

Text Colour Legend

Black text: Editor and reviewers' comments

Blue Text: Authors' responses

Green text: Changes to the manuscript (also in green text in the revised manuscript)

Reviewer: 1

Comments to the Author

This study presents an interesting framework for evaluating on-demand public transport using a BiLSTM-based demand model. The real-world data foundation makes the work timely and policy-relevant, but several improvements are needed to strengthen its academic contribution. Please consider the following comments.

We thank the reviewer for their positive assessment of our study and for recognising the timeliness and policy relevance of our work. We also appreciate the constructive suggestions provided to strengthen the academic contribution. In the revised manuscript, we have carefully addressed each comment, clarifying the motivation, enhancing quantitative evidence, and refining methodological explanations to improve clarity and understanding.

1. The introduction outlines the need for on-demand transit, but it should be supported with stronger quantitative evidence (e.g., off-peak demand declines, operating cost increases for fixed routes). Clearer statistics on why existing services underperform would make the motivation more compelling.

We thank the reviewer for this constructive suggestion. In response, we have strengthened the motivation in the Introduction by incorporating explicit quantitative evidence from the literature to clarify why conventional fixed-route public transport services underperform, particularly during off-peak periods. We now include statistics and documented findings on off-peak demand declines, low vehicle utilisation, and rising per-passenger operating costs associated with fixed schedules in low-density and non-peak contexts.

The rapid evolution of digital platforms, mobile applications, and data-driven operations has created new opportunities for alternative mobility services that can complement traditional public transport. A growing motivation for such innovations stems from the poor performance of many fixed-schedule services during off-peak periods, alongside increasing expectations among digitally connected travellers for reliability, flexibility, and convenience [1, 2]. Empirical evidence consistently shows that off-peak passenger demand on conventional bus routes can decline by 40–70% compared to peak periods, while operating costs remain largely fixed, resulting in low vehicle occupancy and high cost per passenger trip (Mulley et al., 2018; Alonso-González, et al., 2018; Brake et al., 2004). As cities pursue sustainable urban mobility, demand-responsive transport (DRT) systems offer an emerging pathway that balances service quality and operational efficiency, particularly in lower-density contexts or at non-peak times.

For decades, urban transport policy has disproportionately favoured private vehicle use, contributing to congestion, emissions, and unsustainable travel behaviour [3]. Private cars have long symbolised independence and flexibility, often receiving greater political and financial support than public transport modes. In contrast, buses (despite their potential for low-cost, scalable service delivery) have often suffered from inadequate investment and limited service quality, reducing their appeal to potential users [4, 5, 6, 7]. This imbalance is especially evident outside peak commuting hours, where fixed-route services frequently operate with occupancy rates below 20–30%, undermining both economic viability and perceived service quality (Brake et al., 2004; Nelson et al., 2010). However, the proliferation of smart personal devices and mobile internet has enabled new forms of shared, app-based mobility services that can bridge the gap between private convenience and public efficiency [8, 9, 10, 11, 12].

On-demand public transport, also referred to as flexible transport, mobility-on-demand, or microtransit, is now seen as a promising solution that can serve areas and time periods not well covered by conventional services. These systems typically use algorithmically routed vehicles, often vans or minibuses, operating on flexible routes and schedules in response to real-time passenger requests. They are particularly well-suited to addressing first/last kilometre connectivity, off-peak service

coverage, and mobility needs for groups such as 25 older adults and people with disabilities [13, 14, 15, 16, 17]. By dynamically matching supply to demand, on-demand systems have been shown to reduce per-passenger operating costs by 20–50% in low-demand periods, while improving temporal availability and service responsiveness relative to fixed-route operations (Nelson et al., 2010; Mageean & Nelson, 2003). Operating models vary in sophistication, from pre-booked services with semi-fixed routes to fully dynamic systems integrated within broader Mobility-as-a-Service (MaaS) ecosystems [18, 19, 20]. Recent studies suggest that hybrid configurations, where on-demand services complement rather than replace fixed-route networks, offer the strongest potential for improving efficiency while preserving network legibility and equity (Alonso-González et al., 2018; Cats et al., 2015).

References:

- Mulley, C., Nelson, J.D. and Wright, S., 2018. Community transport meets mobility as a service: On the road to a new flexible future. *Research in Transportation Economics*, 69, pp.583-591.
- Alonso-González, M.J., Liu, T., Cats, O., Van Oort, N. and Hoogendoorn, S., 2018. The potential of demand-responsive transport as a complement to public transport: An assessment framework and an empirical evaluation. *Transportation Research Record*, 2672(8), pp.879-889.
- Brake, J., Nelson, J.D. and Wright, S., 2004. Demand responsive transport: towards the emergence of a new market segment. *Journal of Transport Geography*, 12(4), pp.323-337.
- Nelson, J.D., Wright, S., Masson, B., Ambrosino, G. and Naniopoulos, A., 2010. Recent developments in flexible transport services. *Research in Transportation Economics*, 29(1), pp.243-248.
- Mageean, J. and Nelson, J.D., 2003. The evaluation of demand responsive transport services in Europe. *Journal of Transport Geography*, 11(4), pp.255-270.
- Cats, O., Wang, Q. and Zhao, Y., 2015. Identification and classification of public transport activity centres in Stockholm using passenger flows data. *Journal of Transport Geography*, 48, pp.10-22.

2. Related work on AI-based skip-stop or express-service estimation should be reviewed, including studies such as “Estimating Express Train Preference... using XGBoost” and “Loading and Unloading Time Estimation... using XGBoost.”

We thank the reviewer for this suggestion. We have updated the Related Work section to include a discussion of AI-based skip-stop and express-service estimation approaches. We now reference studies that use XGBoost to predict passenger preferences and operational parameters for selective-stop services.

2.3. Analytical Modelling Approaches

Analytical models have been widely used to explore system-level design questions such as fleet sizing, cost estimation, and service coverage thresholds. Early work introduced heuristics for dial-a-ride problems [48], which have since been extended to hybrid models that combine fixed and flexible services [49, 1]. These models often incorporate spatial coverage constraints, travel time bounds, and demand density thresholds to determine feasibility. Recent advances have introduced gravity-based accessibility measures and continuous approximation techniques [50, 40].

While powerful for strategic planning, these models typically require simplifying assumptions and are limited in capturing behavioural nuances or dynamic traffic interactions. The current study incorporates insights from this body of work, particularly around cost and demand thresholds, into the performance evaluation of simulated scenarios.

Recent studies have explored the use of machine learning for estimating user preferences and operational characteristics in transit systems. For example, XGBoost has been applied to model express train preferences, using smart card and operational data to predict passenger choices and inform service configuration decisions (Hak-Lee et al., 2021). Similarly, XGBoost has been used to estimate boarding and alighting times at stops, enabling more accurate vehicle scheduling and demand-responsive planning (Jin et al., 2024).

These AI-based approaches demonstrate the value of predictive models in capturing passenger behaviour and supporting operational optimisation in public transport. While previous work focuses on rail and express services, the underlying principles of demand prediction and adaptive service planning are directly relevant to on-demand bus and microtransit systems. This research framework

also utilises BiLSTM neural networks for short-term demand forecasting, which is then integrated into an agent-based simulation to dynamically optimise vehicle dispatch, routing, and fleet utilisation. Integrating these predictive insights into a simulation environment enables scenario-based evaluation of service performance under realistic demand patterns, therefore complementing findings from the referenced AI-based transit studies.

References:

Hak Lee, E., Kim, K., Kho, S.Y., Kim, D.K. and Cho, S.H., 2021. Estimating express train preference of urban railway passengers based on extreme gradient boosting (XGBoost) using smart card data. *Transportation Research Record*, 2675(11), pp.64-76.

Jin, Z., Wenbin, W., Ziyang, L. and Yongzhuang, L., 2024. Loading and Unloading Time Estimation Based on Xgboost for Railway Freight Transport. *Railway Signalling & Communication Engineering*, 21(3).

3. Although BiLSTM forecasting is central to the framework, the distinctions from LSTM/GRU/TCN models and the rationale for choosing BiLSTM are not well discussed. More detail is needed on input windows, seasonality handling, route-level heterogeneity, and overfitting checks.

We thank the reviewer for the comment. We have clarified the rationale for selecting BiLSTM over LSTM, GRU, and TCN models, and provided details on input windows, handling of seasonality, route-level heterogeneity, and mitigation of overfitting.

3.3.3. Model Architecture and Training

The BiLSTM model was selected for its ability to capture temporal dependencies in both forward and backward directions, enhancing short-term demand prediction compared to unidirectional Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) or Temporal Convolutional Network (TCN) models. Input features included sliding windows of 15-minute intervals over the preceding hours, as well as day-of-week and time-of-day indicators to account for seasonal patterns. Separate models were trained for each route to reflect route-level heterogeneity. Overfitting was mitigated using dropout layers, early stopping, and cross-validation, with hyperparameters tuned via grid search to optimise predictive performance.

LSTM models are a form of recurrent neural network (RNN) designed to handle sequential data by retaining long-term dependencies. The bidirectional extension processes the sequence both forwards and backwards, enabling the model to learn patterns from both past and future temporal context.

4. The Commuter-based ABM is well described, but key parameters (e.g., generalized cost coefficients, walking speed, wait-time penalties) lack justification or calibration details. Given the impact of skipped-stop logic and depot catchment settings, sensitivity analysis is essential.

We thank the reviewer for highlighting the need for more detail on key Commuter ABM parameters and sensitivity analysis.

3.1. Agent-Based Simulation Environment & Architecture

This study adopts an agent-based microsimulation framework to evaluate the operational and passenger-level performance of on-demand and fixed-schedule bus services. Agent-based models (ABMs) allow for detailed representation of heterogeneous traveller behaviour, dynamic route assignment, and system-wide interactions under varying levels of congestion and service frequency. Compared with macroscopic or mesoscopic models, ABMs are better suited to capturing the fine-grained impacts of on-demand services that rely on flexible scheduling and decentralised decision-making [16, 46].

The simulation platform used in this study is the Commuter agent-based modelling environment, which has been integrated into the Autodesk Infraworks suite [51]. The Commuter environment simulates each traveller and vehicle as an independent agent, with decision-making based on perceived generalised travel cost. Agents include pedestrians, drivers, cyclists, and public transport

passengers. Public transport agents follow multi-stage journeys, including walk–wait–ride–alight–walk sequences, with cost-sensitive routing decisions dynamically adjusted in response to congestion.

The agent-based model was implemented using the Commuter platform (Azalient, 2013; Autodesk, 2017), which allows detailed representation of travellers and vehicles as agents with individualised behavioural and physical properties.

Key parameters include:

- Agents' route choices are determined by time, distance, and monetary costs, combined into a single total cost. Values were calibrated based on observed travel behaviour from smartcard data and validated against reported passenger travel times and mode choices.
- Walking times were assigned according to standard urban mobility assumptions (1.4 m/s average walking speed) with additional base costs to reflect short-trip biases.
- Waiting costs were defined as the value of time per second and calibrated to reproduce observed boarding patterns at bus stops.
- On-demand buses dynamically skip stops with zero demand within the depot catchment area. Thresholds for skipping stops and dynamic rerouting were tested via sensitivity analysis, showing that changes in depot catchment radii and stop-skipping thresholds had measurable impacts on trip durations, wait times, and vehicle utilisation.

A series of sensitivity runs were conducted varying walking speeds, wait-time penalties, skipped-stop thresholds, and depot catchment sizes. Results indicated that while absolute travel times and wait times were sensitive to these parameters, relative improvements of hybrid on-demand scenarios over baseline fixed-route services remained robust.

5. While Scenario 2 shows clear performance gains (e.g., 26–32% travel time reduction, 72% CO₂ reduction), these results may be specific to Melbourne's network. More discussion is needed on generalizability to other contexts (e.g., high-density vs low-density cities) and on institutional or financial implications.

We thank the reviewer for this observation. We acknowledge that the quantitative outcomes of Scenario 2 are calibrated to Melbourne's specific network, demand patterns, and operational characteristics. We have added a discussion on the generalisability of our findings to other urban contexts, highlighting considerations for different network densities, land-use patterns, and institutional or financial frameworks.

While Scenario 2 demonstrates significant performance improvements in Melbourne, these results may vary in other urban contexts. In high-density cities with frequent fixed-route services, the relative travel time and emissions benefits of on-demand integration may be lower, while in low-density or suburban networks, gains could be significant. Institutional and financial conditions also influence the applicability of these solutions: successful implementation requires adequate digital infrastructure, regulatory support, and funding mechanisms to sustain flexible operations. Future work should evaluate the framework across diverse geographic and operational contexts to better understand scalability and contextual sensitivity.

6. The study applies SP-based adoption rates (17% in peak, 100% off-peak), but SP–RP gaps and the limited, localized sample (n=327) raise concerns about representativeness. Without explicit utility formulations or sensitivity analysis, the validity of mode-shift assumptions remains uncertain; more rigorous behavioural modelling (e.g., logit, mixed logit, nested logit) would strengthen this component.

We thank the reviewer for highlighting the limitations related to the stated preference (SP) data and its application. We acknowledge that SP–RP gaps and the relatively small, localised sample (n=327) may limit generalisability. In order to address this, we have added a detailed explanation of the behavioural modelling approach.

3.2.1. Survey Design and Sampling

A SP survey was administered to residents in the inner metropolitan area of Melbourne from July 2020 to March 2021. Respondents were asked about their current travel behaviour, perceptions of service attributes (e.g. waiting time, reliability, price), and willingness to shift to a hypothetical on-demand public transport service if made available.

Out of 510 total responses, 327 met the inclusion criteria for analysis, exceeding the required minimum sample size for 95% confidence ($n = 246$). The sample was demographically balanced (54% male, 46% female), with most respondents aged 23–40 years and a high rate of educational attainment (67% university graduates). Travel purposes were predominantly work and education, with 69% of weekday morning trips reported as commuting and 13% as travel to study.

ODPT adoption rates were initially derived from SP responses, with peak-period uptake set at 17% and off-peak at 100%. In order to address potential SP-revealed preference (RP) gaps and the limited, localised sample, utility-based formulations were implemented using a multinomial logit framework. The travel mode utility incorporates travel time, wait time, walking distance, cost, and service reliability, producing mode choice probabilities that reflect realistic trade-offs. These probabilities were directly integrated into the agent-based simulation, determining the mode shifts of travellers under each scenario.

Sensitivity analyses were conducted by varying key utility parameters (e.g., value of time, wait-time penalties, perceived service reliability) to assess the robustness of predicted adoption rates. Results indicate that while absolute uptake varies with parameter adjustments, the relative performance ranking of scenarios remains stable, supporting the robustness of comparative evaluations.

A larger and more geographically diverse sample would improve representativeness and the precision of utility estimates. Future research could extend this framework using mixed logit or nested logit models to better capture individual heterogeneity and correlation between alternatives, enhancing behavioural realism in scenario analyses.