

# GROSS: German Rail Open-Source SUMO Scenario

Juri Penell<sup>1</sup>, and Damian Dailisan<sup>1</sup>

<sup>1</sup>Computational Social Science ETH Zurich, Switzerland

\*Correspondence: Juri Penell, [jpenell@ethz.ch](mailto:jpenell@ethz.ch)

**Abstract.** Microscopic simulation enables reproducible evaluation in intelligent transportation systems, yet most open SUMO scenarios and toolchains remain road-traffic centric, leaving rail underrepresented despite its importance for public transport and its sensitivity to network-wide disruptions. We present the German Rail Open-Source Scenario (GROSS), an open pipeline that combines OpenStreetMap railway infrastructure with GTFS schedules to generate nation-scale rail scenarios for SUMO (Simulation of Urban MObility). Existing conversions often rely on geometry-only stop-to-track matching and inconsistent platform/track assignments, which can create routing anomalies and unstable simulations dominated by teleportation artefacts. GROSS addresses this with topology-aware stop mapping via a hierarchical station model, followed by station-level routing with validation and targeted repair. Across multiple German regions, GROSS reduces average teleportations per vehicle by a factor of 1.7–76.8×, shortens delays compared to the vanilla SUMO pipeline, and it enables end-to-end generation of a Germany-wide scenario with 35 925 trips for comparisons with operator-reported delay statistics. While the remaining long delays highlight limitations in available timetable metadata and rail dispatch modeling, GROSS lowers the barrier to building scalable, fully open rail simulations and to studying delay propagation at country scale.

**Keywords:** Rail simulation, Open-source, Routing

## 1. Introduction

Microscopic traffic simulation plays a central role in intelligent transportation systems (ITS) research, enabling reproducible evaluation of control strategies, infrastructure design, and mobility policies under realistic operating conditions [1]. Among existing simulation platforms, SUMO is widely adopted due to its open-source nature and ability to support large-scale, network-level experiments. Consequently, substantial effort has been devoted to the development of open traffic scenarios that approximate real-world mobility patterns [2]–[7]. Such traffic scenarios are also widely used as a controllable testbed for data-driven, learning-based traffic research [8]–[11].

To date, however, most publicly available large-scale scenarios are predominantly road-traffic oriented. Well-established examples for various cities, such as the Luxembourg [3], Monaco [4], Tokyo [6], and Berlin [5] SUMO traffic scenarios focus primarily

on private vehicular mobility, typically relying on demographic data, traffic counts, or calibrated demand generation pipelines. These scenarios have significantly improved reproducibility and comparability in ITS research, yet they capture only a limited subset of modern mobility systems.

In contrast, rail-based transportation constitutes a fundamental component of mobility in many European countries, particularly in Germany, where regional, suburban, and long-distance rail services form the backbone of daily commuting and intercity travel [12]–[14]. Despite its importance, rail mobility remains underrepresented in open, reusable simulation scenarios. This gap is notable given that railway systems are highly susceptible to network-wide disruptions: infrastructure failures, capacity bottlenecks, and tightly coupled timetables can lead to cascading delays that propagate across large geographic areas [15], [16]. Such phenomena significantly degrade service reliability and hamper passenger mobility, motivating increasing interest in robustness analysis and disruption management [17].

Simulation-based approaches are particularly well-suited for studying these effects, as they allow controlled experimentation under repeatable conditions. However, the lack of openly available, large-scale rail scenarios has limited systematic investigation of railway operations within the ITS community. At the same time, the availability and quality of open-access mobility data have improved substantially. OpenStreetMap provides detailed representations of railway infrastructure [18], while the General Transit Feed Specification (GTFS) offers a standardized description of schedules, stops, and services [19]. Together, these data sources enable the construction of realistic rail simulation environments without reliance on proprietary datasets.

In this work, we present a pipeline for generating country-scale rail simulation scenarios in SUMO [1], [20] using open-access OSM and GTFS data. We apply our pipeline to create train-focused simulation scenarios in Germany. Unlike existing SUMO pipelines and baseline examples, we show that our approach improves operational realism by reducing the amount of deadlocks and artificial teleportations that occur at junctions in the simulation. We then demonstrate that our pipeline can be effectively scaled to the national level and create a Germany wide rail scenario. We characterize the level of realism these scenarios can achieve by comparing them to existing real-world rail statistics from the Deutsche Bahn (DB). Our rail scenario can serve as a complement to other existing car-centric scenarios, leading to a more complete representation of mobility systems and providing a foundation for future studies on railway operations, delay propagation, and multimodal transportation analysis [21].

## 2. Related Work

### 2.1 Open SUMO Traffic and Rail Scenarios

The SUMO community has released a number of widely used open scenarios to foster reproducible experimentation with microscopic traffic simulation [1]. Early work, such as the Bologna scenario, demonstrated end-to-end scenario creation from open data and helped establish common practices for network extraction, demand generation, and calibration [2]. More recent city-scale scenarios—including LuST (Luxembourg) [3]–[6], MoST (Monaco) [4], ToST (Tokyo) [6], and BeST (Berlin) [5]—provide curated networks and demand models that are frequently used as baselines and benchmarks. However,

these scenarios primarily focus on road traffic; rail infrastructure and train operations are either not included or play only a minor role.

Compared to road-traffic scenarios, openly available large-scale rail scenarios for SUMO remain scarce, despite SUMO's support for rail vehicle dynamics and signaling abstractions. This limits systematic investigation of railway-specific phenomena such as delay propagation, conflicts at junctions, and the interplay between timetables and infrastructure capacity.

## 2.2 Timetable Robustness and Delay Propagation

Robustness analysis and disruption management have long been studied in railway operations research. Ref. [17] formalizes notions of timetable robustness and discusses how slack and buffer times relate to delay recovery. At the network level, partial capacity degradations can lead to disproportionate impacts on passenger service, motivating vulnerability- and resilience-oriented evaluation methodologies [16]. Recent work further emphasizes the systemic nature of delay cascades and proposes models and mitigation strategies that account for network-wide coupling effects [15]. Our work complements this literature by providing an openly reproducible, country-scale simulation scenario that enables controlled experiments on delay propagation under realistic operating constraints.

## 2.3 Scenario Construction Pipelines and Practical Limitations

Many SUMO tutorials and scenario-generation pipelines are designed with road traffic in mind, and their rail-related examples are typically small, synthetic, or hand-crafted. In large rail networks, naive conversion of junction topology and right-of-way rules can lead to deadlocks, unrealistic blocking, and excessive teleportations, reducing the validity of operational studies. By focusing on rail-specific network processing and schedule-driven operations, our pipeline aims to bridge the gap between tutorial-scale examples and country-scale railway simulations.

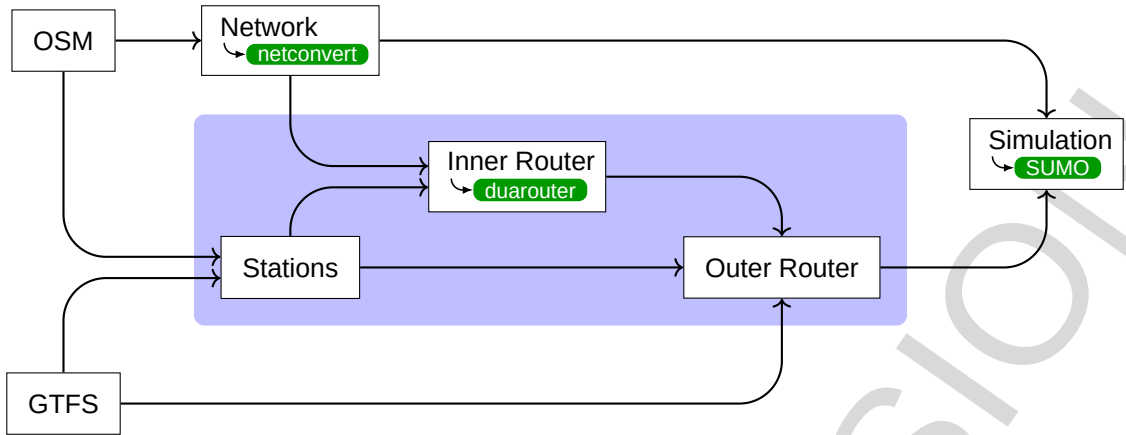
# 3. Methodology

In this section, we describe the methodology and pipeline used to create GROSS (Fig. 1). In the following subsections, we will describe the following: (3.1) Data, (3.2) Baseline, (3.3) Network, (3.4) Stations, (3.5) Routing, and (3.6) Simulation Experiments.

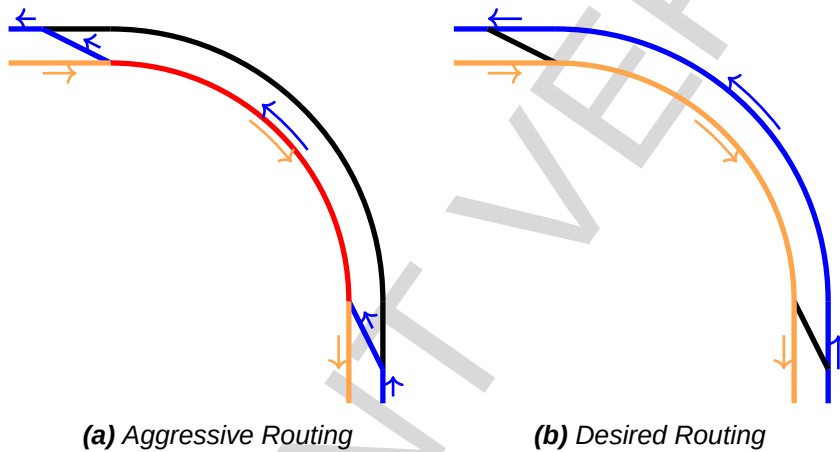
## 3.1 Data

The pipeline relies on two complementary open data sources that together provide both the physical infrastructure and the operational schedule of a rail network: OpenStreetMap (OSM) provides the transport infrastructure [18], and the General Transit Feed Specification (GTFS) provides the passenger train schedules.

The OSM dataset allows us, after intensive processing, to extract information on train tracks, signals, station platforms, stop positions, and other physical infrastructure.



**Figure 1.** Flowchart of GROSS pipeline architecture. Each box represents a discrete processing step, with green highlights identifying the use of a SUMO tool within a step. The blue-shaded region signifies the public transport processing block, which is comparable to `gtfs2pt`.



**Figure 2.** Curve Routing with clockwise route (orange), counter-clockwise route (blue), and both routes conflicting (red)

The GTFS records contain the routes, timetables, and basic stop information. Specifically, we use GTFS feeds provided by Delfi [19], which also include additional metadata (in the form of IFOPT numbers) that help cross-reference the GTFS stops to the OSM stations. Although this additional metadata helped improve train routing in this work, the pipeline is designed to still function if explicit cross-referencing between GTFS and OSM is absent.

### 3.2 Baseline

We benchmark our processing pipeline against the typical workflow one would use following the standard SUMO tools (v1.26), as outlined in the `sumo-berlin` repository of SUMO [22]. Given an OSM network and GTFS schedules, this baseline approach primarily uses `netconvert` to create the rail network with candidate stops, and `gtfs2pt.py` to map the GTFS routes onto the network. We then patch the output of `gtfs2pt` to remove trips that were filtered out due to incorrect tagging (e.g. buses and trams incorrectly tagged as trains), and split trips at segments that would require impossible speeds (Tab. A2).

### 3.3 Network

The rail network used for the simulation is generated using the SUMO tool `netconvert`. We use it to apply changes and convert the OSM data into a simulation network where all tracks are modeled as bidirectional edges because directional edges introduce non-recoverable anomalies. Stop positions, representing the exact location a train comes to a halt at a station, are taken as the simulation stop locations. We penalize crossovers to discourage frequent track changes and stop edges to minimize conflicts of stopping- and through-traffic. Priority is set based on inferred track direction based on using `netconvert` direction priority, and node order [23]. Stop edges were extended to accommodate vehicles stopping within one signal block.

### 3.4 Stations

First, we extract all relevant station information from OSM, building a structured framework of stop areas, platforms, platform edges, and stop positions. These elements are then combined with the stops from GTFS to build spatial clusters. All elements that are within 500 meters of each other are clustered and assumed to be part of the same station.

Within a station, we attempt to attribute all applicable information derived from platforms, platform edges, and the overall station to the stop positions. Using that information, we construct a penalized mapping from each GTFS stop  $h$  to all  $m$  simulation stops:

$$\text{StopMap} : h \rightarrow \{(s_{sim,j}, pen_j)\}^m, \quad 0 < j \leq m. \quad (1)$$

The penalty  $pen_j$  consists of a component for the confidence in assignment  $pen_{ass}$  plus the distance  $dist$  from the GTFS stop to the track the stop position is on. This approach allows us to determine preference for track and stop usage without discarding information and disqualifying alternative stop options.

### 3.5 Routing

The routing of trips is implemented via a two-layered, weighted, nested shortest-path approach. This architecture not only bridges the gap between high-level GTFS schedules and microscopic simulation constraints but also ensures high error resistance. A proper routing algorithm is necessary to avoid simple issues that arise in the SUMO-tools pipeline, such as aggressive routing on one side of parallel tracks (Fig. 2).

Compared to a fixed-stop approach, it handles local data inconsistencies more effectively because it can change multiple parameters simultaneously.

#### 3.5.1 Formal Definitions

We define  $S$  as the set of all stops and  $T$  as the set of all stop times. Let  $R \subseteq \{S \times T\}^*$  be the set of all routes, where the subscripts  $_{sim}$  and  $_{gtfs}$  denote the simulation and GTFS domains, respectively. A scheduled route  $r_{gtfs} \in R_{gtfs}$  with  $n$  stops is defined as:

$$r_{gtfs} = ((h_1, t_1), (h_2, t_2), \dots, (h_n, t_n)) \quad (2)$$

where each pair  $(h_i, t_i)$  represents a stop identifier and its scheduled time.

**Algorithm 1** outerRouter

```

Input: Route  $r \leftarrow ((h_1, t_1), (h_2, t_2), \dots, (h_n, t_n))$ 
Initialize:  $V \leftarrow \emptyset, E \leftarrow \emptyset, w_s \leftarrow \emptyset, w_e \leftarrow \emptyset$ 
for  $i \leftarrow 1$  to  $n$  do
     $S_i \leftarrow \text{StopMap}(h_i)$ 
    for  $(s_{i,j}, p_{i,j}) \in S_i$  do
         $V_i \leftarrow V_i \cup \{s_{i,j}\}$ 
         $w_s(s_{i,j}) \leftarrow p_{i,j} + \text{additionalPenalties}(s_{i,j}, t_i)$ 
     $V = V \cup V_i$ 
for  $i \leftarrow 1$  to  $n - 1$  do
     $S_i \leftarrow \text{StopMap}(h_i)$ 
     $S_{i+1} \leftarrow \text{StopMap}(h_{i+1})$ 
    for  $s_{i,j} \in S_i$  do
        for  $s_{i+1,k} \in S_{i+1}$  do
             $E \leftarrow E \cup \{(s_{i,j}, s_{i+1,k})\}$ 
             $w_e((s_{i,j}, s_{i+1,k})) \leftarrow \text{innerRoute}(s_{i,j}, s_{i+1,k})$ 
            if reversing then
                 $w_e((s_{i,j}, s_{i+1,k})) \leftarrow w_e((s_{i,j}, s_{i+1,k})) + \text{reversalPenalty}$ 
 $g \leftarrow G(V, E, w_s, w_e)$ 
 $\text{path} \leftarrow \text{shortestPath}(g, V_1, V_n)$ 
return  $\text{path}$ 

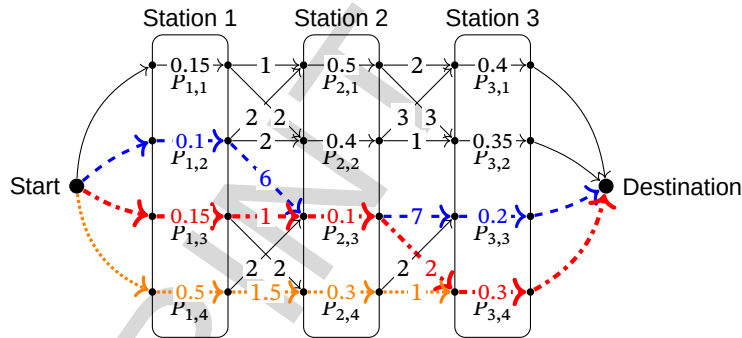
```

▷ Build node set and assign penalties

▷ Node weight (penalty)

▷ Connect subsequent stations

▷ Precomputed Values

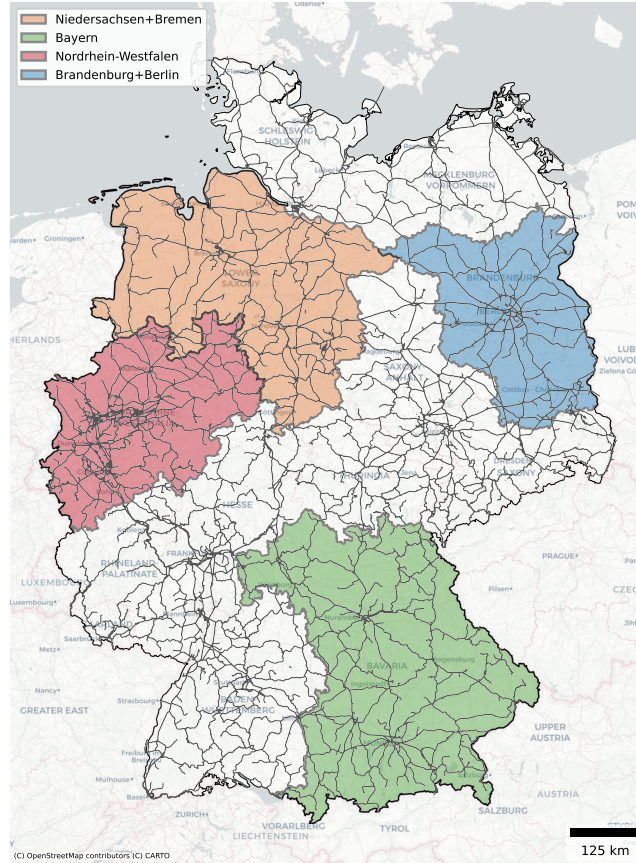


**Figure 3.** Example routing showing a fixed stop routing (blue, dashed), a lowest edge cost traversal (orange, dotted), and routing with both edge and node weights (red, dash–dotted). The top two rows represent tracks with a preferred direction left, and the bottom two rows have a preferred direction right.

**3.5.2 The Two-Layer Approach**

The routing logic is split into an internal micro-router and an external macro-router to balance physical accuracy with network flexibility:

- **Inner Layer (Micro-Routing):** This layer utilizes a `duarouter` to determine the physical traversal cost between specific platform tracks. It defines the function  $\text{innerRoute}(u, v)$ , which calculates the distance or travel time between subsequent stations. If no physical path exists, it returns  $\infty$ . These values are computed before running the outer layer.
- **Outer Layer (Macro-Routing):** This layer constructs a directed graph  $G = (V, E)$  for each route. The nodes  $V$  represent the simulation stops  $s_{i,j}$  mapped from GTFS stop  $h_i$  via the `StopMap` function (Alg. 1).



**Figure 4.** Study areas for this paper. We applied the GROSS pipeline on subregions of Germany to compare its performance using only SUMO tools.

Edge weights  $w_e$  are derived from the inner router's cost, while node weights  $w_s$  represent platform-specific penalties. This redundancy provides error resistance: if inconsistent, wrong, or missing data breaks a connection, the graph is still able to select the holistically best solution (Fig. 3).

### 3.5.3 Node Penalty Calculation

The penalty  $p_{i,j}$  assigned to a simulation stop  $s_{sim,i,j}$  is designed to minimize platform conflict through stochastic and temporal weighting. The penalty is formulated as the sum of three distinct factors:

$$\text{additionalPenalties}(s_{i,j}, t_i) = R + U_{short}(t_i) + U_{long}(t_i) \quad (3)$$

where  $R$  is a random variable sampled from the uniform distribution  $\mathcal{U}(0, 8)$ . The temporal components,  $U_{short}$  and  $U_{long}$ , are defined by linear ramps based on the planned arrival  $t_{arr}$  and departure  $t_{dep}$ .  $U_{short}$  provides a peak of 64 during the planned stop (ramping up over 5 minutes and down over 10 minutes). In contrast,  $U_{long}$  provides a broader baseline penalty of 16, starting 1 hour before arrival and persisting until 2 hours after departure to account for extended platform occupancy.

### 3.5.4 Route Repair

The route repair applies in addition to the flexible routing some adaptations based on issues we have found in the data and processing with 18% of trips affected (Tab. A2).

**Table 1.** Teleports per vehicle and vehicle count by region and scenario (averaged across seeds and dates). Teleportations are reduced by a factor of 1.7 to 76.8× compared to SUMO.

Region	Trips		Teleports/veh	
	SUMO	GROSS	SUMO	GROSS
Bayern	7451	7341	1.015	0.583
Brandenburg	2361	2381	0.736	0.045
Niedersachsen	3165	2976	0.764	0.017
Nordrhein-Westfalen	6295	6199	0.922	0.012
Germany	37710	35925	0.991	0.219

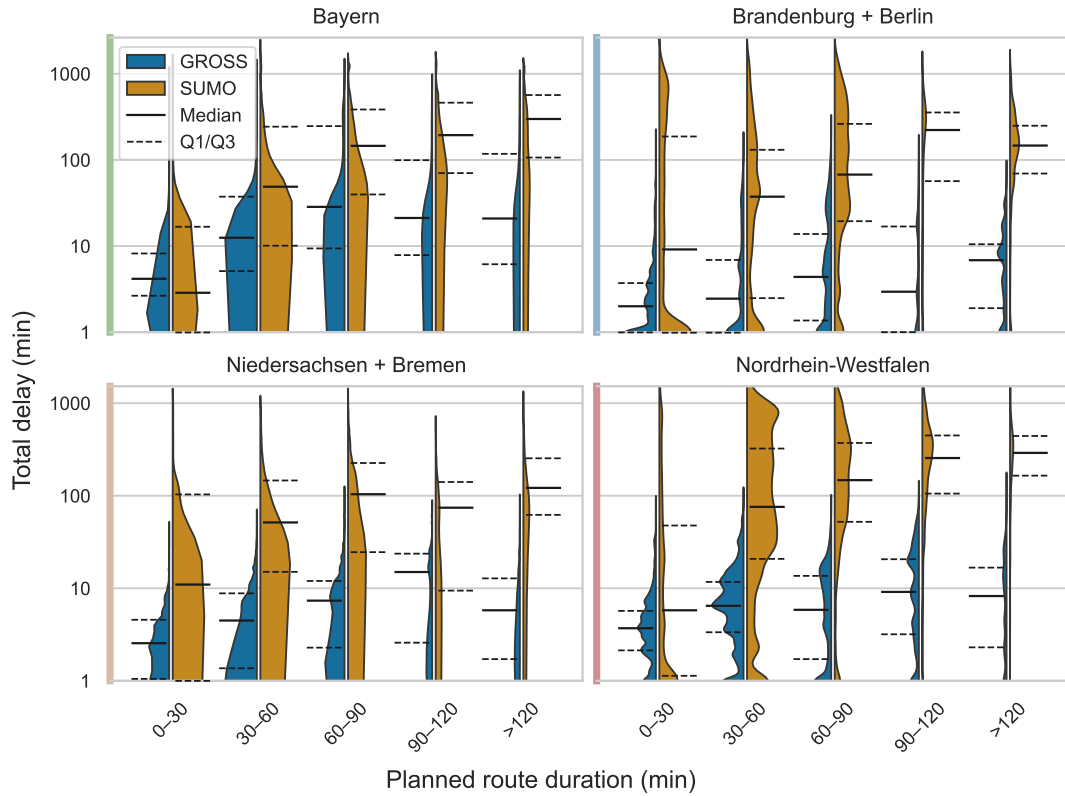
For routes that stop at a station with a single stop platform that, at the planned time, is already occupied, we allow computing the route to the following station. For missing stops (due to stops incorrectly tagged in OSM or outside the simulated networks) or sections that would require an impossible speed to traverse within the scheduled time, we allow skipping the stop by removing the vehicle at the previous stop and inserting it at the following stop after the scheduled travel time (while retaining delays) using the `jump` stop attribute in SUMO.

### 3.6 Simulation Experiments

Using SUMO (v1.26), we created and ran multiple scenarios with the GROSS pipeline on sub-regions of Germany. We sampled the GTFS dataset to create scenarios representing Wednesdays and Fridays between 14–28 January 2026, a period without public or school holidays. We selected Wednesdays as a representative weekday and Fridays to capture additional long-distance routes connecting Germany to other European cities. To highlight the impact of teleportation effects (and the resulting fix), we set `time-to-teleport` to 1800 s and `time-to-teleport.deadlocks` to 300 s. For each scenario, we ran 10 replicates, each with a different seed. We focused on four sub-regions to compare the GROSS pipeline against a SUMO tool pipeline. We focus on the sub-regions of Bayern, Brandenburg (+ Berlin), Niedersachsen (+ Bremen), and Nordrhein-Westfalen to compare the GROSS pipeline against a SUMO tool pipeline (Fig. 4).

## 4. Results

Our processing pipeline enables substantially faster generation of usable SUMO rail scenarios than the vanilla SUMO toolchain (Tab. A1). Across our four evaluation regions (Brandenburg (+ Berlin), Niedersachsen (+ Bremen), Bayern, and Nordrhein-Westfalen), the baseline SUMO workflow requires 121 minutes (Niedersachsen) to 1638 minutes (Bayern) to generate route files alone. In contrast, GROSS completes the route generation in 12 minutes (Brandenburg) and 49 minutes (Nordrhein-Westfalen). At national scale, this gap widens further: GROSS generates a Germany-wide routing in 190 minutes, whereas the baseline `gtfs2pt` route generation took 106 hours.



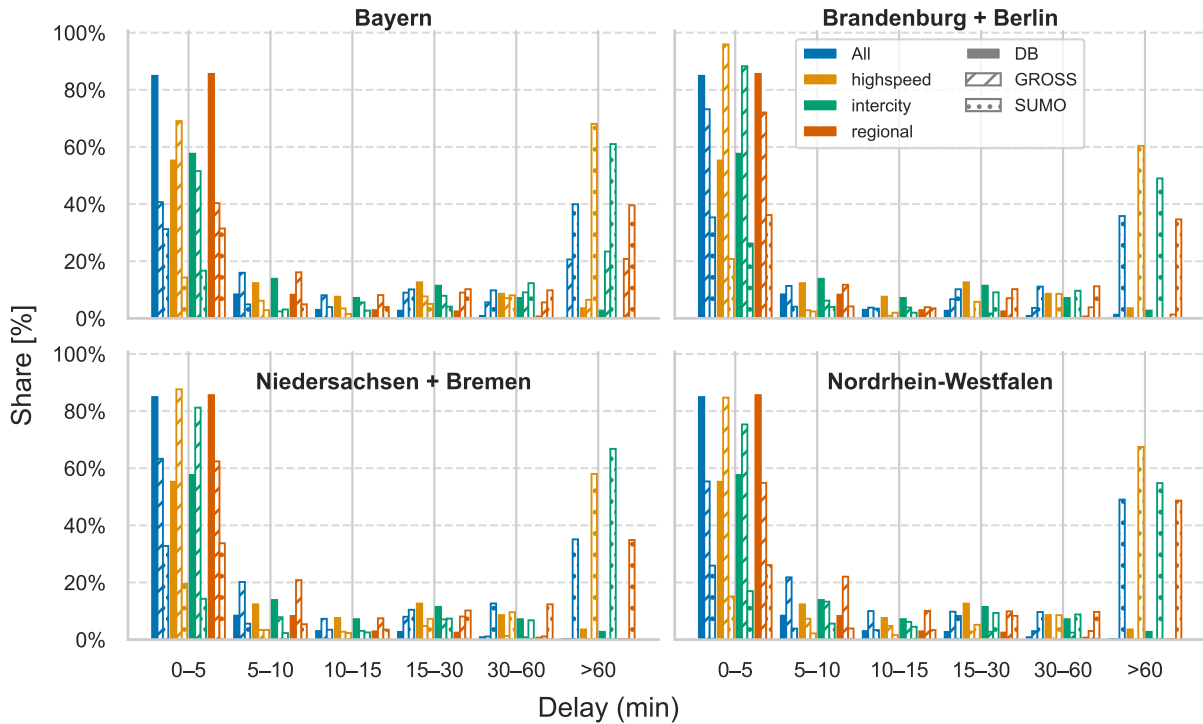
**Figure 5.** Regional-level comparison of vehicle delay distributions, grouped by planned route duration. The ticks show the median and 1st and 3rd quartiles of each distribution. Across all regions, GROSS yields a markedly smaller delay range. Rare outliers (<5% of trips) still occur, with the largest observed delays in Bayern.

#### 4.1 German sub-regions

A central challenge when scaling the vanilla SUMO rail pipeline is the prevalence of teleportation artifacts. To prevent extreme queueing and gridlock, SUMO can teleport vehicles that have been blocked at a junction for longer than a fixed threshold. While this mechanism introduces only minor distortions in road-traffic settings, it is highly unrealistic for rail operations because it breaks vehicle continuity and effectively bypasses capacity constraints. Thus we deliberately evaluate our scenarios with teleportation thresholds of 30 minutes.

Even with a conservative teleportation threshold of 30 minutes, the baseline scenarios exhibit frequent teleports (Tab. 1), which lead to extreme schedule deviations. In particular, the median delay in the baseline reaches 42–88 minutes (Fig. 5), reflecting compounding effects of interactions, missed meets/overtakes, and repeated teleport triggers. Nordrhein-Westfalen suffers from the worst delays, with half of the trips delayed by more than 88 minutes.

GROSS reduces teleports by a factor of 1.7–76.8× relative to SUMO and correspondingly shifts the delay distributions downward across all planned-duration bins (for spatial distribution, see Fig. A1). As expected, longer trips accumulate more variability because they traverse more junctions and are exposed to more opportunities for interaction-induced delay. However, in GROSS these delays remain bounded for the majority of trips (median of 2.5–10 minutes); only a small fraction of services (below 5%) exhibits large tail delays, and these are substantially less severe than in the baseline.

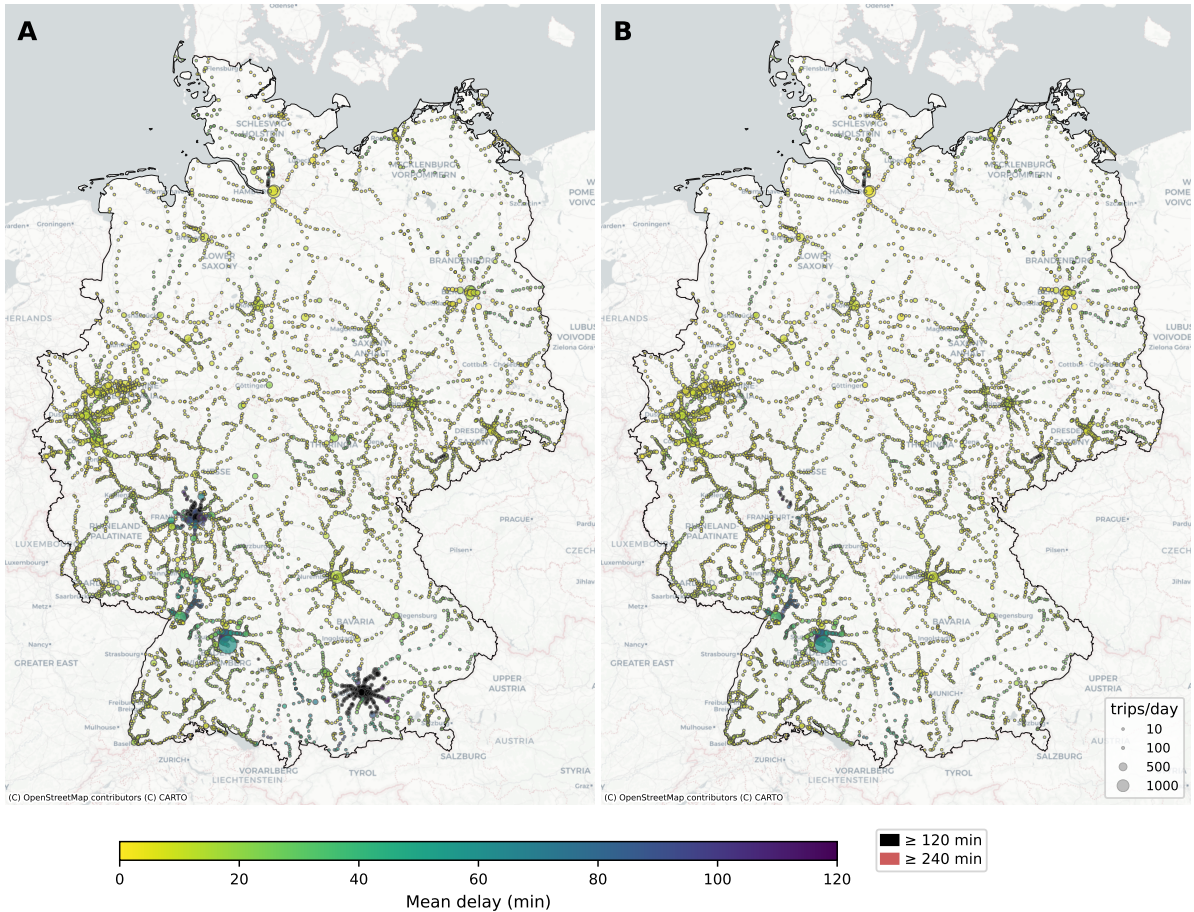


**Figure 6.** Share of vehicle-stops over different delay durations. Compared to station-level delay statistics from Deutsche Bahn (DB) at the national level, the GROSS scenario reproduces the overall shape of the distribution but still exhibits an over-representation of large delays. In contrast, the SUMO pipeline scenario produces a large amount of trains (>30%) that are delayed more than 60 minutes across the four regions.

The teleportation reduction compared to the baseline is primarily driven by GROSS route processing, which (i) reconciles timetable stop sequences with the rail topology, (ii) enforces consistent track usage at stations and along corridor segments to avoid spurious conflicts or deadlocks, and (iii) avoids large detours due to infeasible itineraries (e.g. going from  $A \rightarrow C \rightarrow B$  where  $C$  is an unnecessary non-stopping detour) introduced by noisy schedule-to-network matching. Together, these steps reduce artificial bottlenecks (e.g., multiple trains being forced through an incorrect edge or platform) and produce smoother headway patterns, lowering the probability that trains remain blocked long enough to trigger teleportation.

To contextualize the realism of these synthetic scenario, we compare simulated stop delays against real-world delay distributions reported by Deutsche Bahn (DB), the German national railway operator. We obtain station-level delay statistics for the period December 2025–January 2026 using the data collection scripts by <https://github.com/piebro/deutsche-bahn-data>. For the simulated scenario, we compute per-stop delay as the difference between planned and realized arrival times at each stop, and aggregate delays into the same bins used by the DB statistics.

With this, we are now able to compare the performance of SUMO and GROSS side-by-side in context with the train statistics from DB. In Fig. 6, we see that SUMO suffers from a lot of routes having more than 60 minutes delays. For the region of Bayern, we see that, although it performs better than the SUMO pipeline, our GROSS pipeline has some trouble with the regional train routes, with around 21% having more than 60 minutes delays. We suspect that this is due to the station topology of München Hbf.



**Figure 7.** Mean delay (in minutes) observed at each station cluster for Germany. The sizes of the bubbles also show the number of train routes per day passing through the station cluster. (A) All trips included in analysis. (B) Trips with stops passing through München and Frankfurt am Main filtered out.

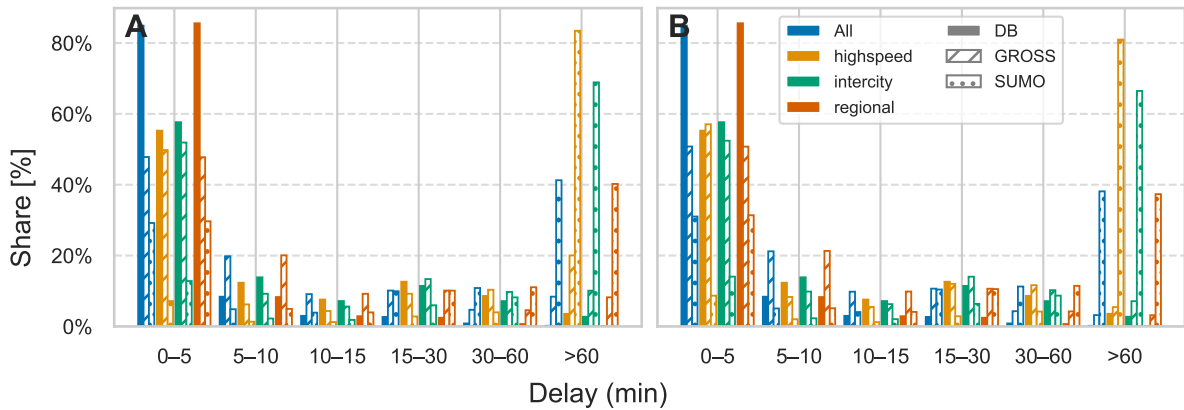
However, apart from Bayern, the other regions perform much better than the SUMO-generated scenarios, albeit still below the real-world statistics of DB.

#### 4.2 Germany Train Scenario

The performance and robustness improvements of GROSS enable us to generate and simulate a Germany-wide rail scenario in SUMO. After running the GROSS pipeline for Germany, we observe that stations in certain regions have disproportionately large delays. These can be seen in the cities of München and Frankfurt am Main (Fig. 7A).

We hypothesize that some of these large delays are linked to challenging local layouts, such as terminating stations (where trains cannot pass through) with a limited through-track capacity. We suspect that this happens as trips get routed via through-tracks that would normally stop at a terminal platform due to shorter connecting tracks.

Instead, we try to see whether this is an isolated effect by excluding all trips that pass through these cities from our analysis. That is, we still simulated Germany with all trips, but filter out this subset of trips (around 15% of stop events) from the analysis (Fig. 7B). Although we show that GROSS substantially reduces unrealistically extreme delay behavior compared to the vanilla SUMO pipeline, Fig. 8 nevertheless shows a delay distribution still produces too many high-delay events compared to the DB profile. In our analysis, these tail delays are often associated with local deadlocks (e.g.,



**Figure 8.** Share of vehicle-stops over different delay durations. We show the stop-level statistics for (A) all trips included in analysis, and (B) trips with stops passing through München and Frankfurt am Main filtered out. Filtering out the trips affects mostly the trips with delays exceeding 60 minutes.

opposing trains meeting on a single-track segment without a feasible resolve) and can cascade to downstream services when headways are tight. Addressing this likely requires more detailed modeling of operational rules and infrastructure constraints (e.g., passing-loop usage, dispatching/priority policies, and single-track conflict resolution), which we leave for future work.

## 5. Discussion

GROSS tackles two recurring practical obstacles when building large-scale rail scenarios in SUMO: long preprocessing times and simulations that become unstable due to routing errors and the resulting teleportation artifacts. By combining topology-aware route construction with systematic validation and targeted repair, the pipeline produces scenarios that are faster to generate and operationally more plausible. This is reflected in substantially reduced teleport rates and a tighter delay distribution, suggesting that a large fraction of previously observed “instability” was caused by inconsistent inputs and brittle routing rather than by unavoidable effects of scale.

The results also clarify what GROSS does and does not solve. The pipeline improves the consistency between the timetable and the network representation, but it cannot compensate for missing information in the underlying data sources. GTFS feeds are primarily meant for passenger-facing schedule dissemination and often omit infrastructure- and operations-relevant detail such as platform-track assignments, passing loop usage, rolling-stock constraints, or operator-specific dispatching rules. Consequently, even after cleaning and repair, the simulated system is best interpreted as an operationally feasible but simplified realization of the published service plan.

Several remaining failure modes are also tied to limitations of the SUMO rail model and the configuration used in this work. In particular, the current setup does not fully capture the richness of railway operations, including explicit timetabled meets on single-track, priority- and rule-based dispatching, and recovery margins that buffer small perturbations. These modeling gaps can manifest as unrealistic deadlocks or exaggerated tail delays in specific network regions. The most prominent issues observed in our experiments were:

- **Coupling and splitting:** coupled trains are represented as independent vehicles, which can amplify delays when they traverse long blocks and compete for scarce capacity.
- **Single-track conflict resolution:** on single-track sections with multiple uncoordinated signal junctions, local look-ahead can admit trains from both ends and trigger deadlocks.
- **Platform assignment and re-planning:** trains are not re-routed online to alternative platforms, so temporary platform conflicts can cascade into system-wide delay propagation.
- **Residual schedule noise:** mistagged or malformed GTFS trips may still be imported and then routed in implausible ways (e.g., repeated reversals), requiring detection and filtering via exception lists.

Despite these limitations, the evaluation demonstrates that the approach scales from subregional deployments to a nation-scale scenario without changing the conceptual workflow. Beyond Germany, GROSS is directly applicable wherever (i) a rail network can be extracted (e.g., from OpenStreetMap) and (ii) service schedules are available in GTFS or can be converted to it. The GTFS standard is widely available from different agencies or at sites like the Mobility Database [24]. Alternatively, EU members are mandated to provide a national access point for NeTeX feeds [25] that can be converted to GTFS using external tools [26].

Although we focused on a rail-only setup, the generated rail network is also intentionally composable with existing SUMO scenarios: it can be merged with city- or region-scale road and active-mobility networks via `netconvert` (using the arguments `-s train.net.xml,notrain.net.xml` to merge networks). This means that GROSS can complement other preexisting, more car- and bus-focused scenarios for building multi-modal simulations in which rail interacts with buses, private vehicles, and pedestrians, thereby lowering the barrier to constructing more holistic digital twins and enabling comparative studies under consistent modeling assumptions.

In this work we found that there is no consistent tagging in regards to how we distinguished light-rail and heavy-rail as heavy-rail being all systems that are interoperable with the national network. Thus we see advantages in simulating all rail traffic (tram, light-rail, heavy-rail, narrow-gauge) to reduce vulnerability to tagging inconsistencies and simplify combined systems (e.g. S-Bahn Karlsruhe operating on the tram, and rail network).

Looking forward, two directions appear most important for closing the remaining realism gap. First, robust station/stop matching across heterogeneous identifiers (e.g., GTFS stop IDs, operator station codes, and external delay statistics) would benefit from standardized mappings and automated alignment routines. Second, reducing the high-delay tail will likely require incorporating additional operational constraints, dispatching priorities, and control logic, especially for single-track sections.

## 6. Conclusion and Future Work

We introduced GROSS, a preprocessing pipeline for generating large-scale SUMO rail scenarios that is both faster than the vanilla SUMO toolchain and substantially more

stable in simulation. Across multiple German regions, GROSS reduces teleportation events and produces significantly lower and less variable delay distributions. At a national scale, the pipeline enables end-to-end generation of a Germany-wide scenario and supports comparison against operator-reported delay statistics while highlighting remaining gaps in the high-delay tail.

Future work will focus on (i) improving data alignment and coverage (e.g., robust station/stop matching and richer infrastructure metadata when available), (ii) improving operational realism through enhanced dispatching and single-track conflict handling, and (iii) tighter integration with multi-modal, city-scale digital twins by merging GROSS rail networks with existing SUMO road and public-transport scenarios using `netconvert`. Together, these directions move toward scalable, transferable, and multi-modal simulation environments that can better support urban planning and policy analysis across diverse geographies.

## Data availability statement

All code used to generate the resulting scenarios in the work will be made available at <https://github.com/ethz-coss/GROSS>.

## Author contributions

J.P.: Conceptualization; methodology; software; data curation; formal analysis; validation; visualization; investigation; writing—original draft; writing—review and editing.

D. D.: Conceptualization; supervision; methodology; software; formal analysis; validation; visualization; investigation; writing—original draft; writing—review and editing.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

J.P. and D.D. would like to acknowledge Prof. Dirk Helbing for his support and insightful feedback.

## References

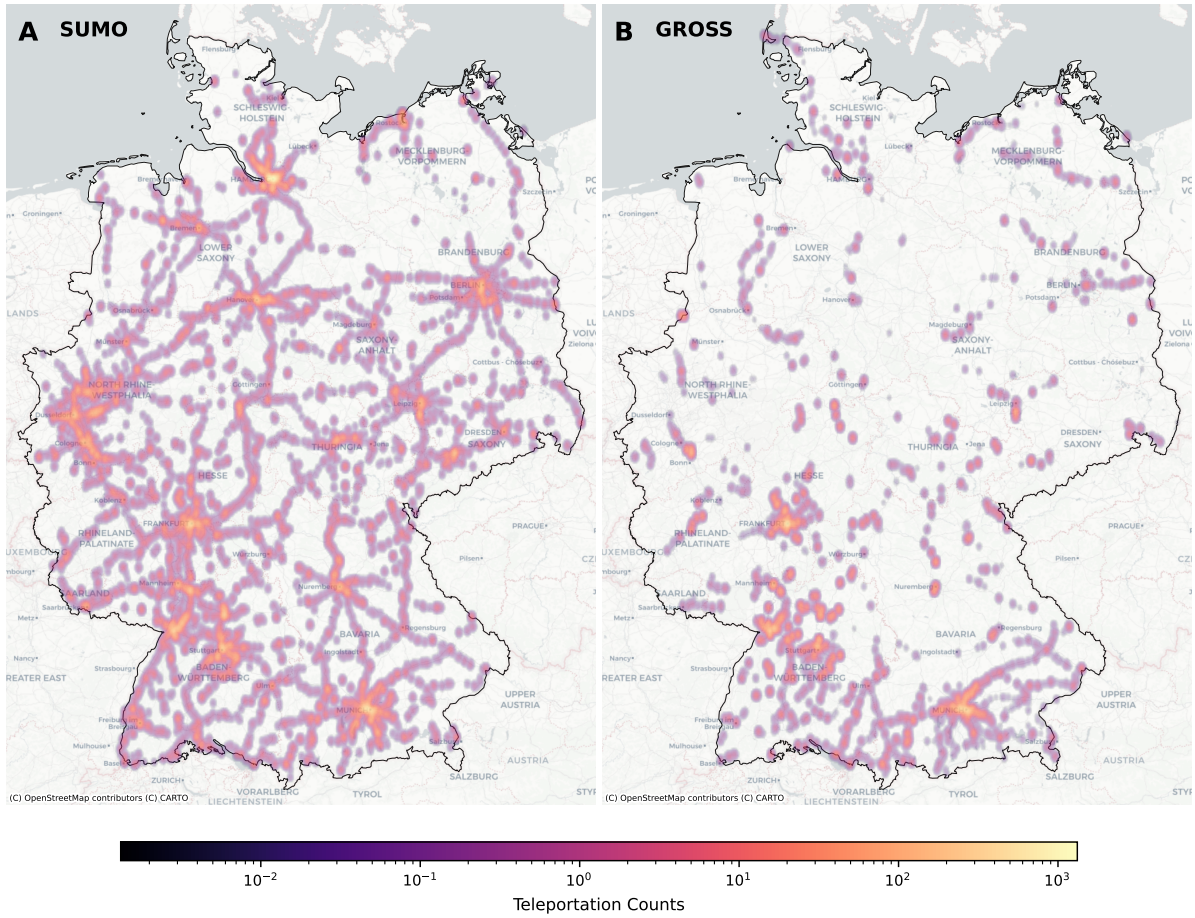
- [1] P. A. Lopez, M. Behrisch, L. Bieker-Walz, *et al.*, “Microscopic Traffic Simulation using SUMO,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2018, pp. 2575–2582. DOI: [10.1109/ITSC.2018.8569938](https://doi.org/10.1109/ITSC.2018.8569938). (visited on 02/09/2026).
- [2] L. Bieker, D. Krajzewicz, A. Morra, C. Michelacci, and F. Cartolano, “Traffic simulation for all: A real world traffic scenario from the city of bologna,” in *Modeling Mobility with Open Data*. Springer International Publishing, 2015, pp. 47–60, ISBN: 9783319150246. DOI: [10.1007/978-3-319-15024-6\\_4](https://doi.org/10.1007/978-3-319-15024-6_4).

- [3] L. Codeca, R. Frank, S. Faye, and T. Engel, "Luxembourg sumo traffic (lust) scenario: Traffic demand evaluation," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 2, pp. 52–63, 2017, ISSN: 1939-1390. DOI: [10.1109/mits.2017.2666585](https://doi.org/10.1109/mits.2017.2666585). [Online]. Available: <http://dx.doi.org/10.1109/MITS.2017.2666585>.
- [4] L. Codeca and J. Harri, "Towards multimodal mobility simulation of c-its: The monaco sumo traffic scenario," in *2017 IEEE Vehicular Networking Conference (VNC)*, IEEE, Nov. 2017, pp. 97–100. DOI: [10.1109/vnc.2017.8275627](https://doi.org/10.1109/vnc.2017.8275627). [Online]. Available: <http://dx.doi.org/10.1109/VNC.2017.8275627>.
- [5] K. Schrab, R. Protzmann, and I. Radusch, "A large-scale traffic scenario of berlin for evaluating smart mobility applications," in *Smart Energy for Smart Transport*. Springer Nature Switzerland, 2023, pp. 276–287, ISBN: 9783031237218. DOI: [10.1007/978-3-031-23721-8\\_24](https://doi.org/10.1007/978-3-031-23721-8_24). [Online]. Available: [http://dx.doi.org/10.1007/978-3-031-23721-8\\_24](http://dx.doi.org/10.1007/978-3-031-23721-8_24).
- [6] Y. Yamazaki, Y. Tamura, X. Défago, E. Javanmardi, and M. Tsukada, "Tost: Tokyo sumo traffic scenario," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, Sep. 2023, pp. 3597–3604. DOI: [10.1109/itsc57777.2023.10422517](https://doi.org/10.1109/itsc57777.2023.10422517). [Online]. Available: <http://dx.doi.org/10.1109/ITSC57777.2023.10422517>.
- [7] J. Argota Sánchez-Vaquerizo, "Getting Real: The Challenge of Building and Validating a Large-Scale Digital Twin of Barcelona's Traffic with Empirical Data," *ISPRS International Journal of Geo-Information*, vol. 11, no. 1, p. 24, Dec. 2021, ISSN: 2220-9964. DOI: [10.3390/ijgi11010024](https://doi.org/10.3390/ijgi11010024). (visited on 02/18/2026).
- [8] L. N. Alegre, *SUMO-RL*, <https://github.com/LucasAlegre/sumo-rl>, 2019.
- [9] M. Korecki, D. Dailisan, and D. Helbing, "How well do reinforcement learning approaches cope with disruptions? the case of traffic signal control," *IEEE Access*, vol. 11, pp. 36 504–36 515, 2023. DOI: [10.1109/ACCESS.2023.3266644](https://doi.org/10.1109/ACCESS.2023.3266644).
- [10] M. Korecki, D. Dailisan, J. Yang, and D. Helbing, "Democratizing traffic control in smart cities," *Transportation Research Part C: Emerging Technologies*, vol. 160, p. 104 511, Mar. 2024, ISSN: 0968-090X. DOI: [10.1016/j.trc.2024.104511](https://doi.org/10.1016/j.trc.2024.104511).
- [11] R. K. Dubey, D. Dailisan, J. Argota Sánchez-Vaquerizo, and D. Helbing, "Fairlane: A multi-agent approach to priority lane management in diverse traffic composition," *Transportation Research Part C: Emerging Technologies*, vol. 171, p. 104 919, Feb. 2025, ISSN: 0968-090X. DOI: [10.1016/j.trc.2024.104919](https://doi.org/10.1016/j.trc.2024.104919). [Online]. Available: <http://dx.doi.org/10.1016/j.trc.2024.104919>.
- [12] R. Buehler, J. Pucher, R. Gerike, and T. Götschi, "Reducing car dependence in the heart of Europe: Lessons from Germany, Austria, and Switzerland," *Transport Reviews*, vol. 37, no. 1, pp. 4–28, Jan. 2017, ISSN: 0144-1647. DOI: [10.1080/01441647.2016.1177799](https://doi.org/10.1080/01441647.2016.1177799). (visited on 02/02/2026).
- [13] D. F. Heuermann and J. F. Schmieder, "The effect of infrastructure on worker mobility: Evidence from high-speed rail expansion in Germany," *Journal of Economic Geography*, vol. 19, no. 2, pp. 335–372, Mar. 2019, ISSN: 1468-2702. DOI: [10.1093/jeg/lby019](https://doi.org/10.1093/jeg/lby019). (visited on 02/02/2026).
- [14] M. Boehm, M. Arnz, and J. Winter, "The potential of high-speed rail freight in Europe: How is a modal shift from road to rail possible for low-density high value cargo?" *European Transport Research Review*, vol. 13, no. 1, p. 4, Jan. 2021, ISSN: 1866-8887. DOI: [10.1186/s12544-020-00453-3](https://doi.org/10.1186/s12544-020-00453-3). (visited on 02/02/2026).
- [15] S. Daniotti, V. D. P. Servedio, J. Kager, A. Robben-Baldauf, and S. Thurner, "Systemic risk approach to mitigate delay cascading in railway networks," *npj Sustainable Mobility and Transport*, vol. 1, no. 1, p. 15, Dec. 2024, ISSN: 3004-8664. DOI: [10.1038/s44333-024-00012-6](https://doi.org/10.1038/s44333-024-00012-6). (visited on 02/09/2026).
- [16] O. Cats and E. Jenelius, "Beyond a complete failure: The impact of partial capacity degradation on public transport network vulnerability," *Transportmetrica B: Transport Dynamics*, vol. 6, no. 2, pp. 77–96, Dec. 2016, ISSN: 2168-0582. DOI: [10.1080/21680566.2016.1267596](https://doi.org/10.1080/21680566.2016.1267596). [Online]. Available: <http://dx.doi.org/10.1080/21680566.2016.1267596>.

- [17] T. Dewilde, P. Sels, D. Cattrysse, and P. Vansteenwegen, "Defining robustness of a railway timetable," in *4th International Seminar on Railway Operations Modelling and Analysis (IAROR) : RailRome 2011, Proceedings*, Ghent University, Department of Industrial Management, 2011. HDL: [1854/LU-1163245](https://hdl.handle.net/1854/LU-1163245). (visited on 02/09/2026).
- [18] OpenStreetMap contributors, *Planet dump* retrieved from <https://download.geofabrik.de>, <https://www.openstreetmap.org>, 2026.
- [19] DELFI e.V., *DELFI-Archive*, [https://archiv.opendata-oeprv.de/VRR/DELFI/Soell-Fahrplandaten%20\(GTFS\)/2025/](https://archiv.opendata-oeprv.de/VRR/DELFI/Soell-Fahrplandaten%20(GTFS)/2025/). (visited on 02/19/2026).
- [20] P. Alvarez Lopez, A. Banse, M. Behrisch, *et al.*, *Simulation of urban mobility (sumo)*, version 1.26.0, Jan. 2026. DOI: [10.5281/zenodo.18406080](https://doi.org/10.5281/zenodo.18406080). [Online]. Available: <https://doi.org/10.5281/zenodo.18406080>.
- [21] T. Litman, *Towards More Comprehensive and Multi-Modal Transport Evaluation*. Victoria Transport Policy Institute, 2013.
- [22] DLR-TS, *sumo-berlin: A SUMO scenario of Berlin*, <https://github.com/DLR-TS/sumo-berlin>, 2026.
- [23] OpenStreetMap Wiki contributors. "Forward & backward, left & right - OpenStreetMap Wiki." (), [Online]. Available: [https://wiki.openstreetmap.org/wiki/Forward\\_%26\\_backward,\\_left\\_%26\\_right#Identifying\\_the\\_direction\\_of\\_a\\_way.27](https://wiki.openstreetmap.org/wiki/Forward_%26_backward,_left_%26_right#Identifying_the_direction_of_a_way.27) (visited on 04/15/2026).
- [24] Mobility Database. "Mobility database," Mobility Database. (), [Online]. Available: <https://mobilitydatabase.org> (visited on 04/15/2026).
- [25] European Commission. "National Access Points — Mobility and Transport — European Commission." (), [Online]. Available: [https://transport.ec.europa.eu/transport-themes/smart-mobility/road/its-directive-and-action-plan/national-access-points\\_en](https://transport.ec.europa.eu/transport-themes/smart-mobility/road/its-directive-and-action-plan/national-access-points_en) (visited on 04/15/2026).
- [26] Entur AS, *Entur/netex-gtfs-converter-java*, Apr. 10, 2026. [Online]. Available: <https://github.com/entur/netex-gtfs-converter-java> (visited on 04/15/2026).

# Appendix

## Teleportation heatmap



**Figure A1.** Spatial distributions of teleportation between SUMO (left) and GROSS (right) across Germany. There are far fewer teleportations for GROSS except for the regions of Karlsruhe, München, and Frankfurt am Main.

## Processing Time Table

**Table A1.** Processing Time to generate a day of rail traffic. For GROSS split into base cost and additional cost per simulation day.

Region	Processing Time (in minutes)		
	SUMO Per Day	GROSS Base	GROSS Per Day
Bayern	1638	30	2
Brandenburg	222	8	1
Niedersachsen	121	15	1
Nordrhein-Westfalen	941	36	2
Germany	6360	369	10

## Repair Table

**Table A2.** Number of stops and trips affected by route repair (averaged across seeds and dates in Germany). Some routes are excluded due to manual review.

Repair Fix	SUMO (%)	GROSS (%)	
	Trips	Stops	Trips
Single-Stop	—	0.038	3.2
Missing Stop	—	3.1	16
Speed	1.4	—	0.0

PRE-PRINT VERSION