampling. We use a 500×500 m grid (639 effective centroids) as a tractability–coverage compromise. This smooths within-neighbourhood variation in station access/egress and can create artefacts where centroid–entrance geometry changes between scenarios (the Lemonnier dot). A denser lattice and/or network-constrained access modelling would sharpen micro-scale heterogeneity and reduce geometry-driven anomalies.

Scenario information scarcity. There is almost no authoritative, stable, public technical documentation at stop-by-stop resolution: station footprints, entrance locations, in-complex transfer links, dwell policies, depot/turnback constraints, and surface restructurings are pieced together from dated releases and press communications. Our GTFS scenario feeds are therefore *internally coherent counterfactuals*, not official plans. Exact implementation choices could move neighbourhood-scale results by dozens of seconds; this is a tractable uncertainty, but it requires the promoter to publish definitive operational data.

No demand model, no boarding loads. We report *generalised time* only. Without a demand model and APC/OD estimates, we cannot weight OD benefits by realised passenger flows; this is precisely the step that would allow for a more credible welfare estimate. Similarly, without capacity and crowding we cannot capture queueing delays, denied boardings, or the relief on overloaded surface lines, effects that are often decisive for welfare. *In this setting we implicitly treat all OD pairs and paths as equally important.* This uniform weighting ignores heterogeneity induced by workplace concentration, rail interchanges, and population density. Introducing OD weights (from APC/smartcard OD matrices or defensible gravity priors) would reweight accessibility toward the north–south spine—Belgium’s busiest axis. So, all else equal, the non-baseline scenarios would receive larger welfare gains relative if this was incorporated, in this sense our estimation can be viewed as a lower bound. This is an expected direction-of-effect; the magnitude depends on the adopted weights and time-of-day mix.

Temporal sampling and reliability. Our robustness is intentionally minimalistic (three instants in three bands), inspired by ergodicity arguments: avoid single-snapshot bias. But we still treat schedules deterministically and do not stochastically model disruptions, blockages, or headway variance propagation. A full reliability appraisal would embed a stochastic service-control model and compute risk-sensitive accessibility over a richer temporal ensemble.

Land-use feedbacks and dynamic network effects. The appraisal horizon that matters for a metro is *decades*, not the first operating year. Station area densification, relocation of activity patterns, and induced demand typically magnify accessibility impacts over time and change the spatial distribution of benefits. Conversely, governance and cost risks can delay or truncate phases; a rigorous option-value treatment would place probabilities over these paths and price irreversibility.

Computational tractability. The analysis required on the order of two CPU-years of computation. This shows the workflow is operating at the edge of the feasible computational frontier. Alternative modelling choices, finer lattices, network-constrained access, capacity and demand integration, would sharply increase runtime. We are therefore aware that other choices are possible and in some respects desirable, but their computational burden makes them currently impractical at metropolitan scale.

5.4. What to conclude and what not to

On a narrow, short-term *time-savings* ledger, Metro 3 is *probably not worth it*. The back-of-the-envelope monetisation above puts the full-line benefit at $\approx \text{€}75.2 \text{ m/year}$; with a 40-year horizon and 4% discount rate this justifies $\sim \text{€}1.5 \text{ bn}$ of capital *before* O&M and only $\sim \text{€}0.3\text{--}0.7 \text{ bn}$ once realistic O&M is included. In other words, time alone cannot carry a multi-billion project.

But that is not the right decision criterion. The question is whether Brussels needs a resilient, grade-separated north–south backbone to (i) enable a coherent surface network redesign, (ii) de-risk climate and mode-share goals, and (iii) reduce structural accessibility deficits in disadvantaged neighbourhoods. Our results speak directly to (iii) largest gains in Forest, Saint-Gilles, and, under the full line, into northern Schaerbeek and Evere, and they show these gains are *robust* across departure-time perturbations and do not degrade reliability. Together, these are the kinds of *distributional* and *resilience* benefits a narrow CBA misses but policy cares about.

Position in the literature. Substantively, we provide the first transparent, reproducible accessibility review of Metro 3. Methodologically, we contribute a generalisable pipeline for GTFS-based *counterfactual* network evaluation under project uncer-

tainty, with an explicit temporal-robustness design inspired by ergodicity economics (multi-instant, multi-day sampling) and risk-adjusted diagnostics. This combination is the paper’s core value: it produces decision relevant evidence where a full welfare CBA is presently infeasible, while advancing good practice for ex-ante accessibility analysis in transport geography.

Acknowledgments

The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation – Flanders (FWO) and the Flemish Government.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used GPT-5 in order to assist with the drafting and structuring of the manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- [1] L. Walker, “Much more costly than expected”: Stopping Metro 3 shouldn’t be taboo, minister says, The Brussels Times
Accessed 2025-09-17 (2024).
URL <https://www.brusselstimes.com/999853/much-more-costly-than-expected-stopping-metro-3-shouldnt-be-taboo-minister-says>
- [2] C. Richards, End of the line for Brussels Metro Line 3?, The Brussels Times
Accessed 2025-09-17 (2025).
URL <https://www.brusselstimes.com/1685904/end-of-the-line-for-brussels-metro-line-3-project>
- [3] RTBF, Travaux du métro au Palais du Midi : il ne fallait pas d’étude d’incidence avant l’ordonnance jugée discriminatoire, RTBF
Accessed 2025-09-17 (2025).
URL <https://www.rtbef.be/article/metro-nord-sud-au-palais-du-midi-il-ne-fallait-pas-d-etude-d-incidence-avant-l-ordonnance-jugee-discriminatoire-11490678>

- [4] BRUZZ, De toekomst van metro 3: MIVB legt verschillende scenario's op tafel, BRUZZ
Accessed 2025-09-17 (2025).
URL <https://www.bruzz.be/actua/mobiliteit/de-toekomst-van-metro-3-mivb-legt-verschillende-scenarios-op-tafel-2025-07-09>
- [5] A. Ferreira, Originally scheduled for 2031 and after 546 million invested, line 3 of the Brussels metro may not see the light of day after all!, Bruxelles Secrète
Accessed 2025-09-17 (2025).
URL <https://bruxellessecrete.com/en/line-3-metro-brussels-works/>
- [6] VRT NWS, Doorgaan, stations schrappen of streep erdoor? Toekomst Metro 3, VRT NWS
Accessed 2025-09-17 (2024).
URL <https://www.vrt.be/vrtnws/nl/2023/05/24/doorgaan-stations-schrappen-of-streep-erdoor-toekomst-metro-3>
- [7] K. T. Geurs, B. van Wee, Accessibility evaluation of land-use and transport strategies: review and research directions, *Journal of Transport Geography* 12 (2) (2004) 127–140. doi:10.1016/j.jtrangeo.2003.10.005.
- [8] G. Vandenbulcke, T. Steenberghen, I. Thomas, Mapping accessibility in Belgium: a tool for land-use and transport policy?, *Journal of Transport Geography* 17 (1) (2009) 39–53. doi:10.1016/j.jtrangeo.2008.04.008.
- [9] K. Lebrun, Travel time by public transport in Brussels: the accessibility of poles of activity, *Brussels Studies* 2018/2 (2018) 1–20. doi:10.4000/brussels.1661.
- [10] K. Martens, *Transport Justice: Designing Fair Transportation Systems*, Routledge, 2017. doi:10.4324/9781315746852.
- [11] R. H. M. Pereira, D. Banister, T. Schwanen, N. Wessel, Distributional effects of transport policies on inequalities in access to opportunities in rio de janeiro, *Journal of Transport and Land Use* 12 (1) (2019) 741–764. doi:10.5198/jtlu.2019.1523.
- [12] M. W. Conway, A. Byrd, M. Van Eggermond, Accounting for uncertainty and variation in accessibility metrics for public transport sketch planning, *Journal of Transport and Land Use* 11 (1) (2018) 541–558. doi:10.5198/jtlu.2018.1074.

- [13] R. H. Pereira, Future accessibility impacts of transport policy scenarios: Equity and sensitivity to travel time thresholds for bus rapid transit expansion in rio de janeiro, *Journal of Transport Geography* 74 (2019) 321 – 332. doi:<https://doi.org/10.1016/j.jtrangeo.2018.12.005>.
- [14] O. Peters, The ergodicity problem in economics, *Nature Physics* 15 (12) (2019) 1216–1221. doi:[10.1038/s41567-019-0732-0](https://doi.org/10.1038/s41567-019-0732-0).
- [15] A. Vanhoyweghen, V. Ginis, C. Macharis, Beyond expected travel time: Unveiling the role of Ergodicity breaking in mobility decisions and effective transportation policy, *Transportation Research Interdisciplinary Perspectives* In press; accepted 25 Sep 2025. Preprint available at SSRN. (2025). URL <https://ssrn.com/abstract=4949275>
- [16] G. Painter, M. Boarnet, M. Swayne, M. Miller, Innovation on job accessibility with general transit feed specification (gtfs) data, Tech. Rep. METTRANS Project 16-08, METTRANS Transportation Center, Los Angeles, CA (Dec. 2018). URL <https://rosap.ntl.bts.gov/view/dot/66861>
- [17] M. Stepniak, J. P. Pritchard, K. T. Geurs, S. Goliszek, The impact of temporal resolution on public transport accessibility measurement: Review and case study in poland, *Journal of Transport Geography* 75 (2019) 52–62. doi:[10.1016/j.jtrangeo.2019.01.007](https://doi.org/10.1016/j.jtrangeo.2019.01.007).
- [18] G. Boisjoly, A. El-Geneidy, Daily fluctuations in transit and job availability: A comparative assessment of time-sensitive accessibility measures, *Journal of Transport Geography* 52 (2016) 73–81. doi:[10.1016/j.jtrangeo.2016.03.004](https://doi.org/10.1016/j.jtrangeo.2016.03.004).
- [19] Société des Transports Intercommunaux de Bruxelles (STIB/MIVB), General transit feed specification (gtfs), accessed: 2025-06-08 (2025). URL <https://stibmivb.opendatasoft.com/explore/dataset/gtfs-files-production>
- [20] De Lijn, General transit feed specification (gtfs), accessed: 2025-06-08 (2025). URL <https://data.delijn.be/products>
- [21] Transport en Commun (TEC), General transit feed specification (gtfs), accessed: 2025-06-08 (2025). URL <https://opendata.tec-wl.be/Current%20GTFS/>

- [22] SNCB/NMBS, General transit feed specification (gtfs), accessed: 2025-06-08 (2025).
URL <https://gtfs.irail.be/>
- [23] OpenStreetMap contributors, Openstreetmap, <https://www.openstreetmap.org>, open data under ODbL (2025).
- [24] Geofabrik GmbH, Belgium — belgium-latest.osm.pbf, <https://download.geofabrik.de/europe/belgium.html>, oSM data up to 2025-09-19T21:39:44Z; accessed 2025-09-21 (2025).
- [25] C. Fink, W. Klumpenhouwer, M. Saraiva, R. Pereira, H. Tenkanen, r5py: Rapid realistic routing with r5 in python, Zenodo10 5281 (2022).
- [26] M. W. Conway, A. Byrd, M. van der Linden, Evidence-based transit and land use sketch planning using interactive accessibility methods on combined schedule and headway-based networks, *Transportation Research Record* 2653 (1) (2017) 45–53. doi:10.3141/2653-06.
- [27] M. W. Conway, A. F. Stewart, Getting charlie off the mta: a multiobjective optimization method to account for cost constraints in public transit accessibility metrics, *International Journal of Geographical Information Science* 33 (9) (2019) 1759–1787. doi:10.1080/13658816.2019.1605075.
- [28] Metro 3 project official portal, Metrolijn 3 in brussel, accessed: 2025-09-17 (2025).
URL <https://metro3.be/>
- [29] R. R. Wilcox, D. M. Erceg-Hurn, Comparing two dependent groups via quantiles, *Journal of Applied Statistics* 39 (12) (2012) 2655–2664. doi:10.1080/02664763.2012.724665.
- [30] M. S. Handcock, M. Morris, *Relative Distribution Methods in the Social Sciences*, *Statistics for Social and Behavioral Sciences*, Springer, New York, 1999. doi:10.1007/b97852.
- [31] J. Frank J. Massey, The Kolmogorov–Smirnov Test for Goodness of Fit, *Journal of the American Statistical Association* 46 (253) (1951) 68–78. doi:10.1080/01621459.1951.10500769.

- [32] C. Chen, A. Skabardonis, P. Varaiya, Systematic identification of freeway bottlenecks, *Transportation Research Record* 1867 (1) (2004) 46–52. doi:10.3141/1867-06.
- [33] K. Lyman, R. L. Bertini, Using travel time reliability measures to improve regional transportation planning and operations, *Transportation Research Record* 2046 (1) (2008) 1–10. doi:10.3141/2046-01.
- [34] H. Föllmer, A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, 4th Edition, De Gruyter, Berlin, Boston, 2016. doi:10.1515/9783110463453.
- [35] G. te Boveldt, K. Van Raemdonck, C. Macharis, A new railway tunnel under brussels? assessing political feasibility and desirability with competence-based multi criteria analysis, *Transport Policy* 66 (2018) 30–39.
- [36] Institut Bruxellois de Statistique et d’Analyse (IBSA), Mini-bru 2024: La région de bruxelles-capitale en chiffres, Tech. rep., Institut Bruxellois de Statistique et d’Analyse (IBSA), Bruxelles, accessed: 2025-09-26 (2024).
URL https://ibsa.brussels/sites/default/files/publication/documents/MiniBru_24-FR.pdf

Appendix A. Back-of-the-envelope monetisation: formulae and sensitivity

Appendix A.1. Set-up and identities

We monetise time savings as

$$B = |\overline{\Delta t}| \times Q_{\text{trips/yr}} \times \frac{\text{VOT}}{3600} \quad (\text{€ per year}). \quad (\text{A.1})$$

Annual benefits B capitalise to breakeven CAPEX via the capital-recovery factor:

$$\text{CRF}(r, n) = \frac{r}{1 - (1+r)^{-n}}, \quad \text{CAPEX}_\star = \frac{B - \text{O\&M}}{\text{CRF}}. \quad (\text{A.2})$$

Everything is linear in $|\overline{\Delta t}|$, Q , and VOT; O&M shifts CAPEX one-for-one divided by the CRF.

We use the paper’s measured mean time gains (partial: 16.5 s; full: 36.1 s), representative trip volumes $Q \in \{400, 500\} \times 10^6/\text{yr}$, and $\text{VOT} \in \{12, 15\} \text{ €/h}$.

Scenario	$ \overline{\Delta t} $ (s)	Q (m trips/yr)	VOT (€/h)	B (m€/yr)
Partial	16.5	400	12	22.0
Partial	16.5	400	15	27.5
Partial	16.5	500	12	27.5
Partial	16.5	500	15	34.4
Full	36.1	400	12	48.1
Full	36.1	400	15	60.2
Full	36.1	500	12	60.2
Full	36.1	500	15	75.2

Table A.4: Annual time-only benefits B under the combinations used or cited in the text.

	$n=30$	$n=40$	$n=50$
$r=3\%$	0.051 019	0.043 262	0.038 865
$r=4\%$	0.057 830	0.050 523	0.046 550
$r=5\%$	0.065 051	0.058 278	0.054 777

Table A.5: Capital-recovery factor $\text{CRF}(r, n)$. Baseline in text: $r=4\%$, $n=40 \Rightarrow \text{CRF} = 0.050523$.

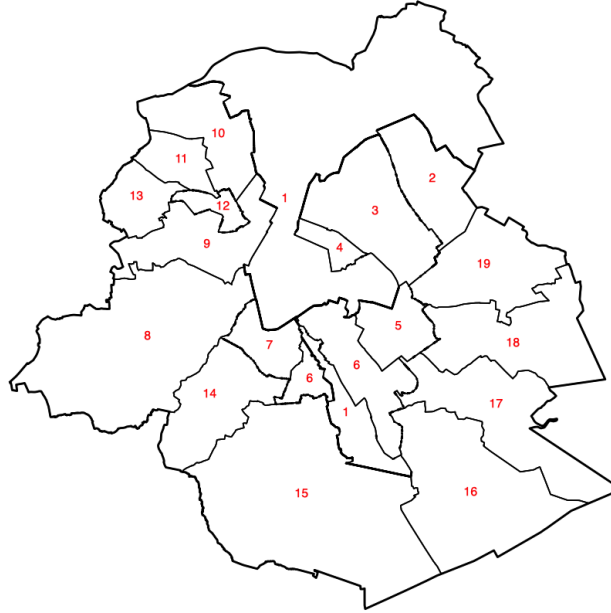
Appendix A.2. From annual benefits to breakeven CAPEX

Illustrative mapping from time-only annual benefit B to $\text{CAPEX}_\star$ for three discount rates (40-year horizon) and O&M bands. Units: billion euros.

CRF case	B (m€/yr)	O&M=0	O&M=40	O&M=60
$r=3\%$, $n=40$	60.2	1.392	0.467	0.005
$r=3\%$, $n=40$	75.2	1.738	0.814	0.351
$r=4\%$, $n=40$	60.2	1.192	0.400	0.004
$r=4\%$, $n=40$	75.2	1.488	0.697	0.301
$r=5\%$, $n=40$	60.2	1.033	0.347	0.003
$r=5\%$, $n=40$	75.2	1.290	0.604	0.261

Table A.6: Breakeven $\text{CAPEX}_\star = (B - \text{O\&M})/\text{CRF}$. The baseline row $r=4\%$, $n=40$, $B=75.2$ reproduces $\sim\text{€}1.49$ bn before O&M and $\sim\text{€}0.70\text{--}0.30$ bn after $\text{€}40\text{--}60$ m/yr O&M.

Appendix B. Brussels communes map



- | | |
|--------------------------|---------------------------|
| 1. Ville de Bruxelles | 11. Ganshoren |
| 2. Evere | 12. Koekelberg |
| 3. Schaerbeek | 13. Berchem-Sainte-Agathe |
| 4. Saint-Josse-ten-Noode | 14. Forest |
| 5. Etterbeek | 15. Uccle |
| 6. Ixelles | 16. Watermael-Boitsfort |
| 7. Saint-Gilles | 17. Auderghem |
| 8. Anderlecht | 18. Woluwe-Saint-Pierre |
| 9. Molenbeek-Saint-Jean | 19. Woluwe-Saint-Lambert |
| 10. Jette | |

Figure B.6: Brussels-Capital Region communes. Numbers correspond to labels on the map.