Useful Data, Not Just More Data: Cost-Aware Floating Car Data Generation in Large-Scale Traffic Simulation

VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER

Paulo Silva, Pavlína Smolková, Kateřina Slaninová, Jan Martinovič (IT4Innovations, VSB – Technical University of Ostrava)

Large-Scale Traffic Simulation

- Traffic simulation can generate millions of Floating Car Data (FCD) records.
- More data does not always mean more useful information.
- FCD output has a cost: storage, memory, I/O and processing.
- We seek useful traffic information without unnecessary output.

How much traffic detail is enough?

FCD records vehicle positions and speeds over time.
Recording less often creates smaller files.

What is retained when output is reduced?

We compare smaller-output settings with a detailed simulation reference to measure retained road coverage, temporal coverage, and speed fidelity.

A useful FCD setting is the smallest output that still preserves the traffic information needed for the analysis.

Floating Car Data (FCD) Generation

GitHub: Ruth Simulator

- F: configured vehicle update period.
- S: configured FCD sampling period.
- Rerouting participation: 80% vs 50%.
- Simulator: Ruth Traffic Simulator on LUMI-C.

Chicago Experiment

- Input: Chicago 2019 AM/PM demand.

6,897,778 input trips
748,509 nodes
1,867,225 edges

F20/S5 = 20 s vehicle-update period; 5 s FCD-sampling period.

Campaign	Setup	Main output
FCD comparison	4 nodes, 8 h, F/S settings	Cost vs usefulness
Scaling preview	1–16 nodes, 1 h	FCD temporal span, CPU busy
Rerouting test	50% vs 80%, 4 nodes, 8 h	Progress and wall cost

Road-edge coverage: reference road segments also observed.
Temporal coverage: same roads in the same 60-second intervals.
Speed correlation: mean-speed agreement in shared intervals.
FCD span: earliest to latest saved FCD timestamps.

Methodology

Fixed simulation input
Chicago demand + cached map

↓

Temporal FCD + rerouting configurations
F20/S5, F20/S10, F20/S20, ...
Rerouting share: 80% vs 50%

↓

Simulation campaigns
FCD comparison + scaling preview + rerouting test

↓

Useful-data balance
Resource cost + retained traffic information

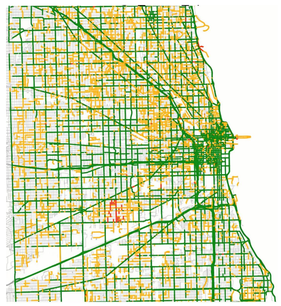
What we compare:
Resource cost + useful traffic information

Resource cost
Wall time
Wall s / sim s
Peak memory
HDF5 size
FCD records

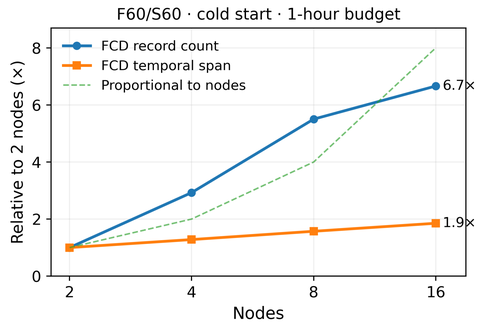
+

Useful traffic information
FCD density
Road-edge coverage
Temporal coverage
Speed-profile similarity

- Comparable runs reuse Chicago demand and a cached map.
- Varying S tests output sampling.
- Changing F or rerouting can change traffic.



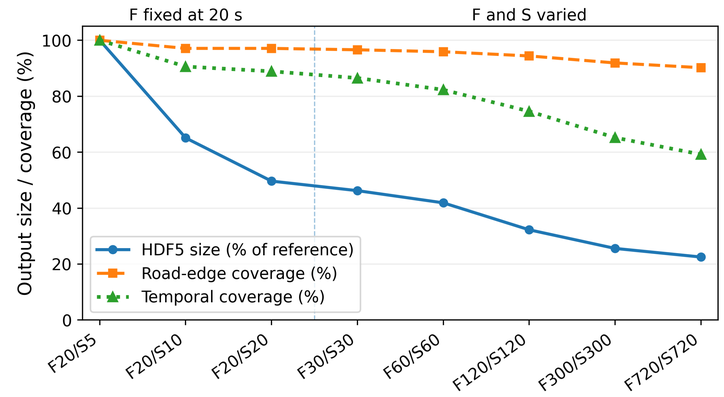
Scaling Preview



- Baseline: 2 nodes, with a one-hour budget.
- 2 → 16 nodes: 6.7× records, but only 1.9× FCD span.
- 1 node: initial routes unfinished; no FCD emitted.

FCD Output vs Usefulness

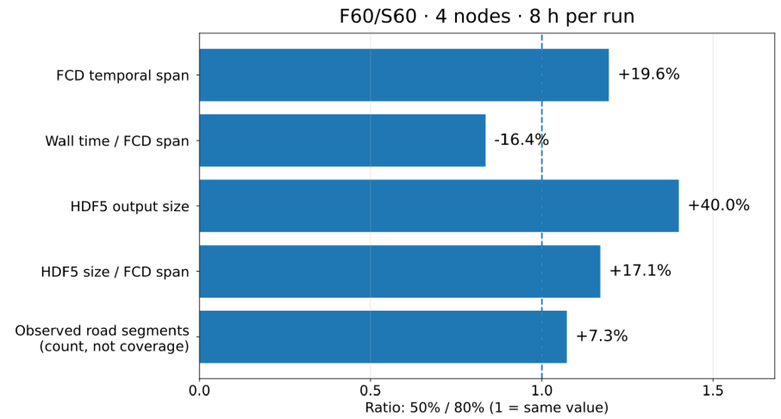
- Reference: the F20/S5 simulation, not measured traffic.
- F20/S20: half the output; 97.1% road and 88.9% temporal coverage.



Setting	Output	Road	Temporal	Speed corr.
F20/S5	1.17 GiB / 28.0 M	100%	100%	1.000
F20/S20	0.58 GiB / 13.9 M	97.1%	88.9%	0.897
F60/S60	0.49 GiB / 11.7 M	95.9%	82.3%	0.770

Example requirement: For a minimum temporal coverage of 88%, F20/S20 provides the smallest output among the settings shown.

Rerouting and Output Progress



Main Findings

- Useful FCD does not require maximum output.
- High road coverage does not guarantee temporal detail.
- 50% rerouting increased FCD span by 18–22%.

Zenodo: Data & Code

This work was also supported by the EPICURE project under grant agreement No 101139786. The project is supported by the European High-Performance Computing Joint Undertaking and its members (including top-up funding by the Ministry of Education, Youth and Sports of the Czech Republic ID: MC2402).

This poster is partially based upon work from the project „Floreon WIA“, VK02020253, supported by the Ministry of the Interior of the Czech Republic from the OPSEC Programme.

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90254).