Towards an integrated approach for the automatic design of feeder bus lines using agent-based simulation and combinatorial optimization

Tarek Chouaki*[1,2], Sylvain Declerq[1], and Sebastian Hörl[1]

[1]IRT SystemX, France
[2]Université Paris-Saclay, CentraleSupélec, Laboratoire Génie Industriel

*Corresponding author

Short summary

This paper explores a two-stage approach for specifying and optimizing feeder bus services for high-capacity rail stations. First, an agent-based simulation is run to identify the service that a new service would generate. Second, an Integer Linear Program is used to construct a bus line in order to satisfy the identified demand. The methodology has been designed to integrate closely with the agent-based transport simulation framework MATSim. The presented use case is based on the Grand Paris Express project in the Paris region, with a specific focus on the Paris Saclay area. The results show that the method is suitable for designing feeder lines for future Metro stations.

Keywords: agent-based, feeder service, integration, linear programming, MATSim, network design, simulation

1 Introduction

In the coming years, several public transport infrastructure projects will be finished in the Île-de-France region around Paris. Besides several new tram lines in the area, the major activity is the Grand Paris Express, which aims to further connect the Paris Metropolitan area by rail and provide tangential travel options. The project contains four new Metro lines and extends two existing ones on a total distance of 200 kilometres, with 68 new Metro stations.

As is the case for other infrastructure projects world-wide, the high-capacity rail infrastructure is planned well in advance, while it will be the duty of the local stakeholders and municipalities to plan the access to the new stations. While planners, traditionally, based the design of the feeder lines on experience and territorial knowledge, digital planning tools are increasingly used for that purpose. In particular, there has been active research on the trade-offs between classic schedule-based services and on-demand offers (Calabrò et al., 2022; Leffler et al., 2021). The latter may especially be interesting under the hypothesis that autonomous driving becomes a reality in the near future (Stevens et al., 2022; Huang et al., 2018).

Simulating door-to-door or stop-based (autonomous) mobility services is a common task in agent-based simulation today that can be applied semi-automatically to generic instances. However, approaches for the network design of bus lines, often, require prior knowledge on stops and stop-based demand Mumford (2013); Nnene et al. (2023). They are, hence, highly dependent on the modeller's knowledge of the territory and, therefore, usually not suited for an automatic assessment of many instances, like in the case of the 68 new Grand Paris Express stations.

2 Methodology

To enable such analyses, we propose a two-step process. The first component is a transport simulation that allows us to obtain the travel demand from and to the stations of a new high-capacity rail service. Not having prior knowledge about the actual design of the bus line at that stage, we make assumptions about the average access time to the stations across the territory. Second, we obtain bus stop candidates from the extracted travel demand, and define a Mixed Integer Linear Program (MILP) to find an optimal sequence of stops. A key element of our analysis chain is that no assumptions on stop locations or travel demand need to be made before running the route optimization, as this information is derived from the simulation results.

Simulation-based demand estimation

Our assessment approach has been designed around the agent-based transport simulation framework MATSim (Horni et al., 2016). However, it is applicable to any other transport simulation that (1) allows defining novel transport modes in a simplified (deterministic) way as defined below, (2) allows the user to extract detailed coordinate-level trip data.

Like in most MATSim use cases, we iterate between a detailed mobility simulation in which agents move in the road and transit networks where they experience potential delays. After, a decision-making step is performed in which agents may change their mode of transport so they can test their new choice in the next simulation cycle. This cycle continues until the global mode shares have stabilized sufficiently.

By default, MATSim agents can choose between *car*, *bicycle*, *walking* and taking *public transport* in these simulations. The latter option is simulated in detail by finding suitable transit stop candidates close to the origin and destination of a particular trip, and then performing a detailed routing in the transit network using the RAPTOR algorithm (Delling et al., 2015). In simulation, the agent will, hence, walk to the access stop, perform one or multiple legs in a transit vehicle, and then walk from the egress stop to the final destination.

The explanations above show that naively integrating a new high-capacity rail line into the simulation would strongly underestimate its demand: Based on the decision-making process of the agents, many persons that, in reality, might benefit from the new connection, would not consider the option as the access or egress walking distance would be too high. In reality, these people would be connected to the station using feeder services.

Therefore, we introduce virtual feeder services into the simulation. Such a service should give agents a realistic picture of travel times and distances from/to their point of origin/destination, but it is simulated in an abstract way.

For each station, we define such a service by (1) the average travel speed of the service, and (2) a radius around each station within which the service can be used. Next to the standard *public transport* option, we add another alternative to the pool of transport modes from which the agents may choose. This *public transport with feeder* mode constructs individual itineraries according to the following logic:

- Choose the closest transit stations for which a feeder service is available and that cover the agent trip's origin and destination location. If no station is found or the same is found for the origin and destination (short trip), the *feeder* option is not available for the agent.
- Obtain the network-routed travel times to access/egress the station.
- Construct an itinerary that contains an access segment that may either be (1) a *direct walk* or a *feeder* leg with the specified travel time, (2) the transit segment, and (3) an egress segment consisting of direct walking or a *feeder*.

During the decision-making phase of MATSim these new travel alternatives are evaluated in competition with the other existing modes. Switching between a feeder and a transit segment is interpreted as a *transfer* in the utility-based choice model, and the travel time is interpreted like standard public transport. Note that there is no notion of service frequency in the virtual feeder service, so also no waiting time can be taken into account in the choice model. However, one may provide a constant hypothesis value for the waiting time at the stop.

Finally, in the mobility phase of MATSim, an agent starting a leg using the *feeder* mode is teleported, i.e., it arrives at the feeder destination after a predefined duration specified by the travel time obtained during the routing phase.

Feeder service optimization

From the simulation, we can obtain all trips for which using a feeder line to a particular stop may be of interest. Based on this information, we (1) derive a set of stop candidates on the territory, and (2) solve a combinatorial optimization problem to find the best set and sequence of stops for a new feeder line.

To derive the stop candidates, we start by placing initial stop candidates on all links of the territory's road network, except for certain exceptions such as dead-ends, highways, and physically

separated primary roads. For each candidate i, we identify all virtual feeder trips that start or end within a configurable distance ρ. It should be set to what is deemed a realistic operational radius of the feeder service, but is rather a parameter to limit computational effort. We note down the sum of incoming and outgoing trips as the flow ϕ_i. We then proceed by finding the stop $i^* = \text{argmax}\phi_i$ with the largest flow and delete all other stops that are within a distance of ρ. We then proceed the same way with the second-largest flow among the remaining stops, and so forth, until no more stops can be deleted.
Based on the selected stops, we obtain the following information from the demand data and by routing stop-to-stop trajectories in the road network:

- Let $\mathcal{S}$ define the set of stops
- Let $F_{ij} \in \mathbb{N}$ define the flows between stops $(i, j) \in \mathcal{S}^2$
- Let $D_{ij} \in \mathbb{R}$ define the distance between stops $(i, j) \in \mathcal{S}^2$

The goal of the route optimization problem is to find a sequence that maximizes the transported flow while minimizing the distance. In our problem formulation, we make use of two problem parameters:

- $\sigma \in \mathbb{R}$ describes the trade-off between flow maximization and distance maximization.
- $T \in \mathbb{N}$ is the length of the bus line in number of stops. It needs to be defined a priori, but multiple values can be tested by repeatedly solving the problem.

The problem makes use of the following binary variables:

- $s_i \in \{0, 1\}$ describes whether the bus line goes through stop $i \in \mathcal{S}$. Only then any flow can be transported from or to the stop.
- $y_{ij} \in \{0, 1\}$ describes whether the flow from stop $i \in \mathcal{S}$ to stop $j \in \mathcal{S}$ is transported by the bus line.
- x_{ijt} describes whether the bus moves from stop $i \in \mathcal{S}$ to stop $j \in \mathcal{S}$ at index $t \in \mathcal{T}$ with $\mathcal{T} = \{1, ..., T\}$ indicating the set of stop indices.

The full optimization problem is stated below:

$$\begin{aligned}
\underset{x,y,s}{\text{maximize}} \quad & \sum_{(i,j)\in\mathcal{S}^2} y_{ij} \cdot F_{ij} - \sigma \sum_{(i,j,t)\in\mathcal{S}^2\times\mathcal{T}} x_{ijt} \cdot D_{ij} && && (1)\\
\text{subject to} \quad & \sum_{i\in\mathcal{S},t\in\mathcal{T}} x_{iqt} = s_q && \forall q \in \mathcal{S} && (2)\\
& \sum_{j\in\mathcal{S},t\in\mathcal{T}} x_{qjt} = s_q && \forall q \in \mathcal{S} && (3)\\
& \sum_{(i,j)\in\mathcal{S}^2} x_{ijt} = 1 && \forall t \in \mathcal{T} && (4)\\
& \sum_{i\in\mathcal{S}} x_{iqt} = \sum_{j\in\mathcal{S}} x_{qj(t+1)} && \forall q \in \mathcal{S}, t \in \mathcal{T}\backslash\{T\} && (5)\\
& 10y_{ij} \leq 9(s_i + s_j) && \forall (i, j) \in \mathcal{S}^2 && (6)\\
& && && (7)
\end{aligned}$$

The objective (1) includes the total transported flow by multiplying whether relation (i, j) is covered by the line with the available flow F_{ij}. This value is compared using the tradeoff parameter σ with the total covered distance, which is obtained by multiplying whether the relation (i, j) is driven at any point along the line t with the stop-to-stop distance D_{ij}.
Constraints (2-3) establish that if a stop q is part of the line ($s_q = 1$), there must be exactly one movement towards q (1) and one leaving q (2).
Constraint (4) requires that only one movement is possible at index along the stop sequence.
Constraint (5) is a continuity constraint that makes sure that if there is an incoming movement to, q at index t there must be an outgoing movement from q at index $t + 1$.
Finally, constraint (6) says that a flow from i to j is covered by the line iff both i and j are part of the line. Note that we assume bidirectional service of the line, i.e. it is served in the order of stops obtained by our problem, and also in the opposite direction.

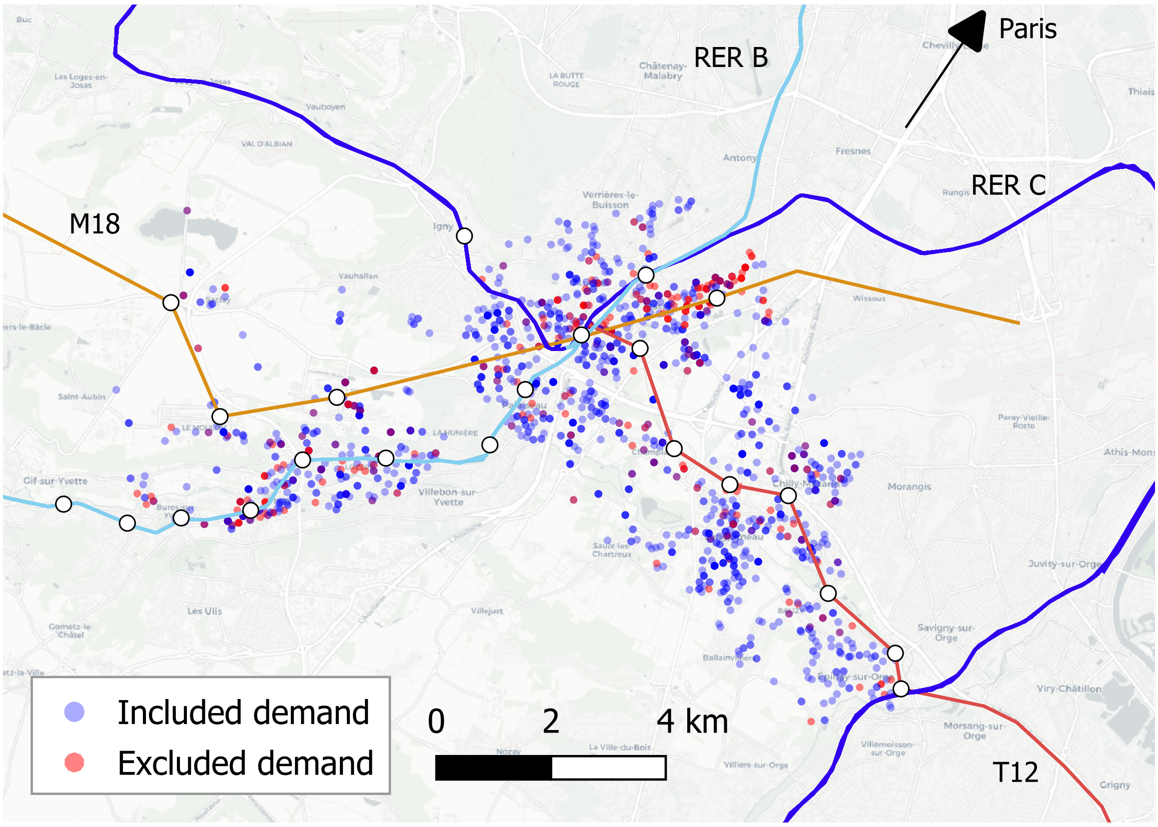

Figure 1: Map of the study area with existing high-capacity lines RER B and C (blue) and future lines Metro 18 (red) and Tram 12 (orange). Background: OpenStreetMap.

Note that in its basic formulation, the problem will form a round-trip due to constraints (2) and (3). However, as described above, we assume a line that is operates in two directions, so we prefer to obtain distinct start and end points. This can be modelled by requesting a line of length $T + 1$ and introducing a virtual stop V. The distance between V and any other stop is fixed to zero. It then requires introducing an additional constraint that forces the line to pass through V:

$$\sum_{j \in \mathcal{S}} x_{Vj0} = 1 \tag{8}$$

There are various ways of simplifying the problem. For the base case of our test scenario, we obtain 244 stops with 41,292 edges. In practice, we reduce the problem by observing that it is very unlikely that the solver will choose two stops that are very far located from each other. For each stop, we keep the edges to the N closest stops according to the distance matrix. With this approach, as long as we define $N \geq T$, the problem is guaranteed to remain feasible.
In practice, for our use case outlined below, a value of $N = 10$ gives 1,550 edges. Such an instance with a line length of $T = 10$ can be solved in less than 10 minutes on a standard computer using a recent version of `CPLEX`. The linear program itself was implemented using the `pyomo` library.

3 Results and Discussion

We make use of the process described above for a use case of the Grand Paris Express. The daily activity chains of the agents are based on an open, disaggregated travel demand for the Île-de-France region around Paris (Hörl & Balac, 2021). The behavioural model has been estimated on a local Household Travel Survey (Hörl, 2023).
The specific use case is the Paris Saclay area, for which we have previously studied the impact of autonomous on-demand feeder services (Chouaki et al., 2023). The area is located in the South of Paris and is connected to the Metropolitan centre by two commuter rail lines (RER B and RER C). Two future transport lines (Metro 18 and Tram 12) have been added to the baseline transport offer, as shown in Figure 1.

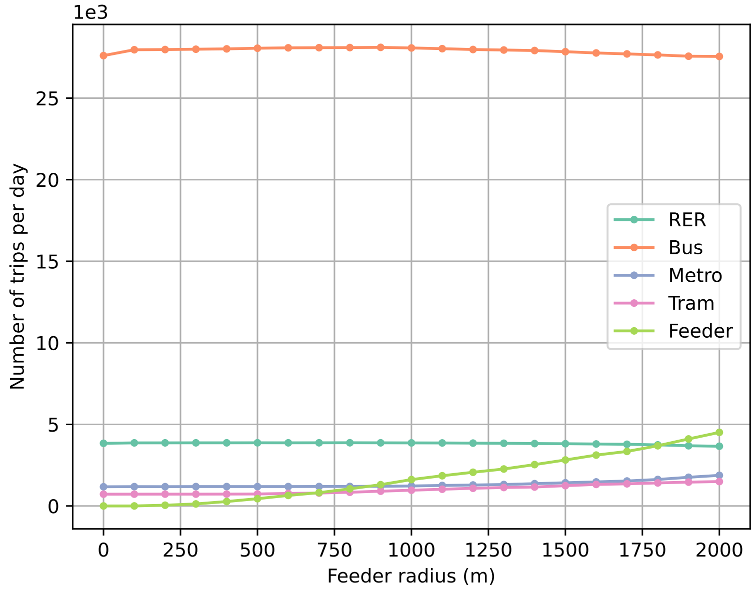

Figure 2: Impact of feeder services on usage of public transport modes

Potential demand

To assess the potential of classic bus services, we define a virtual access service around every high-capacity transit stop (Figure 1). The services are parameterized with an average speed of 15km/h (considering stop times to pick up passengers) and we test various cases with feeder perimeters from 100m to two kilometres.

For computational efficiency, only 10% of the households interacting with the area have been simulated, and the results have been scaled up accordingly. Furthermore, decisions are only allowed for trips that happen entirely in the study area. In particular, this limitation excludes commuters going to and coming from Paris, which will be considered in future simulations on a larger scale.

First, we examine the global mode share in our simulation. We observe no substantial shift from any other mode towards public transport when introducing the feeder services. This observation is due to the fact that in our exploratory use case, agents are only allowed to change their mode of transport for trips within the study area, but not to go to Paris or more remote areas. This merely computational burden will be overcome in future research with currently ongoing developments.

In contrast, Figure 2 shows the passenger volumes on the transport lines in the area. The number of feeder users clearly increases with increasing operating radius, and we observe that the use of trams and Metro slightly increases while the use of buses and commuter rail is reduced. Figure 3 gives a more detailed picture by examining individual transport lines. One can see that the two new lines gain more passengers the wider the range of the feeder becomes. The use of the existing RER C stays roughly the same, while the use of RER B decreases.

What we observe in Figure 2 is that the feeder services that have been configured for the new transport lines T12 and M18 slightly gain new passengers from the introduction of the feeder up until 800m. After, the increase becomes much stronger, but we also see that passengers are attracted from RER B. At that point, the travellers that were previously accessing RER B via existing feeder services make use of the new virtual infrastructure to the new lines.

The disaggregated, agent-based nature of our simulation also allows us to perform detailed analyses on the use of individual stations (Figure 4). For the regional transport hub Massy-Palaiseau the increase depends almost linearly on the feeder radius, while we see a great variety in impact on other stations. This analysis also gives us hints about the importance of the different stations with respect to implementing a feeder service for them or not.

Feeder design

To design the feeder bus lines, we obtain the potential travel demand based on the daily trips performed with the virtual feeder mode and a feeder radius of 2 kilometres. From this set of trips, we remove all those that have used any of the existing high-capacity lines in the baseline case. This way, we avoid that generated feeder bus lines pick up demand that already had sufficient

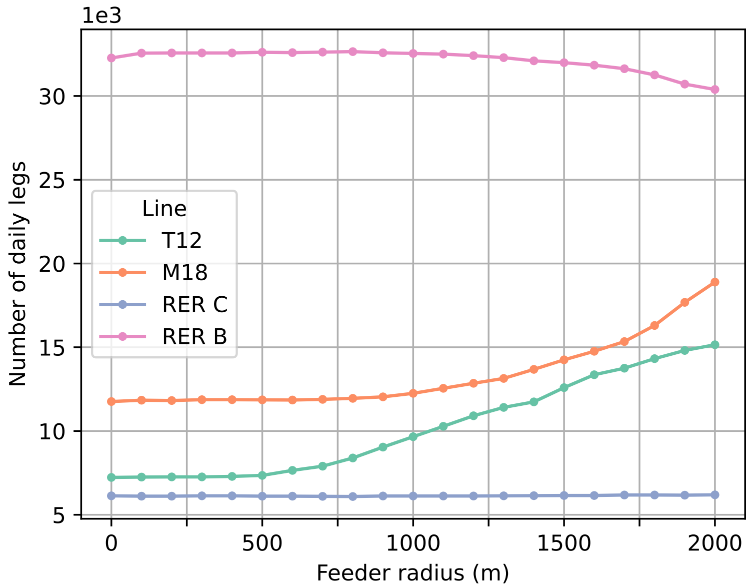

Figure 3: Impact of feeder services on attractiveness of existing and new rail lines

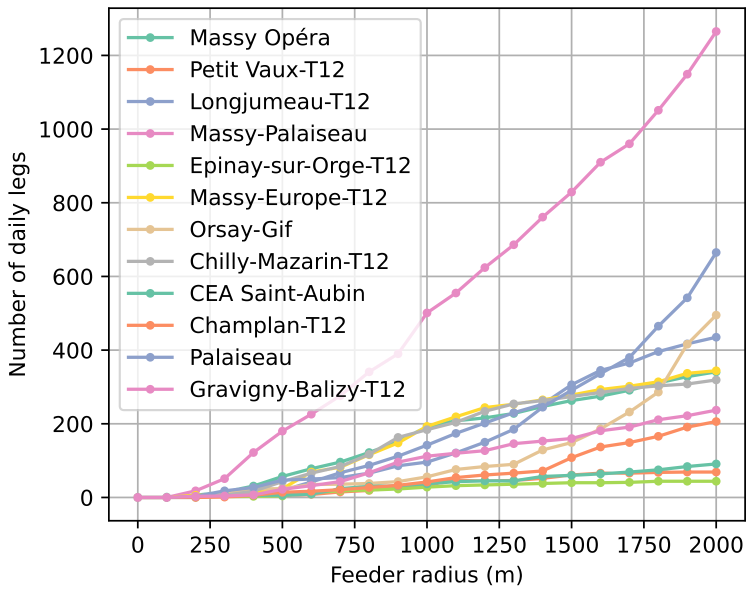

Figure 4: Number of feeder users per rail station

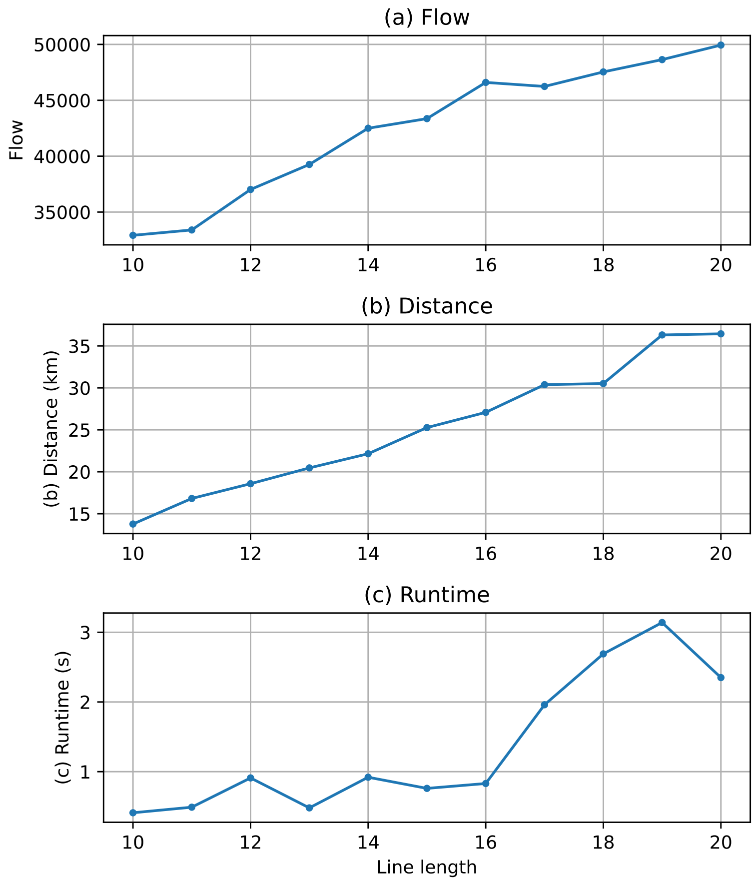

Figure 5: Sensitivity analysis of the line length on the Optimization: (a) satisfied passenger flows, (b) resulting feeder bus line distance and (c) optimization runtime

access to the high-capacity lines. On Figure 1 we show the origins of the remaining trips that are our candidates for using a new feeder lines in comparison to those that have been excluded. We observe that there is, as expected, a demand for feeder operations in currently underserved areas. For our exploration of the approach, we run the optimization with varying line lengths L. Figure 5 shows the transported passengers (a) and the distance (b) dependent on the line length. One can see that both increase with an increased number of stops. However, while the distance increases almost linearly with the number of stops, the flow is slowly saturating towards twenty stops. The gain of passengers per added distance, hence, decreases with longer lines. Furthermore, Figure 5c shows the runtime[1], which stays very low for all instance sizes.

The instances have been solved with $\sigma = 10^-3$, an arbitrary choice that has provided satisfying results in earlier test runs. In a realistic use case, this parameter should be calibrated in terms of infrastructure and operating expenses and the expected revenues of the line.

Figure 6 shows two examples of the obtained lines, for $L = 10$ (a) and $L = 20$ (b). In both configurations, we see that a new line is developed in underserved areas close to the central mobility hub while the longer example additionally covers demand that is close to T12 and especially an area between the new lines M18 and T12.

4 Conclusion

Our experiments show the general feasibility of our approach, in which we connect a detailed demand simulation with virtual feeder services and a route optimization problem. While there is need for further need in tweaking the model parameters, the modelling pipeline provides promising results.

In future research, we plan to further extend the planning pipeline by setting the frequency of the obtained bus lines. This, in turn, will allow us to reintegrate the generated bus schedules in our

[1] Intel(R) Xeon(R) CPU E7-8880 v4 2.20GHz

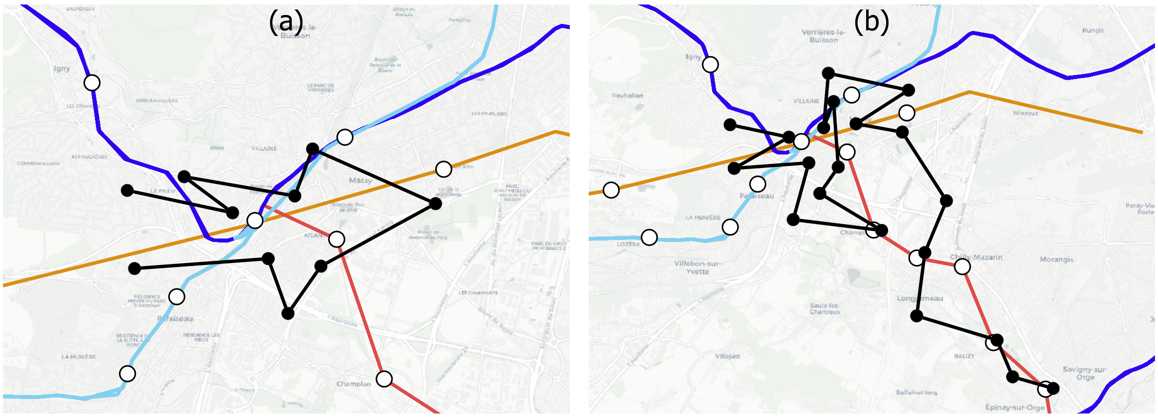

Figure 6: Examples of the generated feeder bus lines with ten (a) and twenty (b) stops. Background: OpenStreetMap.

agent-based simulation. This way, it would even be possible to run a closed loop in which the bus lines get refined iteratively based on the complex decision-making of the traveller agents.

Acknowledgements

This work has been supported by the French government under the “France 2030” program, as part of the SystemX Technological Research Institute and the Anthropolis Chair.

References

Calabrò, G., Le Pira, M., Giuffrida, N., Inturri, G., Ignaccolo, M., & Correia, G. H. d. A. (2022, January). Fixed-Route vs. Demand-Responsive Transport Feeder Services: An Exploratory Study Using an Agent-Based Model. *Journal of Advanced Transportation*, *2022*, 1–20.

Chouaki, T., Hörl, S., & Puchinger, J. (2023, January). Towards Reproducible Simulations of the Grand Paris Express and On-Demand Feeder Services. In *102nd Annual Meeting of the Transportation Research Board (TRB 2023).* Washington D.C, United States.

Delling, D., Pajor, T., & Werneck, R. F. (2015, August). Round-Based Public Transit Routing. *Transportation Science*, *49*(3), 591–604.

Horni, A., Nagel, K., & Axhausen, K. W. (Eds.). (2016). *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press.

Huang, D., Liu, Z., Fu, X., & Blythe, P. T. (2018). Multimodal transit network design in a hub-and-spoke network framework. *Transportmetrica A: Transport Science*, *14*(8), 706–735. (ISBN: 2324-9935 Publisher: Taylor & Francis)

Hörl, S. (2023, June). Towards replicable mode choice models for transport simulations in France. In *9th International Symposium on Transportation Data & Modelling (ISTDM2023).* Ispra, Italy: Joint European Research Center.

Hörl, S., & Balac, M. (2021, September). Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transportation Research Part C: Emerging Technologies*, *130*, 103291.

Leffler, D., Burghout, W., Jenelius, E., & Cats, O. (2021, November). Simulation of fixed versus on-demand station-based feeder operations. *Transportation Research Part C: Emerging Technologies*, *132*, 103401.

Mumford, C. L. (2013). New heuristic and evolutionary operators for the multi-objective urban transit routing problem. In *2013 IEEE congress on evolutionary computation* (pp. 939–946). IEEE.

Nnene, O. A., Joubert, J. W., & Zuidgeest, M. H. (2023). A simulation-based optimization approach for designing transit networks. *Public Transport*, 1–33. (ISBN: 1866-749X Publisher: Springer)

Stevens, M., Correia, G. H. d. A., Scheltes, A., & van Arem, B. (2022, September). An agent-based model for assessing the financial viability of autonomous mobility on-demand systems used as first and last-mile of public transport trips: A case-study in Rotterdam, the Netherlands. *Research in Transportation Business & Management*, 100875.