Operational Micromobility Rebalancing through Crowdsourced Trucks

Linda Belkessa*[1,2], Mostafa Ameli[1,3], Mohsen Ramezani[2], and Mahdi Zargayouna[1]

[1]COSYS-GRETTIA, Univ. Gustave Eiffel, Champs-sur-Marne, France
[2]School of Civil Engineering, The University of Sydney, Sydney, Australia
[3]University of California, Berkeley, USA

Short summary

This paper proposes a city-scale framework for station-based micromobility rebalancing that combines censored-demand reconstruction with hybrid truck coordination. While prior studies have examined operator fleets, uncertain demand, and some crowdsourced relocation, city-scale frameworks that jointly combine censored-demand reconstruction, real delivery-truck trajectories, stochastic participation, and rolling MPC remain limited. To address this gap, the paper introduces **Shortage-Aware Tobit Attention (SATA)** to recover latent demand when low bike inventory suppresses observed departures. The reconstructed demand is converted into shortage and surplus candidates and optimized through a **hybrid service-MPC** that coordinates deterministic operator trucks with stochastic crowdsourced delivery vehicles. Relative to the operator-only baseline, the full stack achieves its strongest service-rate gains in Chicago first-response stress scenarios and Lyon crowd-advantage settings, while profit gains are concentrated in Lyon.
Keywords: micromobility rebalancing; demand forecasting; attention model; censored demand; crowdsourced truck; hybrid fleet optimization

1 Introduction

Station-based shared micromobility systems (shared-μM), especially citywide bikeshare networks, routinely face short-horizon inventory imbalances. Directional commuting, asymmetric trip generation, and time-varying demand leave some stations empty and others saturated, causing lost departures, failed returns, and reduced service availability (Schuijbroek et al., 2017). Rebalancing is the standard response, but the challenge is not only to determine how many vehicles to move. Operators must also decide which shortages require immediate action, which stations can release inventory without creating downstream risk, and which logistics resources can respond within the next decision window. These decisions become harder when the dedicated fleet is small and predicted departures understate true demand during stockouts.

This paper studies city-scale, short-horizon hybrid rebalancing. Dedicated operator vehicles remain the primary resource, while third-party delivery trucks are treated as exogenous carriers for relocation tasks. The idea is that freight vehicles already moving through the city may offer residual carrying capacity on trips that would occur anyway. Prior work on crowd-shipping and hybrid delivery shows that pre-existing trips with spare capacity can be coordinated with professional fleets through explicit matching and routing decisions (Macrina et al., 2020). This suggests that urban delivery flows may complement shared-μM rebalancing.

This motivation is illustrated in Figure 1, which summarizes the Lyon setting from October 2023 to October 2024. On the map, persistent shortage-surplus patterns concentrate along a limited set of urban corridors, several of which coincide with recurrent freight activity and with the most truck-feasible rebalance links. The figure shows that 95% of annual shortage volume lies within 300m of recurrent truck corridors. Furthermore, 71% of inferred rebalance flow has at least 25% corridor overlap. However, hybrid rebalancing is not simply a larger fleet; it is a constrained multimode coordination problem. Shortage-surplus patterns evolve over space and time, feasibility depends on short-horizon forecasts, and crowd resources are only partially controllable (Dell'Amico et al., 2018; Çınar et al., 2024). Unlike dedicated vehicles, third-party truck movements are largely exogenous: their routes, schedules, and participation cannot be fully dictated by the operator. In addition, each candidate transfer must be feasible on the street network under compatibility, timing, capacity, and detour constraints (Dahle et al., 2019; Macrina et al., 2020).

Accordingly, the gap is not whether operator rebalancing, uncertainty-aware planning, or shared logistics can each help in isolation. The unresolved question is whether a city-scale framework can jointly reconstruct censored demand, exploit real delivery-truck trajectories, account for stochastic crowd participation, and re-optimize in rolling-horizon MPC form. The unresolved question is whether a city-scale operational framework can jointly reconstruct censored demand, exploit real delivery-truck trajectories, account for stochastic crowd participation, and re-optimize decisions in rolling-horizon MPC form for station-based micromobility systems.

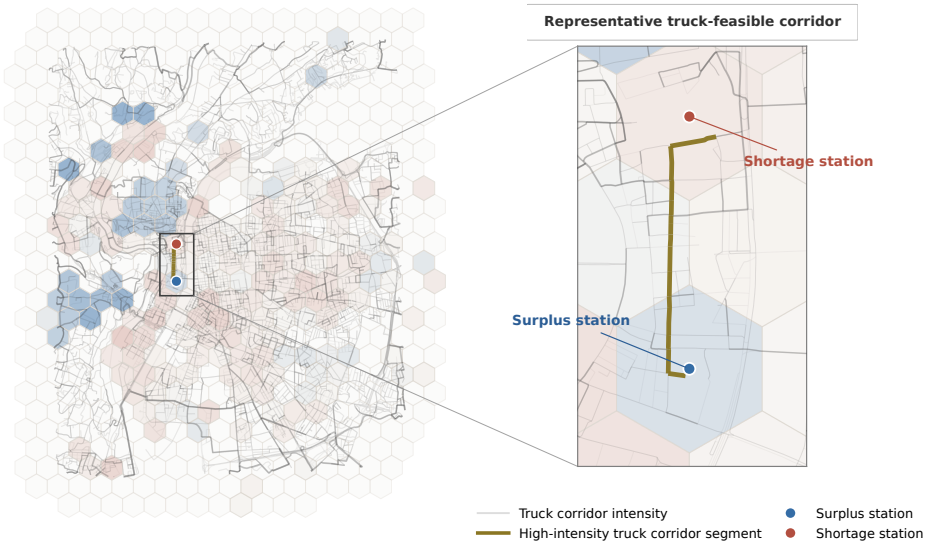


Figure 1: Overlap of major shortage-surplus hubs and delivery truck routes in Lyon over October 2023 and October 2024. Persistent bikeshare imbalance clusters along recurring urban corridors that substantially overlap with recurrent delivery truck activity.

Literature review

The literature relevant to this study lies at the intersection of three streams. The first covers operator-based rebalancing with dedicated operator-controlled vehicles under observed or forecasted imbalances (Dell'Amico et al., 2018). It includes vehicle-oriented formulations with richer routing, assignment, or predictive control logic, such as Jin et al. (2023), but these studies typically assume operator-managed fleets and do not exploit exogenous urban freight trajectories as partially controllable relocation resources. The second covers forecast-informed rebalancing, showing that short-horizon anticipation improves repositioning and that censored departures can bias shortage estimation when low inventory suppresses observed demand (Dell'Amico et al., 2018; Liu et al., 2024). The third concerns shared logistics and crowd-shipping, where residual capacity on pre-existing trips is matched with delivery tasks under detour, timing, and compensation considerations (Cebeci et al., 2024).
Despite this literature, an important gap remains. Existing studies have separately examined operator fleets, uncertain demand, and, in some cases, crowdsourced relocation capacity. Evidence remains limited on city-scale operational frameworks that jointly combine censored-demand reconstruction, real delivery-truck trajectories, stochastic crowd participation, and rolling MPC for station-based micromobility rebalancing.

Objective and contributions

This paper evaluates whether exogenous delivery tours can absorb part of rebalancing under real constraints. We ask whether forecasted shortages and surpluses align with truck routes, what share of rebalancing can be transferred under operational constraints, and how much value censor-aware demand correction adds relative to an operator-only baseline.
This study makes three contributions:

1. we propose **Shortage-Aware Tobit Attention (SATA)**, which combines shortage-gated

attention penalties with an input-dependent Tobit reconstruction layer to correct censored departures.

2. we formulate a **hybrid service-MPC** that jointly optimizes deterministic operator trajectories and stochastic, route-constrained crowd opportunities.
3. we provide a city-scale empirical evaluation that investigates how the integration of crowd-sourced capacity impacts **service availability and user-access metrics**

2 Methodology

We study the problem on a 15-minute station-time grid. For station i at interval t, the operational state includes observed departures y_{it}, available bikes b_{it}, station capacity C_i, and free docks $C_i - b_{it}$. Stockout pressure is quantified through three indicators:

$$m_{it} = \mathbb{1}[b_{it} \le 1], \qquad s_{it} = \mathbb{1}[b_{it} = 0], \qquad \pi_{it} = \mathrm{clip}\big(\tfrac{2-b_{it}}{2},\, 0,\, 1\big)\,. \tag{1}$$

Because departures are mechanically censored when a station is nearly empty, y_{it} is a lower bound on true demand. Throughout the paper, b_{it} denotes available bikes at station i in interval t, and b_i denotes the current inventory when the time index is omitted. The parameter $\underline{b}_i$ is the minimum reserve level at station i, so q_i is the relocatable inventory above that floor, while n_j is the shortage quantity at destination station j. Likewise, $\hat{a}_{it}$ denotes predicted arrivals to station i during interval t, and $\hat{a}_j$ the imminent-arrival term when the time index is omitted.

Representation-level uncensoring with external mobility signals

Existing uncensoring methods act either before training or only at the objective stage, and both are limited. Input-level imputation misses high-frequency station dynamics, while objective-level corrections operate on representations already trained on censored observations y_{it}. As a result, the latent space reflects stockout-induced patterns rather than true demand. Figure 2 illustrates the core intuition of the proposed attention mechanism: censored observations should be discounted in score space and corrected in value space.

We propose **Shortage-Aware Tobit Attention (SATA)**, which embeds uncensoring directly into self-attention. SATA starts from an input-dependent **shortage posterior** $\rho_{it} \in (0, 1)$ computed from inventory, recent stockouts, and contextual covariates. When ρ_{it} is high, observed departures are treated as unreliable. SATA then penalizes shared censorship in the attention scores, reducing spurious relationships between stations that are empty at the same time. We therefore penalize the attention score $\tilde{A}_{ij}$ whenever either the query station i or the key station j shows evidence of censorship:

$$\tilde{A}_{ij} = \frac{Q_i K_j^\top}{\sqrt{d}} - \lambda\, \frac{\rho_i + \rho_j}{2}. \tag{2}$$

The second modification is Tobit-based value reconstruction. When a station is empty, the observed value V_j is a lower bound, so SATA blends it with the conditional mean of a latent Gaussian distribution truncated at the current inventory level:

$$\tilde{V}_j = (1 - \rho_j)\, V_j + \rho_j\, \hat{V}_j^T, \tag{3}$$

where the reconstruction $\hat{V}_j^T$ is derived from the inverse-Mills ratio. This ratio is a statistical correction term representing the expected value of the error in a truncated distribution:

$$\hat{V}_j^T = \mu_j + \sigma_j\, \frac{\phi(\alpha_j)}{1 - \Phi(\alpha_j)}. \tag{4}$$

In implementation, μ_{it} is anchored to the station-time median forecast, while σ_{it}^{base} comes from forecast spread or residual dispersion; the effective scale is widened under stronger shortage evidence and positive external mobility signals.

To resolve the ambiguity of zero-departure intervals, we incorporate lagged taxi activity as an **external mobility covariate**. The standardized signal z_{it}^{taxi} aggregates taxi pickups at the finest geography available in each dataset and serves as an external witness to latent mobility demand. The intuition is that while bike departures y_{it} are physically bounded by inventory b_{it}, taxi pickups are not; therefore, a surge in taxi activity during a bike stockout provides high-confidence evidence

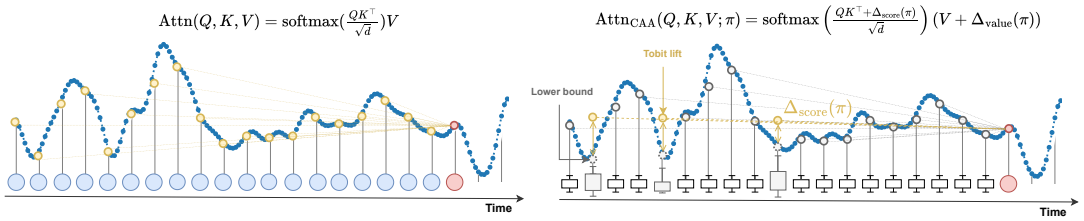


Figure 2: Conceptual comparison between vanilla self-attention and the proposed censorship-aware attention mechanism for demand reconstruction.

of suppressed mobility demand. We formalize this by using z_{it}^{taxi} to inflate the **effective posterior scale** σ_{it}^{eff}:

$$\sigma_{it}^{\text{eff}} = \sigma_{it}^{\text{base}}\big(1 + \gamma_s \rho_{it}\big)\big(1 + \gamma_\tau [z_{it}^{taxi}]_+\big). \tag{5}$$

By widening the distribution's variance during periods of high external mobility and high shortage probability (ρ_{it}), the model acknowledges greater uncertainty in the observed zero. In the present study, taxi activity is used as an offline or lagged contextual covariate to test the value of exogenous mobility information; a live deployment could instead use any streaming urban-mobility proxy available with operational latency. The final uncensored demand $\hat{d}_{it}^{unc}$ is then recovered as the conditional expectation of a lower-truncated Tobit distribution:

$$\hat{d}_{it}^{unc} = \mathbb{E}\big[D_{it} \mid \mu_{it},\, \sigma_{it}^{\text{eff}},\, D_{it} \geq y_{it}\big]\,. \tag{6}$$

This corrected demand signal fundamentally recalibrates the **shortage-to-relocation mapping**. We decompose station states into surplus and shortage candidates based on corrected needs rather than historical observations. A station becomes a **surplus candidate** q_i only if its inventory exceeds a safety reserve $\underline{b}_i$. More critically, a station is identified as a **shortage candidate** n_j if the uncensored demand exceeds the sum of current inventory and imminent arrivals $\hat{a}_j$:

$$n_j = \max\Big(\hat{d}_j^{unc} - \hat{a}_j - b_j,\; 0\Big)\,, \tag{7}$$

where b_j denotes the current bike inventory at shortage candidate station j and $\hat{a}_j$ denotes imminent predicted arrivals to that station over the current control horizon. Grounding n_j in $\hat{d}_j^{unc}$ exposes the optimizer to latent shortage risk and avoids neglecting persistently depleted stations. These candidates define two relocation classes. **Operator tasks** $a \in \mathcal{A}^{op}$ are deterministic moves by dedicated trucks, whereas **crowd opportunities** $o \in \mathcal{A}^{cr}$ are exogenous transfers extracted from active delivery tours when detour, timing, and capacity constraints are satisfied. Candidate opportunities are pre-filtered before entering the optimization model.

Hybrid service-MPC and stochastic participation

Figure 3 summarizes the full methodological pipeline. The framework coordinates rebalancing through a service-oriented Mixed-Integer Program (MIP). The controller is re-optimized every 15 minutes using the latest inventory state, forecast inputs obtained from Belkessa et al. (2026), and currently feasible crowd opportunities. Unlike traditional cost-minimization models, our service-oriented formulation prioritizes protection against stockouts during peak volatility. The objective function jointly optimizes the deployment of deterministic operator tasks (x_a, z_a) and stochastic crowd opportunities (w_o, r_o):

$$\begin{aligned} \max \quad & \sum_{a \in \mathcal{A}^{op}} \Big[\big(\alpha G_a + \beta F_a + \gamma M_a - c_a^{bike} - c_a^{travel}\big)x_a - \lambda_{op}\, c_a^{fix}\, z_a\Big] \\ & + \sum_{o \in \mathcal{A}^{cr}} \Big[\big(\alpha G_o + \beta F_o + \gamma M_o - \lambda_{cr} c_o^{comp} - c_o^{det} + \eta\, \hat{p}_o\big)w_o - \lambda_{pl}\, c_o^{plat}\, r_o\Big]. \end{aligned} \tag{8}$$

Equation 8 balances three dimensions of value. First, G captures immediate shortage relief. Second, F captures forecast-positioning value by moving bikes to stations predicted by SATA to face near-term surges. Third, M captures access quality through walk-minutes saved. By weighting M with destination walking time m_j^{walk}, the optimizer reduces user walking time and prioritizes isolated stations with poor fallback access. Unless otherwise stated, the reserve floor $\underline{b}_i$ and weights $(\alpha, \beta, \gamma, \eta)$ are fixed across experiments.

A critical challenge in this hybrid setting is the inherent uncertainty of the crowd. We model crowd participation as stochastic rather than guaranteed. The planner assigns each feasible crowd offer o an expected acceptance probability $\hat{p}_o$ using a remuneration-detour logistic model calibrated from the public stated-preference crowdshipping dataset underlying (Cebeci et al., 2024). For detour time d_o (minutes) and remuneration r_o (euros), we use

$$\hat{p}_o = \frac{1}{1 + \exp[-(0.070 - 0.052\, d_o + 0.030\, r_o)]}. \tag{9}$$

For the no-remuneration baseline, this reduces to

$$\hat{p}_o = \frac{1}{1 + \exp[-(0.313 - 0.044\, d_o)]}. \tag{10}$$

These coefficients preserve the monotonic effects of detour and compensation while adapting baseline willingness to the short-detour range observed in Lyon and Chicago. This creates a deliberate information asymmetry: the MIP only observes expected participation, whereas the simulator samples realized participation after the four-stage process of visibility, response delay, acceptance/rejection, and cancellation.

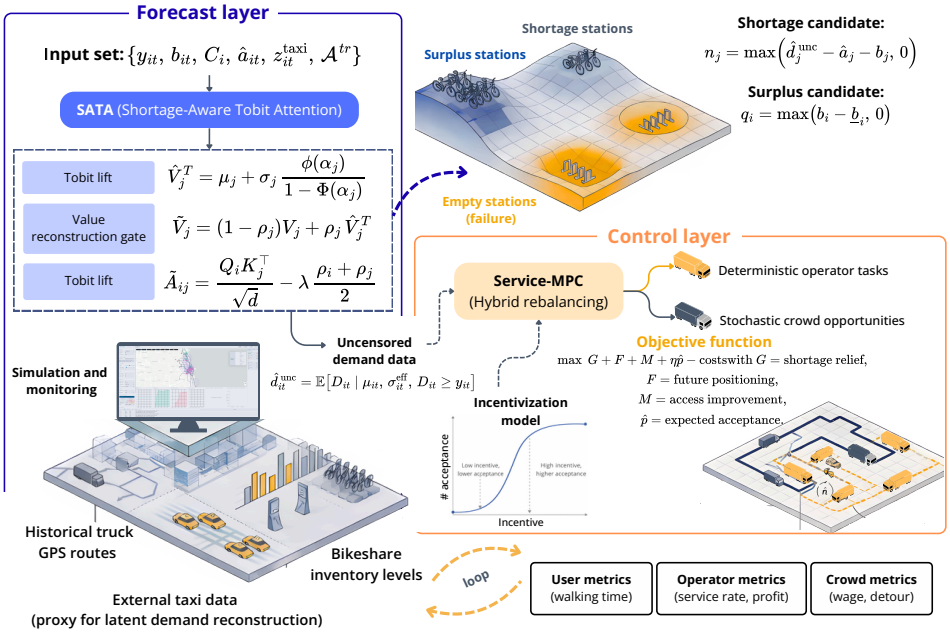


Figure 3: Overview of the proposed hybrid rebalancing framework. SATA reconstructs censored short-horizon demand from multi-source operational data. The resulting shortage and surplus candidates are passed to a hybrid service-MPC that coordinates deterministic operator trucks and stochastic crowdsourced opportunities under feasibility and incentivization constraints. A rolling-horizon simulation loop updates inventories and performance metrics for replanning.

Operational constraints and physical feasibility

The optimization is governed by strict physical flow conservation. The coupling of operator and crowd flows ensures that total extractions and deliveries never exceed the surplus q_i or the corrected need n_j identified by the SATA module:

$$\sum_{a:\text{src}(a)=i} x_a + \sum_{o:\text{src}(o)=i} w_o \leq q_i, \qquad \sum_{a:\text{dst}(a)=j} x_a + \sum_{o:\text{dst}(o)=j} w_o \leq n_j. \tag{11}$$

Station inventory evolves by standard stock-flow balance with capacity truncation:

$$b_{i,t+1} = \min\Big\{C_i, \max[0,\ b_{it} - y_{it}^{serv} + \hat{a}_{it} + \sum_{a:\text{dst}(a)=i} x_a + \sum_{o:\text{dst}(o)=i} \tilde{w}_o - \sum_{a:\text{src}(a)=i} x_a - \sum_{o:\text{src}(o)=i} \tilde{w}_o]\Big\}, \tag{12}$$

where y_{it}^{serv} denotes served departures and $\tilde{w}_o$ denotes realized, rather than planned, crowd execution. This update accounts for served demand, natural arrivals $\hat{a}_{it}$, and the net effect of both deterministic operator moves and stochastic crowd execution. Crucially, the constraint structure preserves the fundamental asymmetry of the system: once the operator flow is dispatched (x_a), it is deterministic. However, crowd flow (w_o) remains subject to stochastic execution and route-feasibility restrictions.

Evaluation protocol

We evaluate these policies in a stock-flow simulator that enforces real-world constraints, such as inventory caps C_i and the refusal of arrivals at saturated stations. The primary comparison is between the full hybrid system and an operator-only baseline under identical information sets, forecast inputs, and simulator rules. This evaluation asks to what extent modest, uncertain crowd capacity can compensate for dedicated-fleet bottlenecks when guided by uncensored demand signals. Furthermore, this controlled comparison allows us to quantify the service gain attributable to SATA's demand recovery independently from the logistics gain provided by the crowd network.

3 Numerical results

Experiments are conducted on two real-world systems. **Chicago Divvy** comprises 658 stations evaluated over two 15-minute decision windows in 2024. Trip records are obtained from the City of Chicago open data portal.[1]. Taxi-pickup proximity signals are drawn from the Chicago Transportation Network Provider dataset.[2] Crowd delivery routes are sourced from the Amazon Last-Mile Routing Research Challenge(Merchán et al., 2024) providing detailed GPS sequences and volumetric capacity for each vehicle. **Lyon Vélo'v** comprises 352 stations evaluated over 2023. Station data are obtained from the Grand Lyon Métropole open API.[3] Crowd routes are drawn from a French last-mile parcel benchmark (Hörl & Puchinger, 2023) covering six carrier networks operating in the metropolitan area: La Poste (31 routes), UPS (16), Chronopost (8), DHL (8), FedEx (7), and Colis Privé (4).

SATA attention inspection

Figure 4 visually confirms the SATA mechanism's ability to reconstruct and prioritize latent demand signals. In the vanilla self-attention baseline (Panel A), attention weights are diffuse and lack clear temporal focus. Because the model processes raw departures y_{it} as ground truth, the censored periods (marked by blue dashed lines) appear as uninformative signals, providing no distinct anchors for the attention mechanism. This results in a smeared representation that fails to capture the true intensity of demand peaks. In contrast, Panel B shows that SATA suppresses attention on jointly censored keys while preserving shortage information through the reconstructed values $\tilde{V}_j$. This reveals the mechanism's two-part logic: censorship is discounted in the score space, while latent unmet demand is restored in the value space through the inverse-Mills-adjusted expectation $\hat{V}_j^T$. The attention mechanism successfully identifies these reconstructed points as the most informative predictors of future demand depth $\hat{d}_{it}^{unc}$.
While the methodology incorporates a penalty term $-\lambda\frac{\rho_i+\rho_j}{2}$ in the attention scores $\tilde{A}_{ij}$ to avoid over-trusting jointly censored tokens, SATA does not discard shortage information altogether. Instead, it discounts misleading co-censorship in the score space while preserving latent unmet demand through the reconstructed values $\tilde{V}_j$. Operationally, this means the system does not react to raw zeros as if they were informative demand observations; it reacts to the reconstructed shortage signal that remains after censoring has been discounted.

Hybrid rebalancing performance

Table 1 extends these insights to the full Lyon and Chicago networks and shows that gains are regime-dependent. Moving from baseline to SATA-enriched models improves the shortage-to-

[1] `https://divvy-tripdata.s3.amazonaws.com`

[2] `https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips-2023-/n26f-ihde`.

[3] `https://data.grandlyon.com/portail/fr/jeux-de-donnees/stations-velo-v-de-la-metropole-de-lyon/info`.

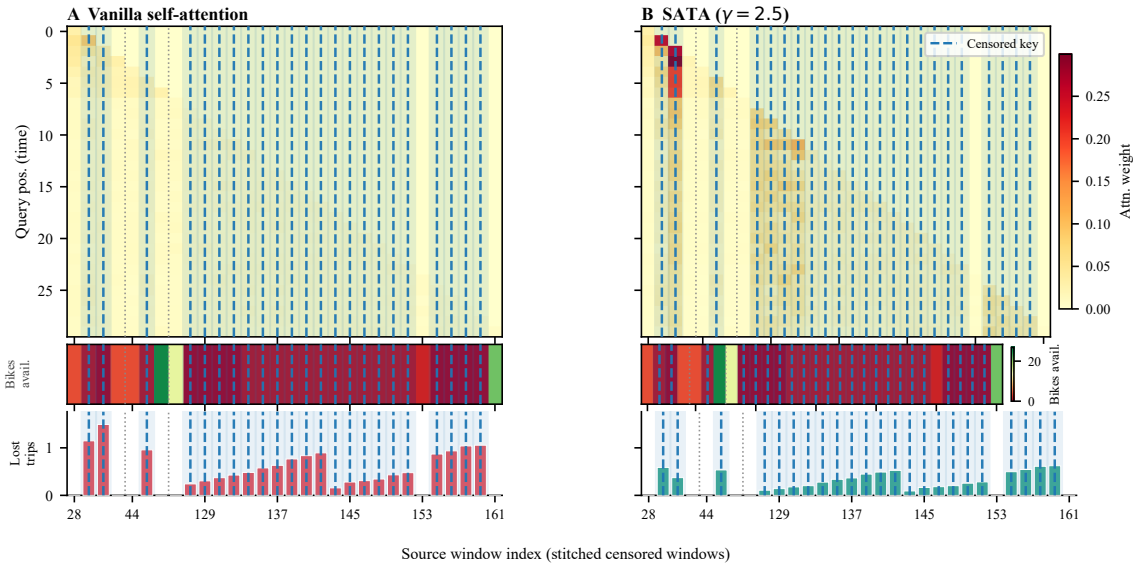


Figure 4: SATA attention inspection on a censored Chicago station sequence. **(A)** Vanilla self-attention assigns non-negligible mass across the full history, including censored key positions. **(B)** SATA ($\gamma = 2.5$) suppresses attention on censored keys and reallocates weight toward informative lookback windows. The lower strips report bikes availability and lost trips.

relocation mapping and fleet efficiency. In the Operator-Only rows, replacing raw departures y_{it} with corrected demand $\hat{d}_{it}^{unc}$ reduces lost trips by identifying shortage candidates n_j in high-traffic zones that appear stable under censored observations.

The largest service-rate gains do not arise uniformly. In **Chicago first-response stress**, service rate rises from 51.7% to 68.0%, the largest relative gain in the table. In **Lyon crowd-advantage proof**, it rises from 73.6% to 83.3%, the clearest case where crowd opportunities improve both service and profit. Gains are smaller in unbiased operating regimes, where the system is less capacity-constrained and fewer crowd opportunities are needed.

Figure 5 illustrates the same mechanism on a selected Chicago corridor. Panel (A) shows the transfer from *Clinton St & Madison St* to *Franklin St & Lake St*, while Panel (B) shows the compounding value of successive modeling enrichments. Under the raw and uncensored variants, both policies remain below 40% recovery, indicating that adding crowd capacity alone or improving the demand signal alone is insufficient. Under the enriched policy, however, the hybrid system reaches a 75% service rate, versus 27% for the operator-only baseline. This gap reflects the core mechanism of the framework: SATA identifies the true shortage candidate n_j, and the hybrid-MPC secures the required crowd capacity before the operator truck, constrained by travel time c_a^{travel}, can respond.

An important cross-city difference concerns profit. In Lyon, the crowd-sensitive regimes show positive profit gains because the accepted offers move more bikes and relieve more shortages per compensated task, so the operational value of the transfers exceeds the associated remuneration and platform costs. In Chicago, however, the crowd-advantage proof regime still reports negative profit changes even when service improves. This indicates that, in that setting, the available crowd offers are less economically favorable: they are either too sparse, too weakly aligned with the most valuable shortage corridors, or too costly relative to the service recovered. In other words, the hybrid layer remains operationally useful in Chicago, but its economic value is more conditional than in Lyon. These findings suggest that uncensoring and uncertainty modeling are functionally complementary. Uncensoring extracts the magnitude of the intervention required for stations in failure mode (shortage). Additionally, uncertainty-aware optimization ensures the dispatch is robust to the volatility of the forecasts.

Table 1: Benchmark of the rolling rebalancing MPC across Lyon and Chicago. Each city-regime block reports the operator-only baseline and hybrid variants using deterministic q50, bootstrap uncertainty, and bootstrap uncertainty with selective uncensoring. The full hybrid stack performs best in Lyon's crowd-sensitive regimes, while Chicago gains concentrate in first-response stress scenarios and uncensoring mainly improves service when crowd deployment is feasible.

City	Benchmark regime	Controller	Demand signal	#episodes	Service rate	Lost-trip Δ, 15m	Lost-trip Δ, interval	Profit Δ, 15m	Crowd bikes	Crowd offers	Mean fare	Mean accept.
Lyon	Crowd-advantage proof	Operator-only	All info states (same execution)	4	73.6%	4.12	12.30	56.74	0.00	0	—	—
		Hybrid MPC	Deterministic q50	4	78.2%	5.87	16.83	64.74	30.07	17	1.28x	68.1%
		Hybrid MPC	Bootstrap uncertainty	4	80.2%	7.96	19.42	68.11	31.94	18	1.35x	73.8%
		Hybrid MPC	Bootstrap + selective uncensoring	4	**83.3%**	**10.74**	**23.91**	**77.35**	38.46	19	1.42x	77.0%
	First-response stress	Operator-only	All info states (same execution)	2	70.8%	1.01	2.36	28.40	0.00	0	—	—
		Hybrid MPC	Deterministic q50	2	76.2%	1.29	4.75	31.87	13.02	7	1.28x	68.0%
		Hybrid MPC	Bootstrap uncertainty	2	77.2%	1.71	5.63	36.92	14.87	7	1.35x	74.6%
		Hybrid MPC	Bootstrap + selective uncensoring	2	**79.2%**	**2.28**	**7.14**	**43.58**	17.94	9	1.43x	79.3%
	Unbiased operating	Operator-only	All info states (same execution)	2	78.5%	0.55	0.55	10.97	0.00	0	—	—
		Hybrid MPC	Deterministic q50	2	78.8%	0.58	0.58	11.02	1.81	1	1.18x	58.4%
		Hybrid MPC	Bootstrap uncertainty	2	79.8%	1.12	1.31	13.26	2.74	2	1.24x	64.9%
		Hybrid MPC	Bootstrap + selective uncensoring	2	**81.8%**	**1.84**	**2.27**	**16.94**	4.36	2	1.31x	71.2%
Chicago	Crowd-advantage proof	Operator-only	All info states (same execution)	4	62.3%	0.14	0.01	-33.88	0.00	0	—	—
		Hybrid MPC	Deterministic q50	4	62.4%	0.14	0.01	-33.59	16.96	12	1.12x	53.4%
		Hybrid MPC	Bootstrap uncertainty	4	63.4%	0.41	0.47	-29.44	19.73	13	1.26x	64.7%
		Hybrid MPC	Bootstrap + selective uncensoring	4	**65.4%**	**0.93**	**1.52**	-22.18	24.56	14	1.40x	76.4%
	First-response stress	Operator-only	All info states (same execution)	2	51.7%	0.79	0.73	9.51	0.00	0	—	—
		Hybrid MPC	Deterministic q50	2	65.0%	1.06	1.26	17.63	5.74	3	1.27x	70.1%
		Hybrid MPC	Bootstrap uncertainty	2	66.0%	1.34	1.74	19.84	6.28	3	1.33x	75.4%
		Hybrid MPC	Bootstrap + selective uncensoring	2	**68.0%**	**1.88**	**2.43**	**22.96**	7.91	4	1.41x	82.6%
	Unbiased operating	Operator-only	All info states (same execution)	2	80.2%	0.32	0.00	0.00	0.00	0	—	—
		Hybrid MPC	Deterministic q50	2	80.2%	0.32	0.48	1.33	0.00	0	—	—
		Hybrid MPC	Bootstrap uncertainty	2	81.2%	0.37	0.52	1.46	1.08	1	1.14x	61.3%
		Hybrid MPC	Bootstrap + selective uncensoring	2	**83.2%**	**0.96**	**1.21**	**4.28**	2.67	2	1.23x	72.8%

Notes. Service rate and mean acceptance are percentages. Lost-trip reductions, profit, and planned-bike counts are totals aggregated within each benchmark. Bold entries mark the best values in each city-regime block for service rate, 15-minute lost-trip reduction, interval lost-trip reduction, and corrected profit.

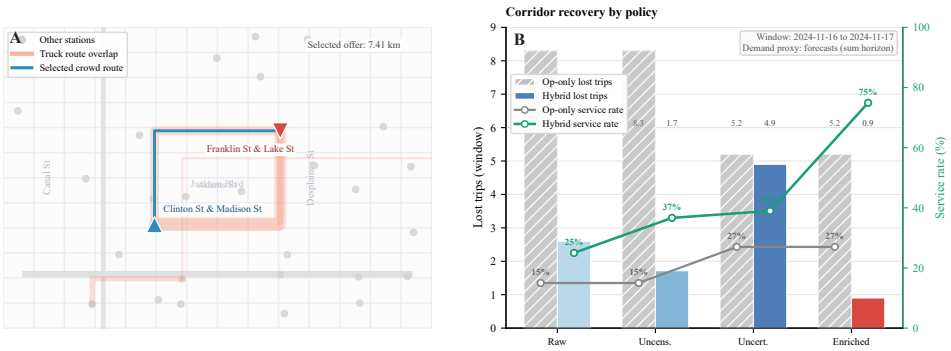


Figure 5: Corridor recovery for a selected Chicago crowd offer. **(A)** Transfer from *Clinton St & Madison St* to *Franklin St & Lake St*; blue is the selected crowd route, orange is truck-route overlap, and the offer length is 7.41 km. **(B)** Lost trips and service rate across progressively richer variants. The hybrid-enriched policy performs best, reaching 75% service and reducing lost trips relative to operator-only.

4 Conclusion

This study shows that hybrid micromobility rebalancing becomes more effective when censored-demand recovery is combined with opportunistic crowdsourced truck capacity. Relative to prior studies that separately examine operator fleets, uncertain demand, and some crowdsourced relocation, this paper contributes a city-scale framework that jointly integrates censored-demand reconstruction, real delivery-truck trajectories, stochastic participation, and rolling-horizon MPC. The service-rate gains are regime-dependent. The strongest gains arise in Chicago first-response stress and Lyon crowd-advantage settings, where either operator capacity is binding or crowd opportunities align with shortage locations. In Lyon, the full stack improves both service and profit. In Chicago, service still improves, but profit can remain negative in some crowd-advantage settings because added crowd actions do not always offset compensation and operating costs.
Overall, SATA improves the identification of latent shortages, while the hybrid service-MPC translates that information into feasible operator and crowd actions under timing, detour, and capacity constraints. Exogenous urban freight flows can complement dedicated fleets, but their economic value depends on the local regime and the feasibility of crowd deployment.

Acknowledgements

This project is supported by the EU Horizon 2020 research and innovation program under the Marie Sklodowska-Curie COFUND grant agreement no. 101034248.

References

Belkessa, L., Ameli, M., Ramezani, M., & Zargayouna, M. (2026). Integrating subway checkout data for shared micromobility demand forecasting: A novel deep learning framework. *Transportation Research Part C: Emerging Technologies*, *187*, 105650. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0968090X26001385` doi: https://doi.org/10.1016/j.trc.2026.105650

Çınar, A. B., Dullaert, W., Leitner, M., Paradiso, R., & Waldherr, S. (2024). The role of individual compensation and acceptance decisions in crowdsourced delivery. *Transportation Research Part C: Emerging Technologies*, *169*, 104834. doi: 10.1016/j.trc.2024.104834

Cebeci, M. S., Tapia, R. J., Nadi, A., de Bok, M., & Tavasszy, L. (2024). Does crowdshipping of parcels generate new passenger trips? evidence from the netherlands. *Transportation Research Record*, *2678*(6), 360-375. Retrieved from `https://doi.org/10.1177/03611981231196149` doi: 10.1177/03611981231196149

Dahle, L., Andersson, H., Christiansen, M., & Speranza, M. G. (2019). The pickup and delivery problem with time windows and occasional drivers. *Computers & Operations Research*, *109*, 122-133. doi: 10.1016/j.cor.2019.04.023

Dell'Amico, M., Iori, M., Novellani, S., & Subramanian, A. (2018). The bike sharing rebalancing problem with stochastic demands. *Transportation Research Part B: Methodological*, *118*, 362-380. doi: 10.1016/j.trb.2018.10.015

Hörl, S., & Puchinger, J. (2023). From synthetic population to parcel demand: A modeling pipeline and case study for last-mile deliveries in lyon. *Transportation Research Procedia*, *72*, 1707-1714. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2352146523009420` (TRA Lisbon 2022 Conference Proceedings Transport Research Arena (TRA Lisbon 2022),14th-17th November 2022, Lisboa, Portugal) doi: https://doi.org/10.1016/j.trpro.2023.11.644

Jin, Z., Wang, Y., Lim, Y. F., Pan, K., & Shen, Z.-J. M. (2023). Vehicle rebalancing in a shared micromobility system with rider crowdsourcing. *Manufacturing & Service Operations Management*, *25*(4), 1394-1415. Retrieved from `https://doi.org/10.1287/msom.2023.1199` doi: 10.1287/msom.2023.1199

Liu, J., Chen, W., & Sun, L. (2024). A data-driven optimization framework for static rebalancing operations in bike sharing systems. *INFORMS Journal on Computing*, *37*(5), 1369-1390. doi: 10.1287/ijoc.2022.0182

Macrina, G., Di Puglia Pugliese, L., Guerriero, F., & Laporte, G. (2020). Crowd-shipping with time windows and transshipment nodes. *Computers & Operations Research*, *113*, 104806. doi: 10.1016/j.cor.2019.104806

Merchán, D., Arora, J., Pachon, J., Konduri, K., Winkenbach, M., Parks, S., & Noszek, J. (2024). 2021 amazon last mile routing research challenge: Data set. *Transportation Science*, *58*(1), 8-11. Retrieved from `https://doi.org/10.1287/trsc.2022.1173` doi: 10.1287/trsc.2022.1173

Schuijbroek, J., Hampshire, R. C., & van Hoeve, W.-J. (2017). Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research*, *257*(3), 992-1004. doi: 10.1016/j.ejor.2016.08.029