Toward Understanding the Dynamics of Commercial FCD Penetration Rates

An Empirical Study in Frankfurt am Main

Amir Babaei,
Prof. Dr. Tobias Hagen

World Conference of Transport Research (WCTR), Toulouse 2026

Frankfurt University of Applied Sciences
amir.babaei@ieee.org

Motivation: The Role of Floating Car Data (FCD)

Floating Car Data (FCD) is indispensable for Intelligent Transportation Systems (ITS) [1].

Applications:

- Real-time incident detection [2]
- Travel time prediction [3]
- Environmental emissions modeling [4]

The Challenge:

- FCD provides speed and travel time reliably.
- However, **vehicle counts are inherently incomplete**. FCD is a non-random sample.

The Scaling Problem

Estimating total network volumes from FCD requires a crucial scaling factor: the **Penetration Rate (PR)**.

The Penetration Rate (PR) Problem

$$PR(l,t) = \frac{\text{FCD Vehicles}(l,t)}{\text{Total Vehicles}(l,t)}$$

Why is PR problematic?

- Often treated as a **single, static value**.
- Low PR can cause "phantom jams" (Video) [5].
- PR is known to vary by time, space, and congestion, but the *full extent of its dynamics* and the *underlying factors* are rarely quantified at scale.

Research Objective

To empirically quantify PR dynamics at scale and systematically investigate the factors that influence its variance.

Study Area: Frankfurt am Main

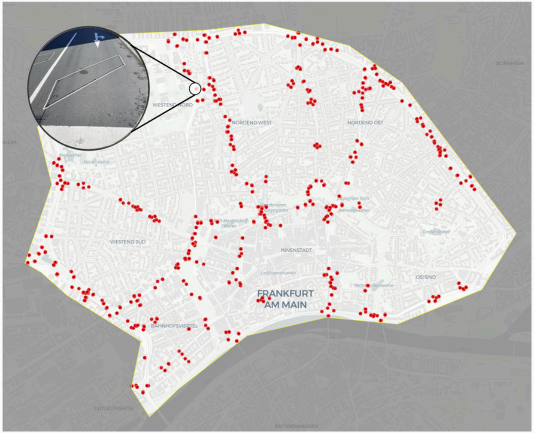


Figure 1: Study area and loop detector locations - map data: OpenStreetMap contributors

Context:

- Major financial and transport hub.
- Dense and heterogeneous network.
- Study focused on the city center (Alleenring).

Infrastructure:

- Highly instrumented with over **500 active loop detectors** serving as ground truth.
- Evaluated over a 4-month period (Nov 2024 - Feb 2025).

Data Sources & Preprocessing

Loop Detector Data (LDD) - (N_{LDD})

- High-resolution, lane-specific.
- Minute-by-minute flow (q) and occupancy (o).
- Preprocessing: Flagged invalid physical constraints, imputed small gaps, aggregated to hourly bins.

Floating Car Data (FCD) - (N_{FCD})

- $\sim$10 Million trips from 1.8M unique devices.
- 81.6% Consumer / 18.4% Fleet.
- Preprocessing: Mapped to OpenStreetMap (OSM) using the Valhalla routing engine.

Figure 2: Raw FCD (sample)trajectories vs. map-matched (daily)volume - map data: OpenStreetMap contributors

Methodology: Data Fusion

To calculate PR, we established a common spatial reference by mapping physical detectors to OSM ways and using **Virtual Trip Lines (VTL)**.

1. **Spatial Mapping:** Group directional lane-detectors to OSM `way_ids`.
2. **FCD Counting:** Count unique probe trips traversing the specific OSM segment.
3. **Directionality:** Applied a hierarchical algorithm (Topological $\rightarrow$ Geometric Vector $\rightarrow$ Proxy Vector) to resolve bidirectional roads.

Resulting Dataset

A spatio-temporal panel dataset spanning $N = 236$ locations over $T = 168$ time periods (hours of the week), encompassing PR, traffic state (occupancy), and OSM attributes.

Descriptive Analysis: Overall Distribution

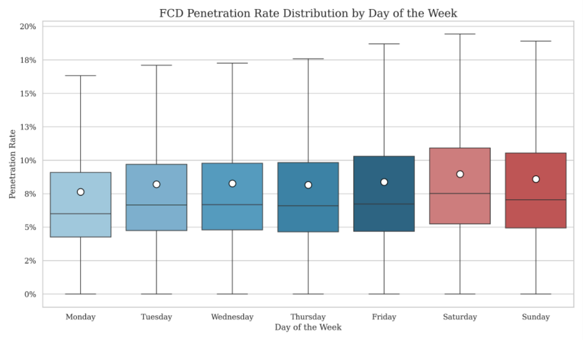


Figure 3: Distribution of PR by Day of the Week.

Key Statistics:

- **Mean PR:** 8.31%
- **Median PR:** 6.74%
- **Coefficient of Variation:** 0.75

Observation

The PR is highly dynamic and positively skewed. A static 8.3% assumption would yield massive errors. Variance is split almost equally between temporal (47%) and spatial (53%) dimensions.

Descriptive Analysis: Temporal Dynamics

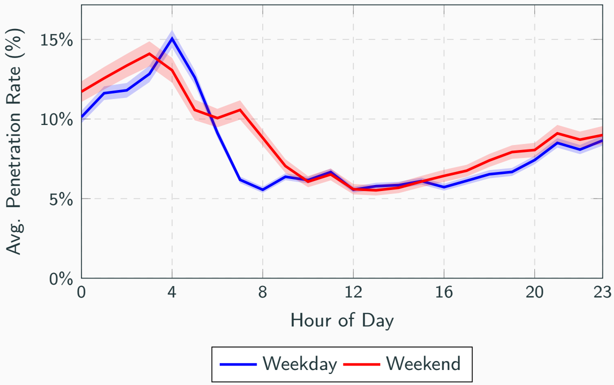


Figure 4: Average hourly PR: Typical weekday vs. weekend.

Diurnal Pattern:

- PR is **inversely related** to typical traffic demand.
- **Highest:** Late-night/early-morning ($> 15\%$).
- **Lowest:** Morning commute ($\approx 5.6\%$).

Weekly Pattern:

- Weekends exhibit significantly higher PR across daytime hours.

Descriptive Analysis: Spatial Variation

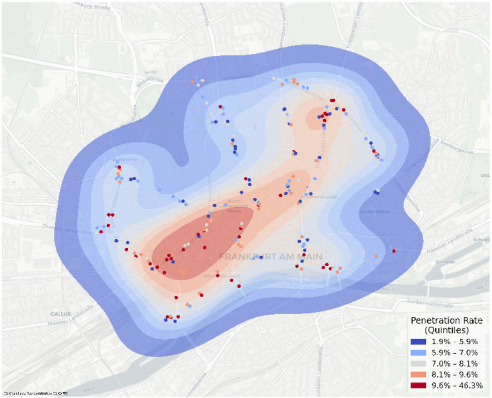


Figure 5: Spatial distribution of FCD Penetration Rate - map data: OpenStreetMap contributors

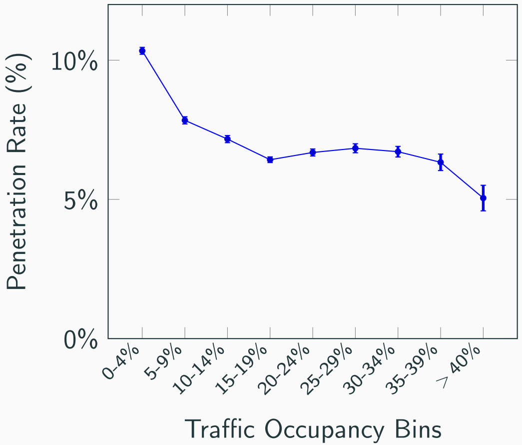
Non-uniform Distribution:

- 5th percentile PR: 3.1%
- 95th percentile PR: 21.5%

Identified Hotspots:

- High PR prominently clustered around the **central train station** (Hauptbahnhof) in the southwest.
- Suggests a strong link between PR and urban functional areas (e.g., taxi fleets serving transit hubs).

Descriptive Analysis: The Impact of Congestion



Traffic State Dependency:

- Free-flow conditions (0–15% occ.): PR averages ∼9.0%.
- Heavy congestion (> 40% occ.): PR plummets to ∼5.0%.

The "Taxi Effect"

Professional drivers and commercial fleets actively avoid congested routes. This route preference inherently biases the sample during peak hours.

Regression Analysis: Addressing Panel Data Challenges

Standard regression (OLS) fails on panel data because it ignores the complex structure of time and space. We evaluated PR across 236 locations over 168 hours.

Key Challenges & Our Solutions:

- **Unobserved Factors:** We used a **Random Effects (RE)** model to account for hidden differences between locations over time.
- **Spatial Spillovers (Ripple Effects):** Traffic at one intersection affects its neighbors. We used a **Spatial (SARAR)** model to capture this dependency.
- **Local Differences:** Congestion impacts different roads differently. We used a **Common Correlated Effects (CCE)** model to calculate unique estimates for every single location.

Regression Results: What Drives Penetration Rates?

Key Significant Findings (Holding other variables constant):

- **Fleet Composition:** A higher proportion of light vehicles (vs. trucks/vans) significantly increases the PR.
- **Congestion (-):** As detector occupancy increases, PR systematically decreases. Confirms the strategic avoidance behavior of commercial FCD drivers.
- **Time Effects (+/-):** Strongest impact. Compared to Monday nights, PR drops nearly 5 percentage points during daytime peak hours. Weekends add roughly 1 percentage point globally.

Spatial Spillovers Exist

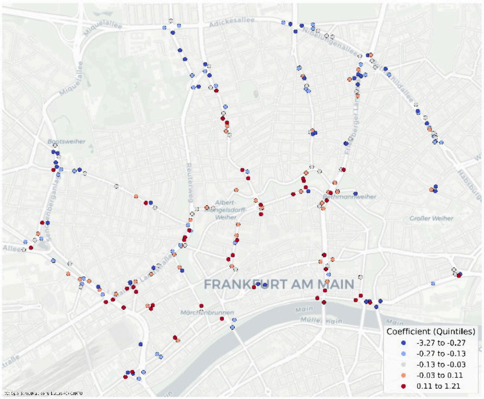
The spatial model (SARAR) confirmed significant positive feedback: an increase in PR at one location positively impacts PR at neighboring locations.

Regression Results: Spatial and Built Environment

Time-Constant Spatial Variables:

- **Proximity to Transit (+):** Distance to the main train station is *negatively correlated* with PR. A 10% decrease in distance yields a 0.26 pp increase in PR.
- **Business Districts (-):** Higher density of office buildings is associated with a *lower* PR.

Interpretation: The commercial FCD fleet is heavily influenced by specific trip purposes (taxis near hubs) and does not adequately represent standard commuter flows in business districts.



Location-specific coefficients for occupancy (CCE model). Red denotes stronger negative impact of congestion on PR near the city center - map data: OpenStreetMap contributors

Conclusion

1. **PR is Not Static:** Treating Penetration Rate as a constant leads to massive scaling errors. It exhibits a CV of 0.75 across space and time.

2. **Congestion Bias:** PR is inversely related to congestion. Commercial fleets actively avoid traffic, meaning FCD represents the smallest fraction of total traffic exactly when ITS applications need data the most.

3. **Spatial Heterogeneity:** Built environment matters. PR surges near major transit hubs and drops in commuter-heavy business districts.

Limitations & Future Work

Limitations:

- Regression relies on a linear framework, potentially missing complex non-linear interactions.
- Inverse-distance spatial weights do not fully capture topological constraints and directional traffic flows.
- Single commercial FCD source in one specific urban layout.

Future Directions:

- Transitioning to Graph Neural Networks (GNNs) to model complex, non-linear spatio-temporal dependencies.
- The ultimate goal: Develop a dynamic, network-wide PR imputation model to recover highly accurate traffic volumes directly from sparse FCD.

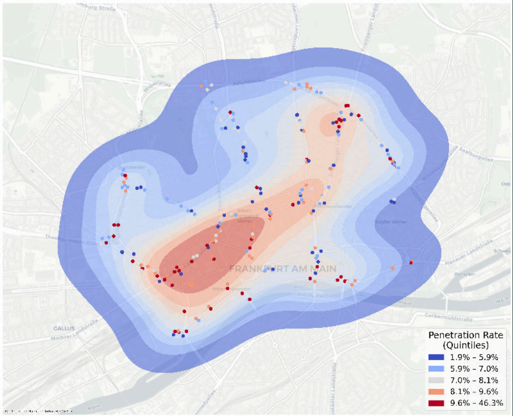
Thank You! Any questions?

Amir Babaei

Frankfurt University of Applied Sciences
Nibelungenplatz 1, 60318 Frankfurt am Main, Germany

Email: amir.babaei@ieee.org
Phone: +49 69 1533-2997

LinkedIn: amirbabaei97



Spatial distribution of FCD Penetration Rate - map data: OpenStreetMap contributors

Backup: Descriptive Statistics of Key Variables

Variable	Mean	Std. Dev.	Min.	Median	Max.
Penetration Rate (%)	8.31	6.24	0.00	6.74	81.02
Occupancy (%)	10.49	9.13	0.00	8.04	49.69
Light Vehicle Proportion	0.87	0.16	0.00	0.92	1.00
Speed Limit (km/h)	45.30	7.61	30.00	50.00	50.00
Dist. to Main Station (km)	1.90	0.96	0.17	1.98	3.35
Gastronomy (in 200m)	15.66	12.83	0.00	14.00	53.00
Retail Stores (in 200m)	18.61	25.35	0.00	10.00	158.00
Offices (in 200m)	6.08	6.52	0.00	4.00	30.00

N = 39,648 observations (236 locations over 168 hours)

Backup: SARAR Model Formulation

The formula for the Spatial Autoregressive Model with Autoregressive Errors (SARAR) is:

$$PR_{it} = \rho \sum_{j=1}^{N} w_{ij} PR_{jt} + \mathrm{X}_{it}\beta + \mu_i + u_{it}$$

With the spatially lagged error term:

$$u_{it} = \lambda \sum_{j=1}^{N} w_{ij} u_{jt} + \varepsilon_{it}$$

Where:

- ρ is the spatial autoregressive coefficient (spillovers).
- λ is the spatial error coefficient.
- w_{ij} are the elements of the spatial weights matrix W (inverse-distance).
- μ_i is the location-specific random effect.

Backup: Model Features (Independent Variables)

Time-Varying Variables:

- Sensor Occupancy (%)
- Proportion of Light Vehicle Trips (%)

Time Fixed Effects:

- Day of the Week (Base: Monday)
- Hour of Day (Base: 0:00–1:00)

Time-Constant Spatial Variables (from OpenStreetMap):

- **Road Infrastructure:** More than one lane (dummy), Speed limit < 50 km/h (dummy)
- **Proximity:** ln(Distance to main train station)
- **POIs (Count within 200m):** Parking Garages, Hotels, Gastronomy, Office Buildings, and Taxi Stands

Backup: Map-Matching Directionality Algorithm

To accurately map FCD trajectories to bidirectional road segments, we employed a 3-tier algorithm:

1. **Tier 1 (Topological Analysis):** Checks if the previous or next traversed way connects to the start or end node of the target way.
2. **Tier 2 (Geometric Vector Analysis):** If Tier 1 fails, calculates the dot product between the vehicle's movement vector (using multiple GPS points) and the way's direction vector.
3. **Tier 3 (Proxy Vector Analysis):** For short segments with only one GPS point, uses the nearest adjacent GPS point to create a proxy vector for the dot product.