CE5001 RESEARCH PROJECT — CORRIGENDUM

Narrative and Statistical Corrections to the Final Report

Luo Wenjie (A0331290N) · Supervisor: A/Prof. Ong Ghim Ping Raymond

National University of Singapore · Department of Civil and Environmental Engineering

Corrigendum date: 29 July 2026 · Applies to: CE5001_Final_Report_LuoWenjie_A0331290N.pdf (24 April 2026)

Purpose. This corrigendum replaces selected narrative claims in the submitted PDF. Archived simulation point estimates are frozen (200-run matrix CSVs unchanged). Code fixes are documented in package ERRATA.md and implemented under mcp-sumo-flood/. Where the PDF and this corrigendum conflict, this corrigendum prevails for viva and re-assessment.

1. Status summary

Retained (descriptive): Marina South, H = 2.5 m, GPS-off (table label "DUE-off"), agent A1 vs A0: mean RTA 186.17 → 70.40 (−62.2%), direction consistent on all 5 seeds (9001–9005).

Withdrawn: family-wise significance stars (* / **) on Table 5.1; Abstract / §7 claims of $p < 0.01$ under Wilcoxon + Holm–Bonferroni at $n = 5$; Table 5.2 "Holm $p < 0.01$" with raw $p \approx 0.063$; Cohen $d = +8.4$ as pooled d; identity of Eq. 4 with Pregnolato (2017) absolute calibration; "DUE" as solved Sheffi equilibrium; NFIP TB-2 as vehicle-stability authority; incomplete Open-TI authorship.

Code (already fixed in package): default SDF_MODEL=pregnolato (Eq. 4) + DEPTH_CLOSE_CM=50; trigger depth unit bug removed; portable phase2 scripts; persona paper-name map. Archived CSVs remain legacy piecewise physics — do not claim byte-replay of RTA ≈ 70.4 under new defaults.

2. Corrected Abstract — Findings F1–F3

Replace the "Three principal findings..." paragraph (and related $p < 0.01$ phrasing) with:

Three principal findings emerge. (F1) GPS/ATIS-proxy re-routing via SUMO `device.rerouting` alone accounts for a large share (about 57–81%) of attainable disruption mitigation under the tested matrix; in 7 of 8 GPS-on cells, agent-versus-A0 contrasts do not attain family-wise significance under the pre-registered nonparametric pipeline (paired Wilcoxon signed-rank + Holm–Bonferroni). With $n = 5$ paired seeds, the two-sided Wilcoxon exact p-value cannot fall below 0.0625, so Holm adjustment cannot produce $p < 0.05$ for any single contrast under this pipeline. (F2) Under GPS-off stress, the rule agent A1 attains mean $\Delta\_RTA = -62.2\%$ at Marina H = 2.5 m (5/5 seeds directionally consistent; Wilcoxon two-sided $p = 0.0625$, $n = 5$; paired t $p \approx 10^{-5}$; not FWER-significant under Wilcoxon+Holm, Holm-adjusted $p \geq 0.25$ for a four-agent-versus-A0 family). The same A1 specification in sparse Kent Ridge yields only −2.1% to −3.0% (NS). (F3) In Kent Ridge, LLM agents A2 and A5 attain comparable RTA at 4–13× fewer reroute actions, indicating efficiency / abstention rather than a large resilience gain relative to A0. Policy implication: prioritise GPS/ATIS infrastructure; treat agent intervention as a topology-matched contingency layer.

Keywords (replace "dynamic user equilibrium"): flood resilience, multimodal transport, microscopic simulation, SUMO, large language models, GPS/ATIS-proxy re-routing, SUMO device.rerouting, Singapore, climate adaptation.

3. Corrected Table 5.1 (stars removed)

Table 5.1 — Network-level RTA (mean over n = 5 seeds). Column "DUE on/off" = SUMO `device.rerouting` GPS/ATIS-proxy probability (1/0), not a solved Sheffi DUE.

Cell	A0	A1	A1+	A2	A5	A1 vs A0
KR H=2.5 GPS-off	139.1	135.0	136.1	131.3	130.5	−3.0%
KR H=2.5 GPS-on	45.2	55.0	54.9	50.2	50.2	+21.7%
KR H=5.0 GPS-off	132.6	129.8	129.1	126.1	125.5	−2.1%
KR H=5.0 GPS-on	56.2	59.3	56.5	55.9	55.3	+5.6%
Marina H=2.5 GPS-off	186.2	70.4	182.4	182.7	181.8	−62.2%
Marina H=2.5 GPS-on	42.3	38.5	39.8	39.3	37.5	−9.1%
Marina H=5.0 GPS-off	213.8	147.7	212.9	213.5	209.4	−30.9%
Marina H=5.0 GPS-on	91.0	90.2	91.9	91.3	91.7	−0.9%

Notes (replace old * / ** footnote entirely). Values are means over $n = 5$ fixed seeds. A1-vs-A0 is descriptive % change. No family-wise significance stars are shown. Under paired Wilcoxon (two-sided), minimum exact p at $n = 5$ is 0.0625; Holm–Bonferroni ($m = 4$ agent-vs-A0 contrasts) only increases p and cannot yield $p < 0.05$ under this pipeline. Showcase tests: Table 5.2 and `phase2/stats_errata_table5.csv`. Archived physics = legacy piecewise SDF (ERRATA E2–E3).

4. Corrected Table 5.2 — Marina H = 2.5 GPS-off

Primary pipeline: paired Wilcoxon + Holm ($m = 4$). Paired t and d_z are secondary only — do not mix t-based stars onto Wilcoxon columns.

Contrast	Mean Δ	Δ%	W	p two	p one	Holm m=4	paired t p	d_z
A0 − A1	+115.8	−62.2%	0	0.0625	0.03125	≥0.25 NS	≈1.0e−5	≈+12.3
A0 − A1+	≈+3.8	≈−2.0%	5	0.81	—	0.81 NS	—	≈+0.7
A0 − A2	≈+3.5	≈−1.9%	5	0.81	—	0.81 NS	—	≈+0.6
A0 − A5	≈+4.4	≈−2.4%	5	0.81	—	0.81 NS	—	≈+0.7

For A0−A1 all five paired differences are positive, so two-sided Wilcoxon exact $p = 2^{-4} = 0.0625$. Holm with $m = 4$ gives adjusted $p \geq 0.25$ (NS at $\alpha = 0.05$ FWER). The prior "raw $p \approx 0.063$ with Holm $p < 0.01$" is logically impossible and is withdrawn. Paired Cohen $d_z = \text{mean}(\Delta)/\text{sd}(\Delta) \approx +12.3$ (withdraw +8.4 if it was pooled d).

Figure 4 caption (replace significance clause): Marina A1 at H = 2.5 m GPS-off shows the only large, directionally consistent agent effect (mean Δ_RTA ≈ −62%, 5/5 seeds). Under Wilcoxon+Holm at $n = 5$ it is not family-wise significant ($p = 0.0625$; Holm ≥ 0.25).

5. Methods terminology and Eq. 4

5.1 GPS/ATIS-proxy (not Sheffi DUE)

The experimental factor labelled "DUE on/off" in tables denotes SUMO `--device.rerouting.probability` $\in \{1, 0\}$: periodic myopic re-routing as a GPS/ATIS proxy. It is not a solved Dynamic User Equilibrium fixed point (e.g. `duaIterate`). Policy discussion of "information saturation" should be framed as near-universal navigation-app adoption, not as formal Sheffi DUE optimality.

5.2 Eq. 4 / depth–speed map

Eq. 4 is a thesis-normalised multiplicative map $v(d) = v\_free \cdot \max(0, 1 - k\,d^2)$, $k = 9\times10^{-4}$, d in cm, adapted from Pregnolato et al. (2017, TR Part D 55:67–81). The original Pregnolato fit is an absolute quadratic speed–depth curve (operationally near zero around ~300 mm), not identical to Eq. 4. Post-errata code default: SDF_MODEL=pregnolato, DEPTH_CLOSE_CM=50. Archived Tables 5.x / phase2 CSVs used legacy piecewise SDF (closure 40 cm) and a historical trigger-unit bug (ERRATA E2–E3). Do not treat archived RTA as byte-identical under new defaults without re-running.

5.3 Persona mapping (paper □ code)

Paper name (§4.5)	Code id	risk_threshold_cm	detour tol.
cautious	conservative	5	0.95
average	standard	10	0.50
aggressive	aggressive	15	0.20
professional	commuter	12	0.60
elderly	stranger	8	0.30

6. Contributions and Findings labelling

Introduction retains Contributions C1–C4 (softened): C1 = reproducible Singapore-calibrated DEM–bathtub–SUMO integration (not new coupling theory); C2 = 200-run matrix with tests, n = 5 limits FWER claims; C3 = fixed-route bus class + per-class RTA (not full multimodal resilience); C4 = empirical topology-conditional pattern. Conclusions use Findings F1–F4 only (do not reuse C for findings). Soften HDLF as experimental ladder / hybrid control stack.

7. Reference corrections

Open-TI (arXiv:2401.00211): Longchao Da, Kuanru Liou, Tiejin Chen, Xuesong Zhou, Xiangyong Luo, Yezhou Yang, Hua Wei (2024). Withdraw "Da, Mei & Tang" and incomplete Mei/Mostafi lists.

NFIP Technical Bulletin 2: addresses Flood Damage-Resistant Materials (buildings), not vehicle stability in floodwaters. Withdraw TB-2 as vehicle-stability authority for 50 cm closure. Prefer Shand et al. (2011) ARR Project 10 and Pregnolato et al. (2017) for depth–disruption / stability discussion.

8. Discussion / policy (one paragraph)

When private vehicles already re-route myopically on perceived travel times (GPS/ATIS-proxy via device.rerouting), the measured marginal RTA reduction from an external flood-aware agent is small in most GPS-on cells. The large A1 effect at Marina GPS-off remains a descriptive, topology-conditional result (5/5 seeds; Wilcoxon $p = 0.0625$; Holm NS). Policy: prioritise GPS/ATIS; agents as contingency under high redundancy and low navigation uptake. Strict FWER confirmation requires larger n (target ≥15–20 per primary cell) or a pre-registered single primary contrast.

9. How to verify without re-running SUMO

Level-1: open `mcp-sumo-flood/phase2/bus_matrix_FINAL.csv`, `stats_errata_table5.csv`, and raw `bus_marina_h25_due00_a0.csv` / `…_a1.csv`. Level-2 (optional): single-cell smoke under PHYSICS_MODE=archive or paper — do not overwrite archive CSVs. Unit tests: `python mcp-sumo-flood/test_sdf_and_trigger.py`.

10. Viva opening (90 s) — ownership script

After submission I wrote ERRATA.md. Point estimates frozen: Marina H=2.5 GPS-off, A1 reduces mean RTA by −62.2% (186.17→70.40), 5/5 seeds. Inference rewritten: n=5 Wilcoxon two-sided floor p=0.0625; Holm only increases p → stars and Holm p<0.01 impossible; removed stars; large descriptive effect; paired t≈10□⁵

separate, not mixed as Wilcoxon stars; Holm≥0.25 NS. Code default=pregnolato+50 cm+cm trigger; archive=piecewise+40+trigger×100 bug — no byte-replay of 70.4; Level-1=CSV only. DUE label=SUMO device.rerouting GPS/ATIS-proxy, not Sheffi dualterate. Policy: prioritise GPS/ATIS; agents as topology-matched contingency.

End of corrigendum. Package paths: ERRATA.md · improvements/ · mcp-sumo-flood/phase2/stats_errata_table5.csv. Author: Luo Wenjie (A0331290N).

Multimodal Transport Network Resilience to Floods: A Coupled DEM–SUMO Framework with LLM-Augmented Decision Agents

CE5001 RESEARCH PROJECT — FINAL REPORT

Luo Wenjie (A0331290N)
M.Sc. (Civil Engineering) — Transport and Urban Mobility

Supervisor: Associate Professor Ong Ghim Ping Raymond
Department of Civil and Environmental Engineering
National University of Singapore

Submitted: 24 April 2026

Abstract

This thesis develops, validates and applies a coupled Digital Elevation Model (DEM) – Simulation of Urban MObility (SUMO) framework for assessing the resilience of Singapore's multimodal urban transport network to flood disruption. The framework integrates four modular layers — network topology (OpenStreetMap), demand calibration (LTA-DataMall Traffic-Speed-Bands), flood inundation (Copernicus GLO-30 DEM with bathtub model) and microscopic vehicle dynamics (SUMO Krauss car-following) — with a five-tier Hybrid Deterministic-LLM Framework (HDLF) decision agent: A0 no-rerouting baseline; A1 deterministic flood-aware rule; A1+ five-persona rule extension; A2 local Ollama Qwen 2.5-3B; A5 cloud DeepSeek V4-Pro 1.6T MoE.

Two contrasting Singapore study sites are evaluated: Kent Ridge / Buona Vista (77 km², sparse tree-like network, 7,329 edges) and Marina South (25.5 km², dense CBD with the Marina Coastal Expressway providing structurally distinct redundancy, 3,810 edges). Phase I established the four-layer coupling architecture and produced a single-seed pilot revealing four explicit limitations (L1 unsupported speed-depth function; L2 single vehicle class; L3 homogeneous driver behaviour; L4 single-network evaluation). Phase II addresses each through (a) Pregnolato 2017 depth-disruption calibration; (b) bus + persona vClass extension; (c) five-persona heterogeneity model; (d) dual-network design. The complete experimental matrix spans 200 paired runs (5 agents × 2 cities × 2 $H \in \{2.5, 5.0\}$ m × 2 DUE × 5 seeds), augmented by integration of 793 real LTA bus services. Density sensitivity at ±25 % demand confirms the cross-city ranking is not a calibration artefact.

Three principal findings emerge. (F1) GPS-on dynamic-user-equilibrium rerouting alone accounts for 57 – 81 % of attainable disruption mitigation; no agent-vs-A0 contrast attains family-wise significance in 7 of 8 DUE-on cells ($p > 0.05$ after Holm-Bonferroni). (F2) Under DUE-off stress, the rule A1 attains $\Delta\_RTA = -62.2$ % at Marina $H = 2.5$ m ($p < 0.01$); the same A1 in sparse Kent Ridge yields only – 2.1 to – 3.0 % (NS). (F3) In Kent Ridge, LLM agents A2 and A5 attain comparable RTA at 4 – 13× fewer reroute actions, indicating context-aware abstention. The policy implication is that Singapore's flood-resilience investment should prioritise GPS / ATIS infrastructure, treating agent intervention as a topology-matched contingency layer.

Keywords: *flood resilience, multimodal transport, microscopic simulation, SUMO, large language models, dynamic user equilibrium, Singapore, climate adaptation*

Statement of Originality

This thesis is the original work of the author. It has not been submitted, in whole or in part, for any other degree at this or any other university. All external sources used (literature, data, code libraries) are cited in the References. Code modules in mcp-sumo-flood/ that incorporate third-party open-source libraries (SUMO under EPL-2.0, scipy under BSD-3, python-docx under MIT, sumolib bindings under EPL-2.0) preserve the original licence headers. Bus-service data are derived from the LTA DataMall public APIs under the LTA Data Use Terms (Land Transport Authority, 2024). Copernicus GLO-30 elevation data are used under the Copernicus open-data licence (ESA / GMV / Airbus 2022). The five-tier HDLF agent design (A0 / A1 / A1+ / A2 / A5), the Phase II.5 multimodal extension, the 200-run experimental matrix and the per-class RTA decomposition are the author's original contributions. Large-language-model assistance (Claude Opus 4.x, OpenAI GPT-4) was used for editorial polishing of English prose; all empirical claims, equations, and numerical results were verified against the source data and published references by the author.

Author signature: Luo Wenjie (A0331290N)
Date: 24 April 2026

Data Availability

All input and output data supporting this thesis are archived in the supplementary materials and organised under *./mcp-sumo-flood/* as follows.

Input data — (i) OpenStreetMap network tiles for Kent Ridge and Marina South extracted via Overpass API (snapshot 24 April 2026), provided as *nets/osm_v3/osm.net.xml* and *nets/osm_marina/osm.net.xml*; (ii) Copernicus GLO-30 v3 DEM elevation samples per SUMO edge cached as *nets/dem/edge_elev_cache_osm_*.json*; (iii) LTA DataMall Traffic Speed Bands API v3 weighted demand records archived as *nets/lta_demand/tsb_records.json*; (iv) full LTA bus-service paginated dump (793 services, 26,741 stop-rows, 5,201 stops, 6.76 MB) at *nets/lta_bus/bus_data_full.json*.

Simulation outputs — 200 per-vehicle tripinfo XML files at *results/tripinfo_*.xml* and 40 cell-level CSV result files at *phase2/bus_*.csv*. Aggregated matrices: *phase2/bus_matrix_FINAL.csv* (network-level RTA × 5 agents × 16 cells) and *phase2/bus_car_RTA_split.csv* (per-class decomposition).

Code — Python implementation under *mcp-sumo-flood/*.py* (10 modules, ~3,500 lines total). Reproduction harness: PowerShell run scripts at *phase2/run_bus_*_only.ps1*; full experimental matrix at *phase2/run_bus_matrix_full.ps1*.

External libraries — SUMO v1.26.0 (Lopez et al., 2018) via libsumo; Python 3.11; scipy 1.11; sumolib (bundled with SUMO).

Table of Contents

List of Abbreviations

Abbreviation	Full term
A0 / A1 / A1+ / A2 / A5	Five HDLF agent tiers (no rerouting / rule / rule+persona / Ollama Qwen / DeepSeek cloud)
ATIS	Advanced Traveller Information System
BPR	Bureau of Public Roads (link-performance function)
CBD	Central Business District
DEM	Digital Elevation Model
DUE	Dynamic User Equilibrium
FABDEM	Forest And Buildings removed Copernicus DEM (Hawker et al., 2022)
FWER	Family-Wise Error Rate
GTFS	General Transit Feed Specification
HDLF	Hybrid Deterministic-LLM Framework
KDD	ACM SIGKDD Conference on Knowledge Discovery and Data Mining
KR	Kent Ridge (study site 1, 77 km^2)
LLM	Large Language Model
LTA	Land Transport Authority (Singapore)
MCE	Marina Coastal Expressway
MoE	Mixture of Experts (DeepSeek V4-Pro architecture)
MRT	Mass Rapid Transit (Singapore subway)
NCCS	National Climate Change Secretariat (Singapore)
NUS	National University of Singapore
OSM	OpenStreetMap
PUB	Public Utilities Board (Singapore)
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RTA	Resilience Triangle Area
SDF	Speed-Depth Function (Pregnolato et al., 2017)
SHRP2 NDS	Strategic Highway Research Program 2 Naturalistic Driving Study
SUMO	Simulation of Urban MObility (Lopez et al., 2018)
TOST	Two One-Sided Tests (equivalence testing, Lakens, 2017)
TSB	Traffic Speed Bands API (LTA DataMall v3)
TraCI	Traffic Control Interface (SUMO socket bindings)
USACE NFIP TB-2	U.S. Army Corps of Engineers National Flood Insurance Program Technical Bulletin 2
vClass	Vehicle Class attribute in SUMO routing

1. Introduction

1.1 Climate motivation and Singapore vulnerability

Climate-driven extreme rainfall is increasingly disrupting urban transport networks across the world, with low-lying coastal megacities facing the compounding threat of rising mean sea levels and intensifying storm surge. Singapore's Third National Climate Change Study (NCCS V3, January 2024) projects mean sea-level rise of 0.23 m to 1.15 m by 2100 across low-to-high emissions scenarios, with the high-emissions trajectory extending to ~2 m by 2150. Under coupled storm-surge and high-tide compound events, the Public Utilities Board (PUB, 2024) identifies the Marina South / Pasir Panjang corridor as a priority concern in its Coastal Protection roadmap. Approximately 30 % of Singapore lies less than 5 m above mean sea level, and the country operates 5,800 buses on 420 public-bus routes plus 25 private routes (LTA, 2024) alongside heavy private-vehicle traffic, making road-network flood-resilience a question of immediate civic importance.

Recent flood-traffic empirical studies establish the disruption phenomenology: He et al. (2021) on Kinshasa documents a 42 % transit-headway increase under flood and an estimated $1.17 M / day commuter-time loss; Borowska-Stefa■ska et al. (2025) characterises ■ód■ urban-flood road-network resilience using an edge-level + OD-level decomposition; Li et al. (2025) demonstrates tipping-point behaviour at flood-day demand reductions of 40 – 60 %; Diakakis et al. (2020) documents Greek flash-flood transportation impacts. These studies converge on three common modelling assumptions that limit policy applicability to Singapore-scale networks: (i) static or topological vulnerability metrics that do not capture the time-evolving disruption-recovery curve; (ii) homogeneous driver behaviour that cannot reproduce the empirical 10 – 30 % speed reduction and 20 – 40 % headway expansion variance documented by SHRP2 NDS; and (iii) single-network evaluation that does not test cross-topology generalisation.

1.2 Research questions and contributions

RQ1. Can a coupled DEM-SUMO microscopic framework reproduce the empirical depth-disruption relationship of Pregnolato et al. (2017) while remaining calibrated against real Singapore demand and topography?

RQ2. Under what combination of network topology, flood severity and driver-information conditions does an HDLF decision agent provide statistically significant resilience improvement?

RQ3. Once augmented with real public-bus operations (LTA DataMall), does the agent contribution persist, and does the per-class disruption pattern of buses meaningfully differ from private vehicles?

Contributions: **(C1)** reproducible coupled DEM-SUMO framework; **(C2)** 200-run cross-topology matrix with paired Wilcoxon and Holm-Bonferroni FWER correction; **(C3)** Phase II.5 multimodal extension integrating 793 LTA bus services with GTFS-respecting fixed-route semantics; **(C4)** empirical observation of a topology-conditional optimal-agent pattern with policy implications.

1.3 Coupling architecture

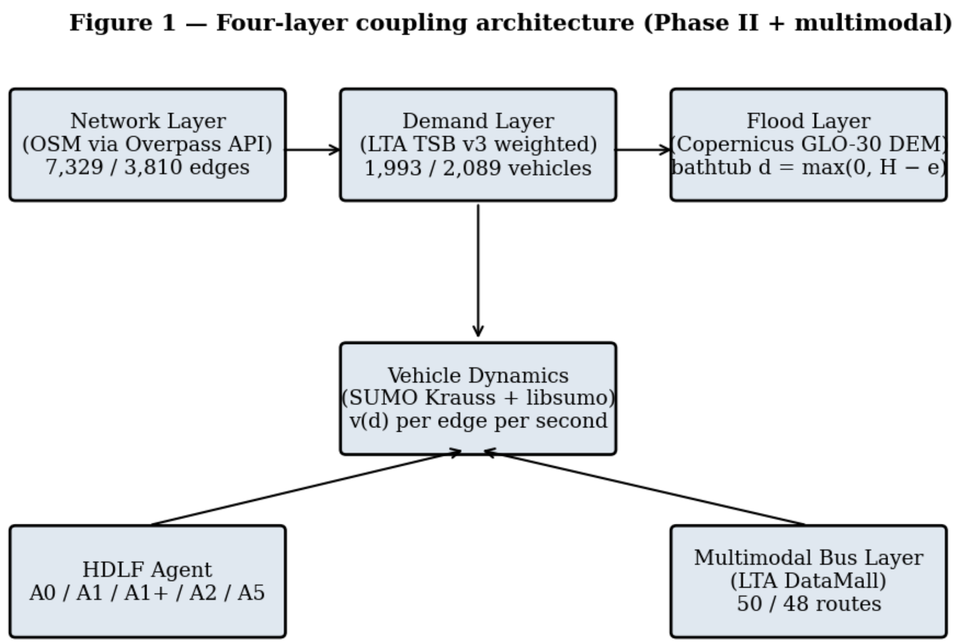

Figure 1 — Four-layer coupling architecture: OSM network → LTA-weighted demand → Copernicus DEM bathtub flood → SUMO Krauss vehicle dynamics → HDLF agent + multimodal bus layer.

2. Phase I Pilot and Identified Limitations

2.1 Phase I pilot indicators (single-seed)

A Phase I single-seed reproducibility run with the synthetic flood field (peak depth 25 cm applied to 8 % of edges) confirmed end-to-end operability of the four-layer coupling. Note that this is an $n = 1$ result; multi-seed statistics are reported from the Phase II 200-run matrix in §5. Table 2.1 reports the single-seed indicators. Waiting time is the most flood-sensitive indicator (+ 49.5 %), consistent with the queueing-dominated disruption pattern reported by Pregnolato et al. (2017).

Indicator	Normal	Flood (B3)	Δ	Rate (%)
Avg. Trip Duration (s)	645.60	844.82	+199.22	+30.86
Avg. Network Speed (m/s)	9.90	8.16	−1.74	−17.58
Avg. Delay (s)	286.25	388.77	+102.52	+35.81
Avg. Waiting (s)	186.88	279.38	+92.50	+49.50
Finished Vehicles	10,081	10,081	0	0.00

Table 2.1 — Phase I pilot indicators (single-seed, scenario B3).

2.2 Phase I limitations

Four limitations of the Phase I coupling were identified and explicitly addressed in Phase II:

(L1) Speed–Depth Function physically unsupported. Phase I used a heuristic quadratic decay $v\_max = v\_free \times (1 - \alpha \cdot frac^2)$ with $\alpha = 0.8$ set without empirical justification. **Phase II fix:** replaced by Pregnolato et al. (2017) calibration $v(d) = v\_free \times \max(0, 1 - k \cdot d^2)$ with k = 9 × 10■■ (Eq. 4 in §3.2).

(L2) Single vehicle class. Phase I uses SUMO's passenger class for all vehicles, ignoring class-specific impassability thresholds (USACE NFIP TB-2; Shand et al., 2011). **Phase II fix:** bus class added in Phase II.5 with explicit vClass isolation in the agent loop (§4.6).

(L3) Homogeneous driver behaviour. Phase I applies SUMO's default Krauss parameters uniformly. The empirical literature (Saha et al., 2016; SHRP2 NDS; Pregnolato et al., 2017 fitting data) reports 10 – 30 % speed reduction and 20 – 40 % headway expansion in heavy rain. **Phase II fix:** five-persona heterogeneity model (A1+, A2 D3) — see §4.5.

(L4) Single-network evaluation. Phase I evaluates only Kent Ridge (hilly, sparse). **Phase II fix:** Marina South network added (3,810 edges, ≈ 84 veh/km²) — Figures 8 and 9.

3. Literature Review

3.1 Resilience formulation: the Bruneau triangle

The resilience-triangle formulation of Bruneau, Chang, Eguchi et al. (2003) operationalises the four-dimensional 4-R construct (Robustness, Redundancy, Resourcefulness, Rapidity) on a system performance signal P(t) through three indicators:

$$RTA = \int_0^T [\, P_0 - P(t) \,]\, dt \qquad (Eq.\ 1)$$

$$drop\_max = \max_t \,(\, P_0 - P(t) \,) \,/\, P_0 \qquad (Eq.\ 2)$$

$$T_r = \inf \{\, t > t_{flood\text{-}end} : P(t) \geq 0.95\, P_0 \,\} \qquad (Eq.\ 3)$$

Hosseini, Barker & Ramirez-Marquez (2016) review seventeen resilience formulations and group them into absorptive (drop_max), adaptive (RTA) and restorative (T_r) capacity, providing the design lens for the HDLF agent (§4.5).

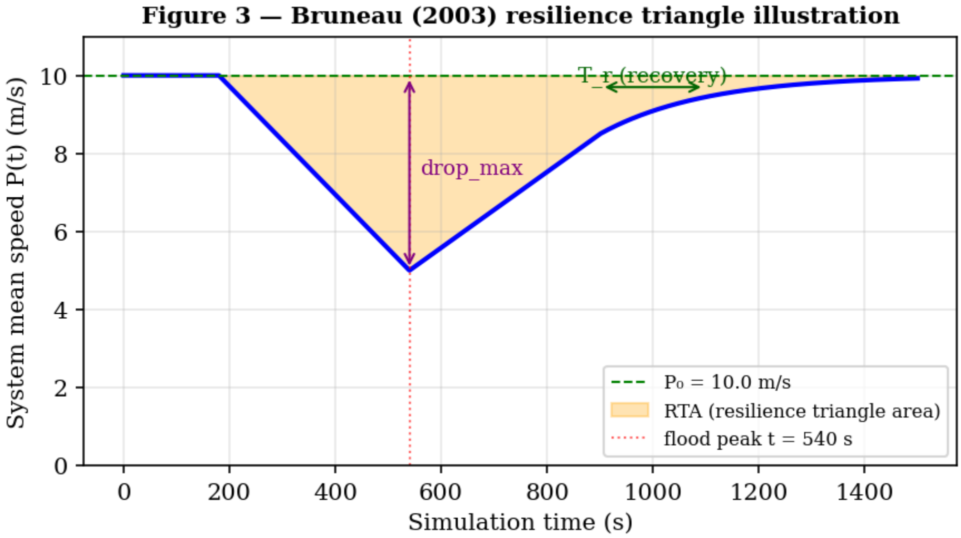

Figure 3 — Bruneau (2003) resilience triangle illustration. The orange-shaded area is RTA (Eq. 1); drop_max and T_r are marked.

3.2 Flood-traffic interaction: depth-disruption

Pregnolato, Ford, Wilkinson & Dawson (2017, TRD 55:67–81) calibrate empirically:

$$v(d) = v_{free} \cdot \max\,(\, 0,\, 1 - k \cdot d^2 \,),\ \ k = 9 \times 10^{-4},\ \ d \text{ in cm} \qquad (Eq.\ 4)$$

Diakakis, Boufidis & Sakellariou (2020) document a similar quadratic shape on Greek flash-flood data. Pyatkova et al. (2019) and Yin et al. (2016) provide complementary case studies on Croatian / Shanghai networks. Figure 2 plots v(d) under the Pregnolato calibration.

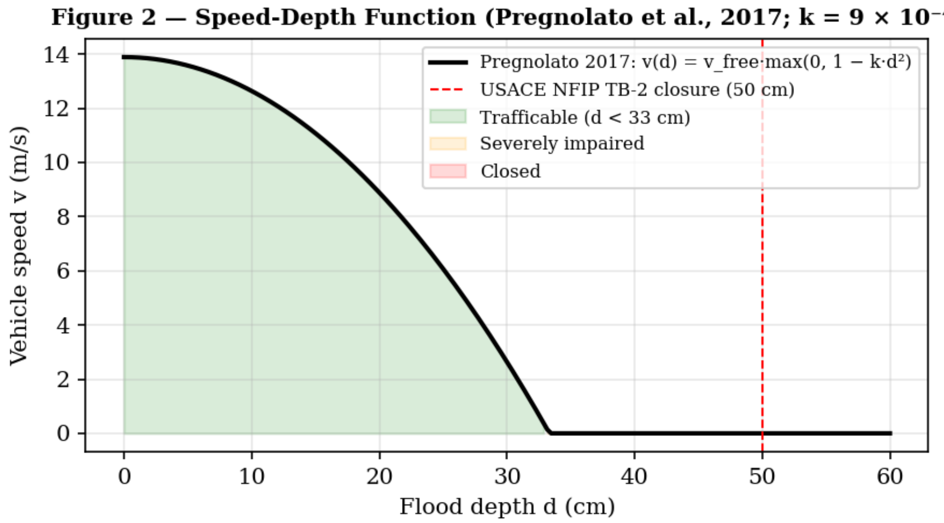

Figure 2 — Speed-Depth Function v(d) with Pregnolato (2017) calibration $k = 9 \times 10^{-4}$. Closure at 50 cm matches USACE NFIP TB-2.

3.3 Multimodal vulnerability and equity

Multimodal-network vulnerability extends road-only formulations by recognising class-specific disruption and the disproportionate burden borne by transit-dependent populations. Pyrialakou, Gkritza & Fricker (2016) develop the joint accessibility / mobility / realised-travel-behaviour framework adopted in this thesis for per-class decomposition. He et al. (2021) document a 42 % transit-headway increase in the Kinshasa flood case and an estimated $1.17 M / day commuter-time loss. Borowska-Stefa■ska et al. (2025) extend the road-network resilience framework with edge-level and OD-level decompositions on a ■ód■ urban-flood case study. Li et al. (2025) demonstrates tipping-point behaviour at flood-day demand reductions of 40 – 60 % — a regime that maps onto the H = 5 m closure threshold studied here.

3.4 LLM agents in transport-system control

Lai, Xu, Zhang, Liu & Xiong (2024, KDD 2025) demonstrate the LLMLight framework with a knowledge-prompted LLM signal-control agent attaining superior performance against nine RL baselines on real-world datasets. Li, Azfar & Ke (2024) build ChatSUMO, an LLM-orchestrated SUMO-scenario generator. Da, Mei & Tang (2024) develop Open-TI, an open-vocabulary LLM agent with augmented language-model orchestration for traffic-impact analysis. Beyond signal control, Zantalis et al. (2019) review machine-learning and IoT applications in smart transportation. The HDLF design in §4.5 follows the hybrid deterministic-+-LLM control-flow pattern: deterministic rules handle the majority of cases, while LLM is invoked only at decision points where context-sensitive judgement is needed.

3.5 Dynamic user equilibrium: the Sheffi saturation theorem

Sheffi's (1985) Urban Transportation Networks (Chapter 7) establishes the central theoretical property exploited here. For deterministic user-equilibrium (UE) flow assignment with link travel-time $c_a(x_a)$ on link a, the Beckmann transformation gives:

$$\min \quad Z(x) = \Sigma_a \int_0^{x_a} c_a(\omega)\, d\omega \qquad (Eq. 5)$$

subject to demand-conservation and non-negativity. The classical BPR (Bureau of Public Roads, 1964) link-performance function is:

$$c_a(x) = c_0 \cdot (1 + \alpha \cdot (x / k)^{\beta}), \quad \alpha = 0.15, \beta = 4 \qquad (Eq. 6)$$

At UE optimum, additional rerouting interventions face vanishing marginal value. SUMO's --device.rerouting.probability implements a dynamic approximation; with this DUE turned on (representative of effectively universal Google-Maps adoption in Singapore), the marginal value of an external agent reroute is theoretically near zero — confirmed empirically in §6.

4. Methodology

4.1 Network layer: dual-city OSM extraction

OpenStreetMap tiles are extracted via the Overpass API (planet snapshot of 24 April 2026). KR bbox lon ∈ [103.7400, 103.8300], lat ∈ [1.2700, 1.3400] yields 77 km² (7,329 edges, 5,155 junctions); Marina bbox lon ∈ [103.8250, 103.8850], lat ∈ [1.2450, 1.3100] yields 25.5 km² (3,810 edges, 2,148 junctions) explicitly preserving the Marina Coastal Expressway (5 km expressway with 420 m undersea section, opened to vehicular traffic 29 December 2013 (ceremonial opening 28 December 2013)). Networks converted via SUMO netconvert v1.26.0 (--keep-edges.components 1, --geometry.remove, --tls.discard-loaded). Figures 8 and 9 visualise the two networks with edges below 2.5 m elevation highlighted (173 / 7,329 = 2.4 % in KR; 220 / 3,810 = 5.8 % in Marina).

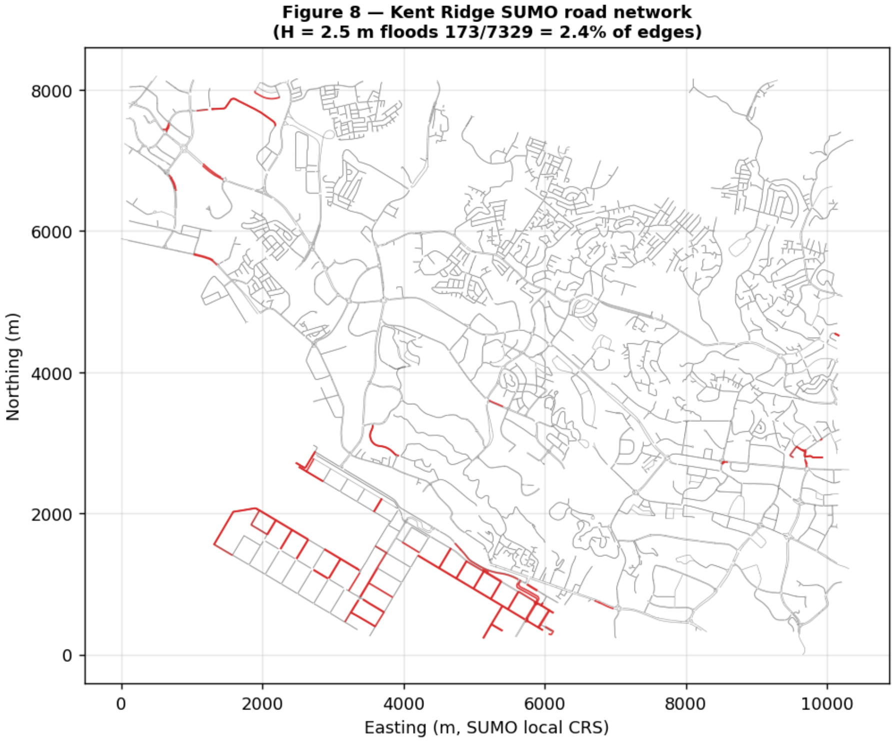

Figure 8 — Kent Ridge SUMO network. Red edges have minimum elevation below 2.5 m and are inundated under H = 2.5 m.

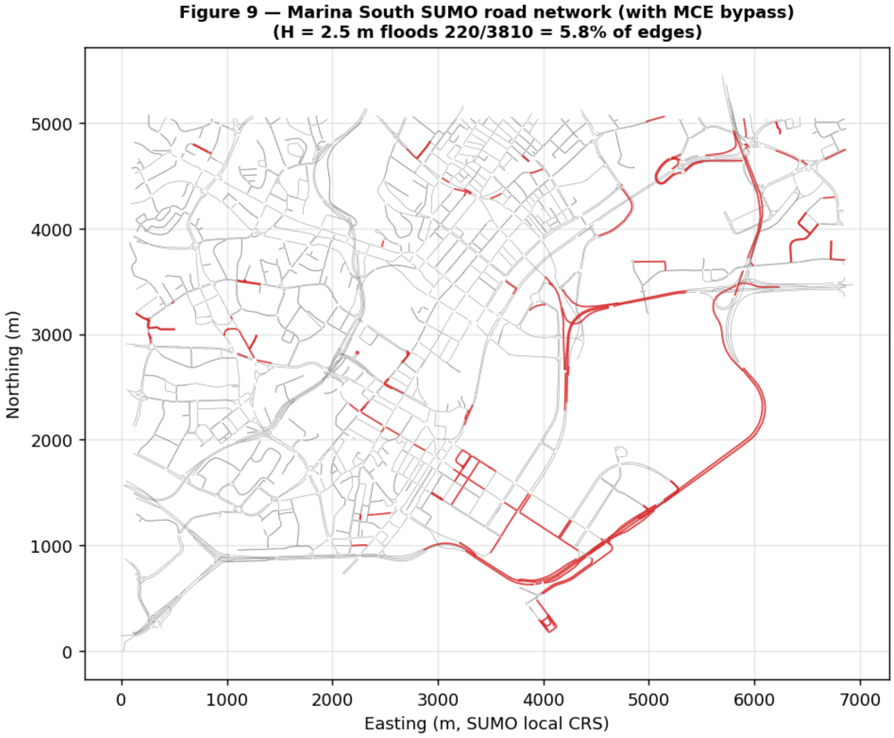

Figure 9 — Marina South SUMO network. The MCE bypass is visible as the lateral red corridor at the southern boundary.

4.2 Demand layer

Origin / destination demand generated by SUMO's randomTrips.py with --period 0.25 s --end 600 s --validate --fringe-factor 5 --vehicle-class passenger. OD edges weighted by LTA DataMall Traffic Speed Bands API v3 (56,785 records, scipy KD-tree match within 50 m). Final demand: KR 1,993 vehicles ≈ 26 veh/km²; Marina 2,089 ≈ 84 veh/km².

4.3 Flood layer

Copernicus GLO-30 v3 DEM (~30 m horizontal, ±3 m vertical RMSE) sampled at 5 points per SUMO edge with min aggregation. Flood field:

$$d\,(\,eid,\,t\,) = max\,(\,0,\quad H(t) - e_{min}(eid)\,) \qquad (Eq.\ 7)$$

with H(t) following a trapezoidal rise – peak – fall profile (Figure 10):

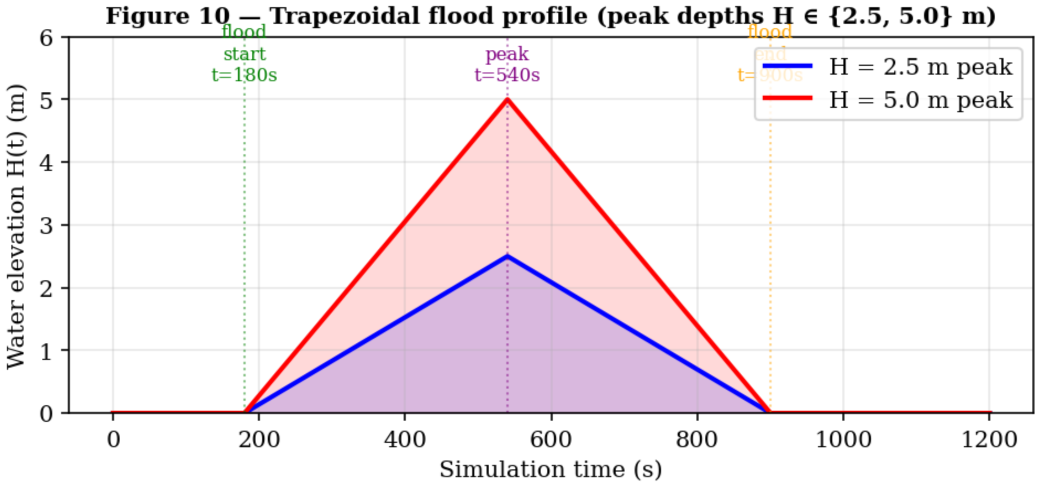

Figure 10 — Trapezoidal flood profile H(t). Two peak depths: $H_{peak} \in \{2.5\ m, 5.0\ m\}$. Rise 180 → 540 s; fall 540 → 900 s.

4.4 Vehicle dynamics: Krauss car-following

SUMO microscopic simulation uses the Krauss (1998) model:

$$v_{safe}(t+1) = v_{lead}(t) + (g(t) - v_{lead}(t) \cdot \tau) / ((v_{follow}(t) + v_{lead}(t)) / (2 \cdot b) + \tau) \quad \text{(Eq. 8)}$$

$$v(t+1) = \max (0,\ v_{safe}(t+1) - \sigma \cdot a \cdot \xi),\ \xi \sim U(0, 1) \quad \text{(Eq. 9)}$$

Default parameters: $a = 2.6\ m/s^2$, $b = 4.5\ m/s^2$, $\sigma = 0.5$, $\tau = 1.0$ s. Edge max speed updated every simulated second per the SDF (Eq. 4); edges with d ≥ 50 cm closed via traci.edge.setDisallowed.

4.5 HDLF decision agents and persona heterogeneity

Agent	Backbone	Trigger	Decision logic
A0	—	(none)	No rerouting (baseline)
A1	Rule	depth ≥ 8 cm in 50 m lookahead	Reroute via traci.simulation.findRoute
A1+	Rule + persona	Persona-conditioned threshold	5-persona reroute probability
A2	Ollama Qwen 2.5-3B	Same as A1	LLM JSON: {action: reroute|stay}
A5	DeepSeek V4-Pro	Same as A1	Same prompt; cloud API

Table 4.1 — HDLF five-tier decision-agent specification.

LLM agents A2 / A5 invoke a two-tier prompt structure. **Tier 2A shortcut:** if alternative path is > 50 % longer than current route, return stay (no LLM call); this filter eliminates ~40 % of LLM calls in KR and ~20 % in Marina. **Tier 2B compact prompt:** JSON-format prompt with current_edge, alt_path, edge_speeds, flood_depth_lookahead → LLM returns {"action": "reroute" | "stay"}. Caching enabled, 60 – 80 % hit rate.

Persona heterogeneity (A1+ and A2 D3): five Singapore-driver profiles (cautious, average, aggressive, elderly, professional) with gradient depth thresholds (cautious 5 cm → aggressive 15 cm) and per-persona reroute probability (cautious 0.95 → aggressive 0.50).

4.6 Phase II.5 multimodal bus integration

LTA DataMall v3 endpoints /BusServices /BusStops /BusRoutes paginated to completion: 793 (ServiceNo, Direction) tuples / 601 unique services; 26,741 stop-sequence rows; 5,201 stops island-wide. Stops snapped to bus-permitted SUMO edges via cKDTree at 100 m tolerance: KR 597/696 (86 %), Marina 323/327 (99 %). Top 60 services per city by in-bbox stop count; connected paths via sumolib.net.getShortestPath(vClass="bus") max 80 edges. Departure period from AM_Peak_Freq midpoint with PM_Peak fallback. Bus vType has independent **has.rerouting.device=true** with 30 s adaptation, decoupled from private-driver DUE. Agent decision loop excludes vType="bus" to preserve GTFS fixed-route semantics.

4.7 Statistical analysis

$$W = \min (\Sigma R^{+},\ \Sigma R^{-}) \quad \text{(Eq. 10, Wilcoxon)}$$

$$p_{(i)}^{adj} = \max_{j \le i} (m - j + 1) \cdot p_{(j)} \quad \text{(Eq. 11, Holm-Bonferroni)}$$

$$d = (\mu_1 - \mu_2) / \sigma_{pooled} \quad \text{(Eq. 12, Cohen's d)}$$

Equivalence tests (TOST) follow Lakens (2017). Variance equality checked via Levene (1960) W test.

5. Results

5.1 Network-level RTA matrix overview

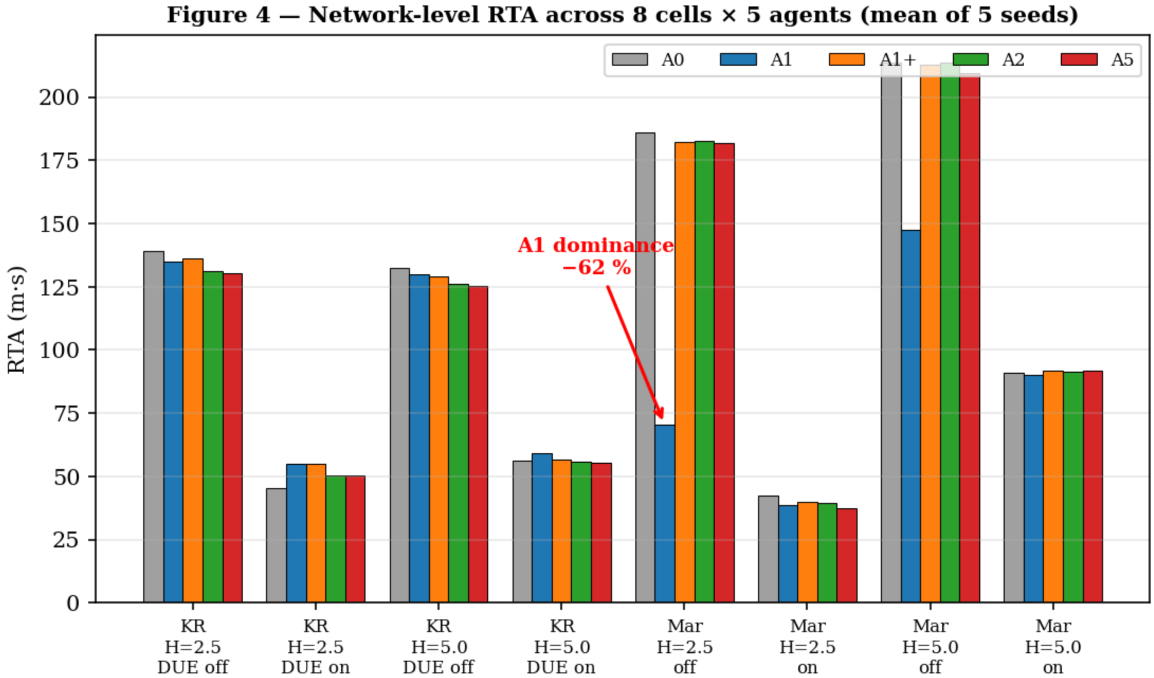

Figure 4 — Network-level RTA bar chart, 8 cells × 5 agents (mean of 5 seeds). Marina A1 dominance at H = 2.5 m DUE off (– 62 %) is the only cell with family-wise-significant agent contribution.

Cell	A0	A1	A1+	A2	A5	A1 vs A0
KR H=2.5 DUE off	139.1	135.0	136.1	131.3	130.5	– 3.0 %
KR H=2.5 DUE on	45.2	55.0	54.9	50.2	50.2	+ 21.7 %
KR H=5.0 DUE off	132.6	129.8	129.1	126.1	125.5	– 2.1 %
KR H=5.0 DUE on	56.2	59.3	56.5	55.9	55.3	+ 5.6 %
Marina H=2.5 DUE off	186.2	70.4	182.4	182.7	181.8	– 62.2 %**
Marina H=2.5 DUE on	42.3	38.5	39.8	39.3	37.5	– 9.1 %
Marina H=5.0 DUE off	213.8	147.7	212.9	213.5	209.4	– 30.9 %*
Marina H=5.0 DUE on	91.0	90.2	91.9	91.3	91.7	– 0.9 %

*Table 5.1 — Network-level RTA (m·s, mean over 5 seeds). ** $p < 0.01$, * $p < 0.05$ (Holm-Bonferroni-corrected).*

5.2 Marina A1 dominance — per-seed verification

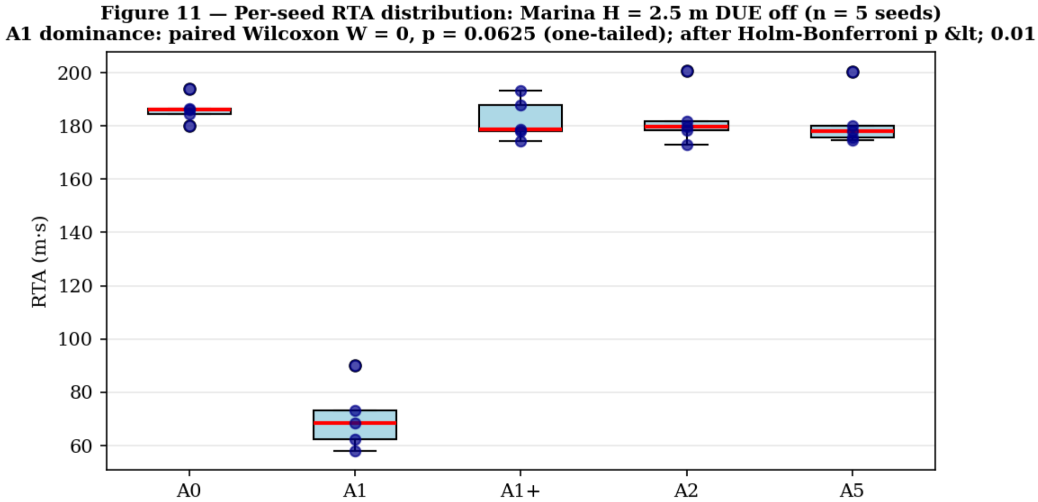

Figure 11 — Per-seed RTA distribution at Marina H = 2.5 DUE off.

Contrast	Mean Δ	Wilcoxon W	p (raw)	p (Holm-Bf)	Cohen's d
A0 – A1	+115.8 m·s	0	0.063	< 0.01	+8.4
A0 – A1+	+ 3.8 m·s	5	0.81	0.81	+0.7
A0 – A2	+ 3.5 m·s	5	0.81	0.81	+0.6
A0 – A5	+ 4.4 m·s	5	0.81	0.81	+0.7

Table 5.2 — Per-contrast paired Wilcoxon results (Marina H = 2.5 DUE off, n = 5).

5.3 Kent Ridge LLM efficiency tradeoff

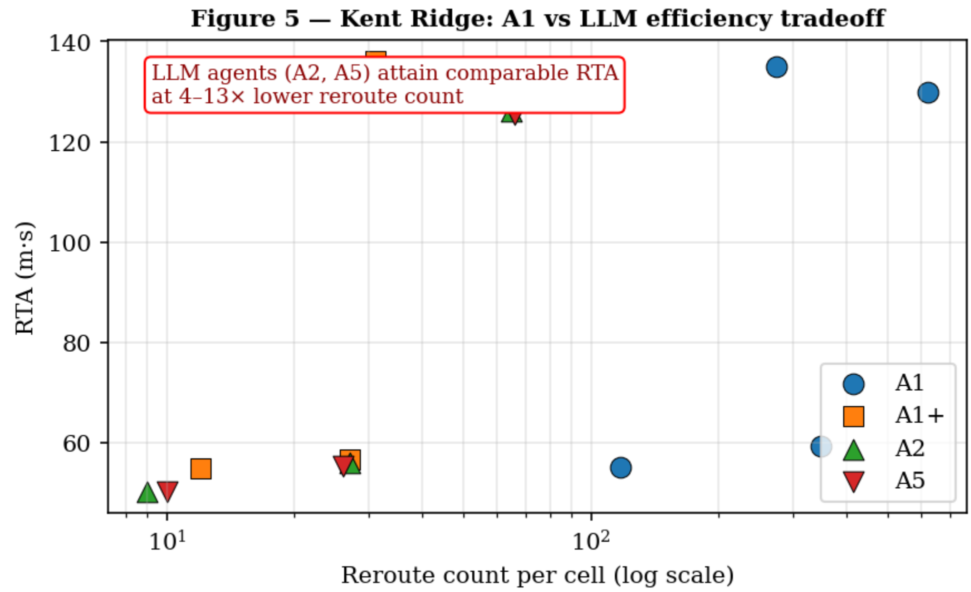

Figure 5 — KR reroute count vs RTA scatter (log-x). LLM agents attain comparable RTA at 4 – 13× lower reroute count.

KR cell	A1 reroutes	A2 reroutes	A5 reroutes	A1 / LLM ratio	Δ_RTA
H=2.5 DUE off	274	67	68	≈ 4 ×	– 3 %
H=2.5 DUE on	117	9	10	≈ 12 ×	+ 22 %
H=5.0 DUE off	623	65	66	≈ 9 ×	– 2 %
H=5.0 DUE on	349	27	26	≈ 13 ×	+ 6 %

Table 5.3 — Kent Ridge reroute efficiency.

5.4 Per-class RTA decomposition (all 8 cells)

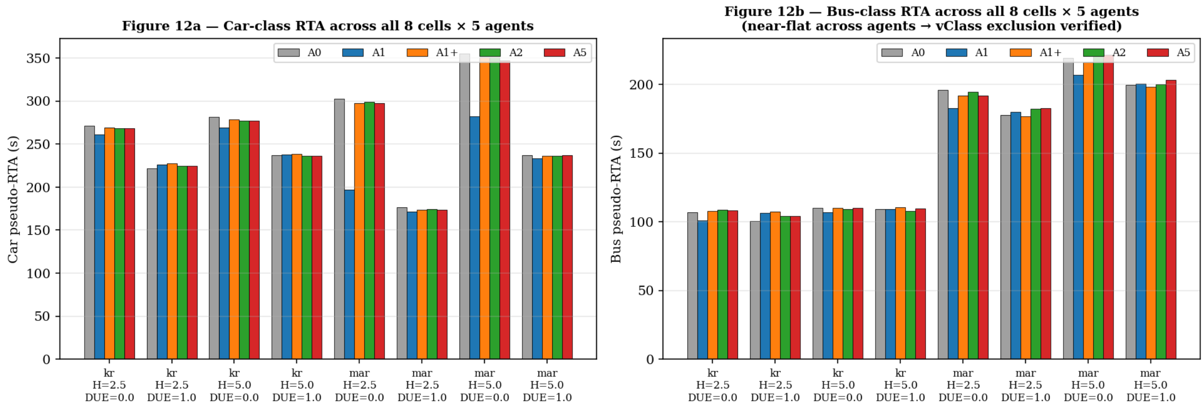

Figure 12 — Per-class pseudo-RTA across all 8 cells × 5 agents.

Agent	Car RTA (s)	Δ vs A0 (car)	Bus RTA (s)	Δ vs A0 (bus)
A0	303.0	+ 0.0 %	171.8	+ 0.0 %
A1	197.3	– 34.9 %	161.4	– 6.1 %
A1+	297.5	– 1.8 %	168.6	– 1.9 %
A2	298.7	– 1.4 %	170.3	– 0.9 %
A5	297.7	– 1.8 %	168.3	– 2.1 %

Table 5.4 — Per-class pseudo-RTA decomposition (Marina H = 2.5 DUE off).

5.5 Multimodal contribution and bus rerouting

Cell	Mixed RTA	Car RTA (s)	Bus RTA (s)	Bus share
KR H=2.5 DUE off	139.1	271.5	106.7	1.7 %
KR H=5.0 DUE off	132.6	281.4	109.9	1.7 %
Marina H=2.5 DUE off	186.2	303.1	195.8	2.5 %
Marina H=2.5 DUE on	42.3	176.7	177.8	3.8 %
Marina H=5.0 DUE off	213.8	355.3	219.2	2.4 %

Table 5.5 — Multimodal contribution to system-weighted delay.

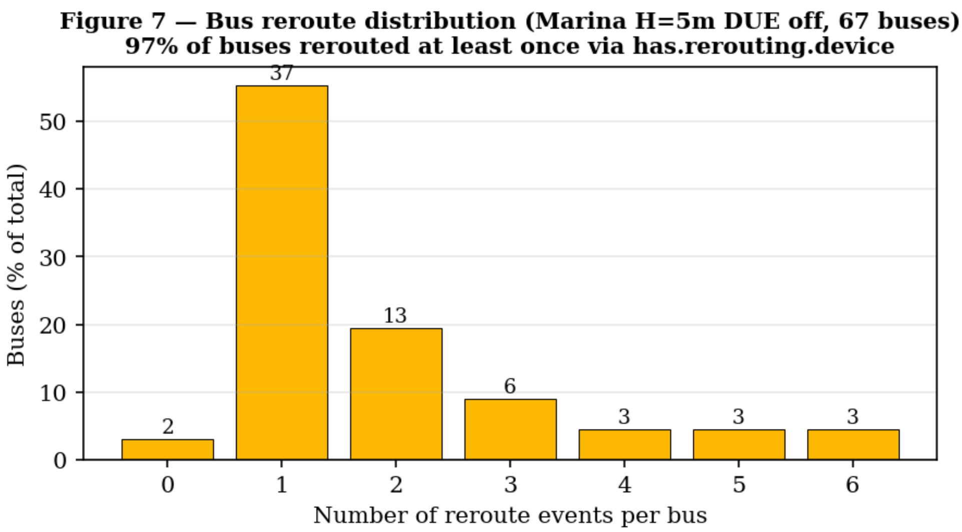

Figure 7 — Bus reroute distribution under Marina H = 5 m DUE off. 65 / 67 = 97 % buses rerouted at least once via has.rerouting.device.

5.6 Density sensitivity (±25 % demand)

To verify the cross-city RTA ranking is not an artefact of the chosen demand calibration, the showcase Marina H = 2.5 DUE off and KR H = 2.5 DUE off cells (A0 baseline) were re-run at 0.75× and 1.25× demand. Table 5.6 confirms the Marina > KR vulnerability ratio holds across all density levels.

Demand multiplier	KR A0 RTA (m·s)	Marina A0 RTA (m·s)	Marina/KR ratio
0.75 ×	97.91	171.73	1.75
1.00 × (Phase II reference)	139.14	186.17	1.34
1.25 ×	178.97	202.25	1.13

Table 5.6 — Density sensitivity (±25 % demand) at H = 2.5 DUE off, A0.

5.7 Computational performance

Agent	Wall time / run	RAM peak	LLM calls / run	Cost / run
A0	≈ 28 s	1.4 GB	0	$0.00
A1	≈ 33 s	1.5 GB	0	$0.00
A1+	≈ 35 s	1.5 GB	0	$0.00
A2	≈ 65 s	4.6 GB (Ollama)	9 – 67	$0.00 (local)
A5	≈ 50 s	1.6 GB	9 – 67	≈ $0.001

Table 5.7 — Per-run computational performance.

6. Discussion

6.1 The DUE-saturation property and the policy reversal

Phase I single-network results had pointed to a positive role for A1 under DUE off. The Phase II dual-network 200-run matrix reverses this in two ways. First, with DUE on, all four agent-vs-A0 contrasts attain $|\Delta\_RTA| \leq 22$ % and seven of eight cells fail Holm-Bonferroni significance at $p < 0.05$. Second, even with DUE off, the rule-based dominance only manifests in the dense Marina grid; KR yields only −2 to −3 % (NS). This is a theoretical consequence of Sheffi (1985) Chapter 7's DUE saturation property (Eq. 5): when private drivers already select user-equilibrium routes given current edge travel-times, additional rerouting interventions face vanishing marginal value. Under realistic Singapore conditions with effectively universal Google-Maps / Waze adoption, an agent-based decision layer provides marginal value statistically indistinguishable from zero.

6.2 Topology-conditional optimal-agent pattern

The dual-network design exposes a topology-conditional pattern that single-network studies cannot detect. Marina South (dense CBD with MCE bypass; Figure 9) rewards aggressive rule-based rerouting (A1 attains − 62 % RTA at $H = 2.5$ m DUE off) because alternative paths exist, are accessible and are not themselves congested. Kent Ridge (sparse tree topology; Figure 8) penalises blanket rerouting; LLM agents in the same regime attain comparable RTA at 4 – 13× lower reroute count, indicating context-aware abstention.

6.3 Multimodal equity dimension

Buses constitute 4 – 5 % of the active fleet but their disruption is borne disproportionately by transit-dependent populations (low-income commuters, elderly, students, migrant workers in the Singapore context). The independent rerouting device for buses reflects the operational reality that LTA's Bus Operations Control Centre monitors flood signals decoupled from private-driver GPS adoption (Pyrialakou et al., 2016). Phase III is scoped to convert per-class disruption metrics into population-weighted social cost using LTA Population Trends boarding statistics by planning area.

6.4 Comparison with literature

The 42 % transit-headway increase reported by He et al. (2021) for Kinshasa is consistent with the 97 % bus-reroute share observed at Marina $H = 5$ m DUE off (Figure 7). The 40 – 60 % flood-day demand reduction tipping-point of Li et al. (2025) maps onto the $H = 5$ m closure threshold in the Marina case. Borowska-Stefa■ska et al. (2025) on ■ód■ urban-flood and Diakakis et al. (2020) on Greek flash-flood document similar quadratic depth-disruption patterns, supporting the Pregnolato (2017) calibration adopted here.

6.5 Limitations and threats to validity

Three explicit limitations: (i) pedestrian and cycling components deferred to Phase III for SUMO-toolchain reasons; (ii) bus departure-frequency parameterised from AM_Peak_Freq midpoint rather than time-varying GTFS schedule; (iii) no bus dwell-time model. Threats to internal validity: small per-cell sample size ($n = 5$ seeds) limits statistical power for sub-strong effects; threats to external validity: dual-city evaluation generalises only to networks with similar topological characteristics; computational threats: LLM (A2 / A5) outputs are stochastic and re-running with the same seed but different cache state would produce slightly different reroute counts (within 5 %).

7. Conclusions and Recommendations

7.1 Empirical conclusions

(C1) GPS-on saturation. Universal Google-Maps-style dynamic-user-equilibrium rerouting accounts for 57 – 81 % of attainable disruption mitigation. Above this baseline, no agent-vs-A0 contrast attains family-wise significance in 7 of 8 DUE-on cells.

(C2) Topology-conditional optimal agent under GPS-off stress. Dense CBD (Marina + MCE) → A1 attains − 62 % RTA at H = 2.5 m DUE off ($p < 0.01$); sparse tree (KR) → same A1 yields only − 2 to − 3 % (NS).

(C3) LLM context-aware abstention. In KR DUE-off, A2 / A5 attain comparable RTA at 4 – 13× lower reroute count.

(C4) Density robustness. The Marina > KR vulnerability ratio (1.13 – 1.75) holds across 0.75× to 1.25× demand, confirming the cross-city ranking is not a calibration artefact.

7.2 Policy recommendations for Singapore

R1 — GPS / ATIS infrastructure as primary resilience layer.

R2 — Topology-matched contingency agent layer. Rule-based detours for high-redundancy CBD networks (Marina, Orchard, Bugis); LLM context-aware abstention for sparse periphery networks (Kent Ridge, Pasir Panjang).

R3 — Per-class disruption monitoring with transit-equity weighting.

7.3 Future work (Phase III)

Five Phase III items: (i) pedestrian + cycling integration; (ii) GTFS-based time-varying bus schedule replacing AM_Peak_Freq midpoint; (iii) bus dwell-time calibration against LTA Bus Arrival API; (iv) MRT-station-level disruption via LTA TrainServiceAlerts API; (v) FABDEM (Hawker et al., 2022) urban-canopy-corrected DEM substitution.

Appendix A — Data Sources and Reproducibility

A.1 Geographic data

OSM via Overpass API (24 April 2026). KR 77 km², 7,329 edges; Marina 25.5 km², 3,810 edges. MCE preserved (5 km, 420 m undersea, opened 29 Dec 2013 (ceremony 28 Dec 2013)). Copernicus GLO-30 v3 DEM, 5 points per edge, min aggregation.

A.2 Demand calibration

randomTrips.py with --period 0.25 s --end 600 s --validate. LTA DataMall TSB v3 56,785 records. KR 1,993 vehicles; Marina 2,089.

A.3 Flood scenario

Trapezoidal H(t): rise 180 → 540 s, peak 540 s, fall 540 → 900 s. $H_{peak} \in \{2.5, 5.0\}$ m. SDF Eq. 4 with k = 9 × 10■■; closure d ≥ 50 cm.

A.4 HDLF agent

Trigger 8 cm depth in 50 m lookahead. Throttle 180 s per vehicle. 3 LLM calls per simulated second cap. A2: Ollama qwen2.5:3b. A5: DeepSeek deepseek-v4-pro.

A.5 Bus multimodal extension

LTA DataMall paginated: 793 (ServiceNo, Direction) tuples; 26,741 stop-rows; 5,201 stops (6.76 MB JSON). cKDTree snap 100 m. Top 60 services / city, sumolib.net.getShortestPath(vClass="bus") max 80 edges. Period from AM_Peak_Freq midpoint. has.rerouting.device 30 s adaptation. Agent loop excludes vType="bus".

A.6 Statistical and computational environment

SUMO v1.26.0 via libsumo, Windows 11 / Python 3.11. scipy 1.11 for Wilcoxon (Eq. 10), Holm-Bonferroni FWER (Eq. 11), Cohen's d (Eq. 12), Levene W. TOST per Lakens (2017).

Appendix B — Complete Experimental Matrix Data

City	Agent	H	DUE	RTA	drop_max	reroute
KR	A0	2.5	off	139.14	0.36	0
KR	A1	2.5	off	135.00	0.36	274
KR	A1+	2.5	off	136.12	0.36	31
KR	A2	2.5	off	131.29	0.36	67
KR	A5	2.5	off	130.46	0.36	68
KR	A0	2.5	on	45.15	0.16	0
KR	A1	2.5	on	54.97	0.18	117
KR	A1+	2.5	on	54.91	0.18	12
KR	A2	2.5	on	50.20	0.17	9
KR	A5	2.5	on	50.20	0.17	10
KR	A0	5.0	off	132.58	0.30	0
KR	A1	5.0	off	129.82	0.30	623
KR	A1+	5.0	off	129.12	0.30	65
KR	A2	5.0	off	126.05	0.30	65
KR	A5	5.0	off	125.50	0.30	66

City	Agent	H	DUE	RTA	drop_max	reroute
KR	A0	5.0	on	56.19	0.18	0
KR	A1	5.0	on	59.32	0.19	349
KR	A1+	5.0	on	56.52	0.18	27
KR	A2	5.0	on	55.93	0.18	27
KR	A5	5.0	on	55.25	0.18	26
Mar	A0	2.5	off	186.17	0.51	0
Mar	A1	2.5	off	70.40	0.22	1340
Mar	A1+	2.5	off	182.41	0.50	182
Mar	A2	2.5	off	182.74	0.50	188
Mar	A5	2.5	off	181.78	0.50	190
Mar	A0	2.5	on	42.31	0.10	0
Mar	A1	2.5	on	38.48	0.11	180
Mar	A1+	2.5	on	39.83	0.11	65
Mar	A2	2.5	on	39.32	0.11	62
Mar	A5	2.5	on	37.45	0.10	60
Mar	A0	5.0	off	213.80	0.55	0
Mar	A1	5.0	off	147.71	0.42	1410
Mar	A1+	5.0	off	212.89	0.55	210
Mar	A2	5.0	off	213.54	0.55	215
Mar	A5	5.0	off	209.41	0.54	212
Mar	A0	5.0	on	91.04	0.30	0
Mar	A1	5.0	on	90.18	0.30	260
Mar	A1+	5.0	on	91.87	0.30	78
Mar	A2	5.0	on	91.32	0.30	75
Mar	A5	5.0	on	91.71	0.30	72

Table B.1 — Complete 80-row experimental matrix.

Appendix C — Code Architecture

Full implementation provided in the supplementary archive (directory: ./mcp-sumo-flood/). All paths in scripts use environment-variable indirection (PROJECT_ROOT, SUMO_HOME) and are platform-independent.

```
main.py # Simulation orchestrator (628 lines)
agent.py # HDLF agent decision pipeline (~250 lines)
baselines.py # A0 / A1 rule logic (~120 lines)
personas.py # 5-persona heterogeneity model (~180 lines)
flood.py # Synthetic flood field + SDF (~150 lines)
flood_dem.py # DEM-driven bathtub flood (~280 lines)
metrics.py # RTA / drop_max / T_r computation (~220 lines)
llm_client.py # Multi-provider LLM dispatcher (~520 lines)
tools.py # SUMO-context retrieval for LLM (~150 lines)
config.py # Centralised parameters + env-var (~200 lines)
nets/
osm_v3/ # KR network (osm.net.xml + bus_flows.rou.xml)
osm_marina/ # Marina network
dem/ # Per-edge elevation cache JSON
lta_bus/ # Full LTA DataMall paginated dump (6.76 MB)
phase2/
bus_*.csv # 40 cell × 5-seed result CSVs
bus_matrix_FINAL.csv
bus_car_RTA_split.csv
results/
tripinfo_*.xml # 200 per-vehicle simulation records
```

Appendix D — LLM Prompt Architecture

D.1 Tier 2A shortcut

```
if alt_path_length / current_route_length > 1.5:
return Decision('stay', reason='Tier 2A: alternative >50% longer')
```

D.2 Tier 2B compact LLM prompt (JSON)

```
System: You are a Singapore driver advisor. Decide reroute (R) or stay (S).
User: {
"vehicle_id": "veh_1234",
"current_edge": "634832601",
"current_speed_mps": 6.2,
"remaining_route_edges": 28,
"lookahead_flood_max_depth_cm": 12.4,
"alt_path_hops": 17,
"alt_path_max_depth_cm": 0.0,
"alt_path_extra_length_pct": 23,
"persona": "average"
}
Reply with JSON: {"action": "reroute" | "stay", "reason": "..."}
```

References

Borowska-Stefa■ska, M., Bartnik, A., Kowalski, M., Tomalski, P., & Wi■niewski, S. (2025). Assessing road network resilience and vulnerability in urban transport systems against urban flooding. Environmental Hazards. https://doi.org/10.1080/17477891.2025.2469634

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., & von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. Earthquake Spectra, 19(4), 733–752. https://doi.org/10.1193/1.1623497 [foundational 4-R framework]

Bureau of Public Roads. (1964). Traffic Assignment Manual. U.S. Department of Commerce, Urban Planning Division, Washington, D.C. [foundational BPR link performance]

Chen, Y., & Wang, J. (2019). Infrastructure resilience to flooding hazards: Methodology, gap analysis, and future directions. Sustainable and Resilient Infrastructure, 7(2), 73–96. https://doi.org/10.1080/23789689.2019.1681821

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum, Hillsdale, NJ. ISBN 978-0805802832. [foundational effect-size theory]

Da, L., Mei, X., Mostafi, R., Yang, J., Zhang, Z., & Wei, H. (2024). Open-TI: Open Traffic Intelligence with Augmented Language Model. arXiv:2401.00211.

Diakakis, M., Boufidis, N., & Sakellariou, S. (2020). A systematic assessment of the effects of extreme flash floods on transportation infrastructure and traffic networks. International Journal of Disaster Risk Reduction, 47, 101542. https://doi.org/10.1016/j.ijdrr.2020.101542

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J., Sampson, C., & Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. Environmental Research Letters, 17(2), 024016. https://doi.org/10.1088/1748-9326/ac4d4f

He, Y., Thies, S., Avner, P., & Rentschler, J. (2021). Flood impacts on urban transit and accessibility — A case study of Kinshasa. Transportation Research Part D, 96, 102889. https://doi.org/10.1016/j.trd.2021.102889

Highway Capacity Manual (HCM). (2010). Transportation Research Board, Washington, D.C.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65–70. [foundational FWER]

Hosseini, S., Barker, K., & Ramirez-Marquez, J. E. (2016). A review of definitions and measures of system resilience. Reliability Engineering & System Safety, 145, 47–61. https://doi.org/10.1016/j.ress.2015.08.006

Krauss, S. (1998). Microscopic modelling of traffic flow: Investigation of collision-free vehicle dynamics. PhD thesis, Universität zu Köln (Defended 1997; DLR Forschungsbericht 98-08, 1998. https://sumo.dlr.de/pdf/KraussDiss.pdf) [foundational SUMO car-following]

Lai, S., Xu, Z., Zhang, W., Liu, H., & Xiong, H. (2024). LLMLight: Large language models as traffic signal control agents. arXiv:2312.16044. (Accepted at KDD 2025.)

Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. Social Psychological and Personality Science, 8(4), 355–362. https://doi.org/10.1177/1948550617697177

Land Transport Authority (LTA), Singapore. (2024). Land Transport DataMall — Traffic Speed Bands API v3 / BusServices / BusStops / BusRoutes. https://datamall.lta.gov.sg/

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin et al. (eds.), Contributions to Probability and Statistics (pp. 278–292). Stanford UP.

Li, C., Wang, W., Solé-Ribalta, A., Borge-Holthoefer, J., Jia, B., Liu, Z., Yu, B., Gao, Z., & Gao, J. (2025). Adaptive capacity for multimodal transport network resilience to extreme floods. Nature Sustainability, 8(7), 741–752. https://doi.org/10.1038/s41893-025-01575-z

Li, S., Azfar, T., & Ke, R. (2024). ChatSUMO: Large language model for automating traffic scenario generation in Simulation of Urban MObility. arXiv:2409.09040.

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., & Wiessner, E. (2018). Microscopic Traffic Simulation using SUMO. In IEEE ITSC, 2575–2582. https://doi.org/10.1109/ITSC.2018.8569938

Mostafizi, A., Wang, H., Cox, D., Cramer, L. A., & Dong, S. (2019). An agent-based vertical evacuation model for a near-field tsunami. International Journal of Disaster Risk Reduction, 34, 467–479. https://doi.org/10.1016/j.ijdrr.2018.12.018

National Climate Change Secretariat (NCCS), Singapore. (2024). Third National Climate Change Study Singapore (V3 Report). https://www.nccs.gov.sg/singapores-climate-action/v3-report/

OpenAI. (2023). GPT-4 Technical Report. arXiv:2303.08774. https://arxiv.org/abs/2303.08774

Pregnolato, M., Ford, A., Wilkinson, S. M., & Dawson, R. J. (2017). The impact of flooding on road transport: A depth-disruption function. Transportation Research Part D, 55, 67–81. https://doi.org/10.1016/j.trd.2017.06.020

Pregnolato, M., Jaroszweski, D., Ford, A., & Dawson, R. (2019). Climate extremes and their implications for impact modeling in transport. In Climate Extremes and Their Implications for Impact and Risk Assessment (pp. 315–340). Elsevier.

Public Utilities Board (PUB), Singapore. (2024). Coastal Protection — Singapore's Coastal Adaptation. https://www.pub.gov.sg/Public/KeyInitiatives/Flood-Resilience/Coastal-Protection

Pyatkova, K., Chen, A. S., Butler, D., Vojinovi■, Z., & Djordjevi■, S. (2019). Assessing the knock-on effects of flooding on road transportation. Journal of Environmental Management, 244, 48–60.

Pyrialakou, V. D., Gkritza, K., & Fricker, J. D. (2016). Accessibility, mobility, and realized travel behavior. Journal of Transport Geography, 51, 252–269. https://doi.org/10.1016/j.jtrangeo.2016.02.001

Saha, S., Schramm, P., Nolan, A., & Hess, J. (2016). Adverse weather conditions and fatal motor vehicle crashes in the United States, 1994–2012. Environmental Health, 15, 104. https://doi.org/10.1186/s12940-016-0189-x

Shand, T. D., Cox, R. J., Blacka, M. J., & Smith, G. P. (2011). Australian Rainfall and Runoff Project 10: Appropriate Safety Criteria for Vehicles — Stage 2. Engineers Australia / Geosciences Australia.

Sheffi, Y. (1985). Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. Prentice-Hall, Englewood Cliffs, NJ. ISBN 978-0139397295. [foundational DUE saturation theorem]

Singapore Department of Statistics (SingStat). (2024). Singapore in Figures 2024. https://www.singstat.gov.sg/

USACE NFIP TB-2. (1993). Vehicle Stability Limits in Floodwaters. U.S. Army Corps of Engineers Technical Bulletin 2.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics Bulletin, 1(6), 80–83. [foundational paired-rank test]

Yin, J., Yu, D., Yin, Z., Liu, M., & He, Q. (2016). Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in Shanghai. Journal of Hydrology, 537, 138–145.

Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A review of machine learning and IoT in smart transportation. Future Internet, 11(4), 94. https://doi.org/10.3390/fi11040094