Check for updates

RESEARCH ARTICLE

REVISED Range-risk estimation for electric vehicles in low-coverage mountain road segments: A digital-twin study with prequential beacon-based Federated Learning

Yeray Mezquita[1], Jesús Emmanuel Vidal Cuevas[2], Diego Valdeolmillos[1], Albano Carrera[1]

[1]Deep Tech Lab, AIR Institute, Salamanca, Castilla y León, 37003, Spain
[2]Universidad Politecnica de Pachuca, 43830 Zempoala, Hidalgo, Mexico

v2 **First published:** 04 Aug 2026, **6**:275.1
https://doi.org/10.12688/openreseurope.24549.1
Latest published: 09 Nov 2026, **6**:275.2
https://doi.org/10.68336/mrn9fq87

Abstract

Electric vehicles operating on isolated mountain roads may receive over-optimistic range estimates precisely where connectivity and charging alternatives are scarce. This paper proposes a beacon-centric Vehicle-to-Everything and Federated Learning architecture for range-risk estimation in low-coverage mountain-road segments. Entry and exit beacons maintain a local segment-specific model, broadcast it to approaching vehicles, and update it from traversal evidence contributed by vehicles after leaving the segment, without requiring raw telemetry to be centralised. The approach is evaluated in a digital twin of a 10.6 km fragment of the Puerto de la Quesera mountain pass using 12 simulated traversals from two electric-vehicle models under dry, wet and snow/low-temperature conditions. This study presents a proof of concept: its scale demonstrates the mechanism but not population-level accuracy. Road grade is derived from vehicle pitch; against the digital-twin mesh geometry, the average resulting error is 0.36% of route energy and the worst error is 0.74%. The evaluation is route-level, matching the operational decision made before entering the segment. Generic estimators systematically underestimate the energy required to complete the section: the nominal-consumption estimator produces a mean absolute error of 0.777 kWh, while a conventional non-local physics estimator produces 1.975 kWh. A segment-specific residual model reduces the leave-one-route-out error to 0.220 kWh. More importantly, the proposed alert-oriented evaluation shows that false-safe decisions are reduced from 14.58% for the non-local estimator and 8.33% for the nominal estimator to 1.04% for the segment-specific residual model. In the prequential beacon simulation, after eight previous vehicles, FedProx reduces the false-safe rate to 3.38%. Cross-vehicle generalisation remains the main limitation: the leave-one-vehicle-out error rises to

Open Peer Review

Approval Status ? ?

	1	2
version 1 04 Aug 2026	? view	? view

1. **Ahmed Ihsan Simsek**, Fırat Üniversitesi, Elazığ, Turkey

2. **Maksymilian Mądziel**, Rzeszów University of Technology, Rzeszów, Poland

Any reports and responses or comments on the article can be found at the end of the article.

1.568 kWh, and both vehicles share one energy model with different parameters. These results support the feasibility of local, privacy-preserving range-risk intelligence for challenging road segments.

Keywords
Electric vehicles; range estimation; federated learning; mountain roads; digital twin; V2X; beacon networks.



This article is included in the Horizon Europe gateway.

Corresponding author: Albano Carrera (acarrera@air-institute.com)

Author roles: Mezquita Y: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Original Draft Preparation; **Vidal Cuevas JE**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation; **Valdeolmillos D**: Conceptualization, Funding Acquisition, Investigation, Project Administration, Resources, Supervision, Visualization, Writing – Review & Editing; **Carrera A**: Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Project name "A catalyst for EuropeaN ClOUd Services in the era of data spaces, high-performance and edge computing (NOUS)", Grant Agreement Number 101135927. Funded by the European Union.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Mezquita Y *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Mezquita Y, Vidal Cuevas JE, Valdeolmillos D and Carrera A. **Range-risk estimation for electric vehicles in low-coverage mountain road segments: A digital-twin study with prequential beacon-based Federated Learning** Open Research Europe , **6**:275.2 https://doi.org/10.68336/mrn9fq87

First published: 04 Aug 2026, **6**:275.1 https://doi.org/10.12688/openreseurope.24549.1

REVISED **Amendments from Version 1**

The revised version 2 addresses the reviewers' comments. No new experiments were conducted, and no previously reported results were changed. All newly added quantitative results were derived from the twelve traversal logs previously published in Version 1.

Text. The proof-of-concept scope, the origin of the road-grade signal, and the limitations of cross-vehicle generalisation are now specified in the abstract. In the Introduction and Section 2.6, homologated (WLTP) range is distinguished from dynamic on-board range estimation. In Section 3.3.2, the pitch-derived road grade is quantitatively validated against the digital-twin mesh geometry, yielding a mean absolute difference of 0.039 degrees and route-energy differences of 0.36% on average and 0.74% at worst. The unfiltered nature of the pitch signal is clarified, and its residual relative to longitudinal acceleration is decomposed. In Section 3.4.1, "Coulomb counting" is corrected to "energy-based integration". The shared longitudinal model structure used for both vehicles is also highlighted, clarifying that the results evaluate the methodology rather than production-vehicle behaviour. This limitation is reiterated in Sections 4.1 and 5. Section 3.5 describes the aggregation of federated results. Section 4.5 specifies the number of independent observations underlying the alert decisions and explains the non-local estimator's zero risk recall. In Section 3.3.1, the tool name is corrected to *mapperV3*.

Tables, figures, and references. Two tables have been added: a comparison of centralised and federated models using common metrics (Section 4.4, Table 9), and a sensitivity analysis of alert results with respect to the reserve threshold (Section 4.5, Table 12). A FedAvg false-safe column has also been added to Table 11.

Data and code. In Section 6, the descriptions of the archived record, logged variables, and their units have been corrected. The software repository remains unchanged.

Any further responses from the reviewers can be found at the end of the article

1. Introduction

Electric vehicles (EVs) are increasingly deployed on road networks that combine dense, well-connected corridors with extensive stretches of sparse infrastructure, low traffic flow, and harsh or rapidly changing environmental conditions (Jung et al., 2023). The former basically encompasses motorways, urban areas, and well-connected suburbs, while the latter concerns mountain passes, rural valleys, forest roads, or cross-border links. Road segments that simultaneously combine complex orography, variable micro-climates, and intermittent or absent cellular coverage are termed here *challenging segments*: on such sparsely connected segments, both safety-critical perception and energy-aware planning become more difficult (Prasad and Gudipalli, 2025). Two distinct quantities are involved here and are worth separating. The rated range of a vehicle is derived from a standardised type-approval procedure such as the Worldwide-Harmonized Light Vehicles Test Procedure (WLTP (United Nations Economic Commission for Europe, 2022)), whose laboratory conditions are known to diverge systematically from real-world consumption, with reported gaps of the order of 15–20% on average and substantially larger on demanding terrain or in cold weather (Weiss et al., 2024; Yuan et al., 2024). The remaining range displayed on board is a different quantity, which is produced by a dynamic estimator; contemporary production implementations may combine a rated-consumption baseline with recent energy consumption, predicted speed, ambient conditions and, when navigation is active, route-related information, and the strategy may differ between the displayed range and the route-specific consumption forecast. The concern addressed in this work is therefore not that on-board estimators implement type-approval data directly, in fact they generally do not, but that on a low-coverage segment every input that would specialise such an estimator to that segment is degraded at once: a type-approval baseline is generic by construction, a recent-consumption average reflects the road already driven rather than the climb ahead, and route-related information served from the cloud is unavailable precisely where cellular coverage is lost. Micro-climatic phenomena (fog in a valley, ice in shaded curves, sudden snowfall, extreme temperatures) and local terrain (steep grades, tight turns) can degrade on-board sensor performance and invalidate range estimators that rely on standardised cycles or coarse-grained maps (Mao et al., 2025); intermittent connectivity additionally undermines range estimators that depend on cloud services and frequent end-edge-cloud exchanges (Mulpuri et al., 2025). As a result, drivers may receive outdated or over-optimistic information about remaining range and road conditions precisely where mistakes are less forgiving (Ghazali et al., 2025).

Current Vehicle-to-Everything (V2X) and connected-vehicle architectures are largely optimised for urban or highway environments with continuous cellular coverage and high traffic density (Clancy et al., 2024; Wang et al., 2025). In these settings, vehicles can offload perception tasks to the cloud and benefit from dense probe data fed by cooperative perception pipelines (Wan et al., 2025); frequent over-the-air updates and high-capacity downlinks are similarly enabled by wide-area cellular and emerging 6G V2X technologies (Noor-A-Rahim et al., 2022). In contrast, low-coverage segments often fall back to purely on-board intelligence: models are rarely updated once deployed, and any local knowledge acquired by one vehicle (e.g., unexpected ice, wildlife crossings, temporary lane closures) is not systematically shared with subsequent vehicles. Conventional approaches therefore underutilise the potential of V2X to create a

local collective intelligence around difficult road sections, and they offer limited guarantees on how perception and energy models evolve over time in such conditions (Maglogiannis et al., 2022).

The V2X system presented in this paper addresses the mentioned gap by introducing a beacon-centric architecture tailored to low-coverage, safety-critical road segments. Instead of assuming continuous wide-area connectivity, a small number of physical beacons are deployed at the entry and exit of the section of interest. Whenever direct line-of-sight between the entry and exit beacons is available, no further infrastructure is required to keep the two endpoints synchronised; otherwise, the link between them is closed by a small chain of intermediate point-to-point repeaters that act purely as backbone relays and do not participate in V2X exchanges with vehicles. The two end beacons maintain reliable backhaul (when available) and act as local hubs for V2X communication, a role increasingly recognised for roadside units in modern intelligent transportation systems (Bilal et al., 2023; Gong et al., 2023). These beacons broadcast up-to-date summaries of road and weather conditions, receive measurements and model updates from vehicles traversing the segment, and coordinate how this knowledge is aggregated and redistributed. In this way, vehicles entering the section can adapt their perception pipelines, driving strategies, and range estimates in advance, while vehicles leaving the section contribute fresh evidence to improve the system for others.

To make this interaction efficient and privacy-preserving, the proposed architecture couples the beacon infrastructure with an edge Artificial Intelligence (AI) and Federated Learning (FL) pipeline. Lightweight time-series and vision models run directly on vehicles and beacons to forecast energy consumption and risk signals and to interpret local perception cues. Rather than streaming raw sensor data, clients participate in FL rounds in which model updates are aggregated first at beacons and, when possible, forwarded to a cloud server. This hierarchical, communication-aware design is well aligned with vehicular non-Independent and Identically Distributed (non-IID) data distributions, intermittent links, and strict latency constraints (Chellapandi et al., 2024; Zhou et al., 2023); the same design enables continuous adaptation of models to terrain, micro-weather, and traffic conditions without exposing raw, potentially sensitive data beyond the local environment (Du et al., 2020; Zhang et al., 2024).

A central focus of the architecture is context-aware range-risk estimation. The system explicitly incorporates segment-level context, including road geometry, micro-weather, pavement state and recent traversal evidence, recognising their strong influence on EV energy consumption and on the reliability of the information available to drivers (He et al., 2024; Weiss et al., 2024). In the proposed architecture, the same beacon mechanism can also disseminate safety-critical perception cues, such as black ice, debris, wildlife or sudden visibility loss. However, the experimental validation reported in this paper focuses specifically on EV range-risk estimation: reducing unexpected range depletion and avoiding over-optimistic final State of Charge (SOC) predictions before a vehicle enters a connectivity-poor section.

The architecture proposed here is developed within the NOUS reference architecture, which provides the roles, agents, and common services on which the deployment is built. Vehicles and beacons are modelled as Data Space Participants (DSPs) instantiated through the NOUS Data Agent (NDA), while the regional/cloud backend is conceived as a Data Space Infrastructure Provider (DSIP) instantiated through the NOUS Infrastructure Agent (NIA); identity, authorisation, governance, and discovery are delegated to the NOUS common services (Table 1). This alignment follows the broader principles of sovereign data spaces, in which participants retain control over their data and policy-driven sharing replaces centralised data lakes (Jarke et al., 2019; Otto, 2022).

Table 1. System components instantiated by the beacon-centric V2X/FL system: roles, agents, and responsibilities of each component within the NOUS reference architecture.

Component	Role	Agent	Responsibility
Vehicle (on-board inference + FL client)	DSP	NDA	Keeps raw data locally; performs inference; produces evidence summaries and FL updates; publishes/consumes data/services under policy control.
Entry/Exit Beacon (local hub + mini-aggregator + broadcast)	DSP (opt. DSIP)	NDA (opt. NIA)	Aggregates updates locally (hierarchical FL); disseminates alerts/summaries; optionally exposes edge compute/network as managed infrastructure.
Intermediate Repeater (connectivity relay)	DSIP	NIA	Provides policy-aware networking/relay and connectivity between far-edge and regional/cloud resources.

Table 1. *Continued*

Component	Role	Agent	Responsibility
Regional/Cloud Backend (sync + model/artifact repository)	DSIP	NIA	Hosts model/artifact repositories; optional higher-capacity aggregation/analytics; provides infrastructure/platform services for deployment and synchronisation.
Identity, Authorisation, Governance	Common Services	—	Identity and Access Management (IAM), certification/attestation hooks, policy decision/enforcement points, and service gateway for controlled access.
Discovery (agents/apps/models)	Common Services	—	Registries/catalogues enabling discovery of participants, services, and reusable applications.
Federated Learning support	DSP + Common Services	NDA	FL orchestration and lifecycle at participant level (NDA), complemented by cross-space coordination services when needed.

Within this context, the contributions of this paper are threefold:

- A beacon-based V2X architecture for low-coverage road segments is proposed, integrating on-board sensing, roadside context probes, and reliable edge nodes at the section boundaries. The design supports both safety-oriented alerts and energy-aware planning while remaining deployable in infrastructure-scarce environments.

- An edge AI and federated learning pipeline is designed to treat vehicles and beacons as FL clients, support hierarchical and asynchronous aggregation, and incorporate communication-efficient and privacy-preserving mechanisms. The pipeline is tailored to EV energy prediction and risk-aware perception under non-IID, spatio-temporally evolving data.

- Through a digital-twin evaluation of a real mountain-pass segment, the proposed system is shown to reduce route-level range-estimation error and, more importantly, false-safe autonomy decisions when compared with generic or non-local estimators. Communication-layer and perception-layer KPIs are identified as future deployment targets.

The remainder of this article is organised as follows. Section 2 reviews the state of the art in sensing and context modelling, edge AI, federated learning, V2X communication, embedded hardware, and grade- and weather-aware energy estimation. Section 3 presents the proposed V2X architecture, detailing the beacon deployment, data flows, learning protocols, and implementation choices. Section 4 reports the experimental setup and main results, and Section 5 discusses the implications and limitations of the proposed approach and outlines directions for future work.

2. State of the Art

The proposed beacon-centric V2X and federated learning architecture for low-coverage road segments draws on advances in seven complementary research areas. The first concerns sensing and context inputs that feed both perception and energy models; the second and third deal with edge AI for immediate inference and with federated learning protocols suitable for vehicular networks. The fourth pillar is the V2X communication substrate that determines how information actually flows between vehicles, beacons, and the cloud, while the fifth focuses on the embedded hardware that must execute models within strict thermal and latency envelopes. The sixth area gathers the literature on terrain, weather, and environmental factors that drive electric-vehicle energy consumption and dictate the difficulty of any range estimator. The final pillar collects the datasets, simulators, and benchmarks needed to evaluate such systems reproducibly. This section is aimed at reviewing each of these pillars. The experimental contribution of the present paper focuses on the energy and range-risk branch of this architecture; perception, communication-latency and embedded-hardware KPIs are discussed as deployment requirements but are not measured in the current experimental batch.

2.1. Sensing and context inputs

A connected vehicle approaching a low-coverage segment relies on a heterogeneous mix of on-board, roadside, and geospatial signals. On-board sensing typically combines a camera, LiDAR (Light Detection and Ranging) or radar suite with inertial and satellite navigation and with vehicle-bus telemetry such as speed, torque, state of charge, and auxiliary power draw. Recent multi-modal perception studies show that combining these modalities is essential for reliable scene understanding, especially under adverse weather and low illumination, but they also highlight that on-board perception alone degrades in occluded or cluttered environments and benefits substantially from cooperative sources of evidence

(Wan et al., 2025; Xu et al., 2023). Roadside and beacon signals complement vehicle telemetry with measurements of temperature, humidity, pavement wetness, micro-weather, infrastructure vibration, and traffic counts; cooperative perception datasets that include infrastructure-mounted sensors confirm that such signals improve detection range and reliability in vehicle-infrastructure scenarios (Yu et al., 2022).

Geospatial context is equally important. Digital Elevation Models (DEMs), grade and curvature profiles, lane geometry, and segment topology are needed to interpret on-board signals at the level of individual road links. High-definition map pipelines now routinely fuse DEMs with vector road geometry, performing semantic and geometric annotation, lane-level segmentation, and topology extraction prior to deployment (Bao et al., 2023). The corresponding data-preparation chain in the present setting includes time-and-space alignment of heterogeneous streams, map matching of Global Navigation Satellite System (GNSS) traces, segmentation of road links, and outlier filtering, which are well-established preprocessing steps in vehicular sensor fusion (Wan et al., 2025).

A recurring theme across this literature is the trade-off between raw data and lightweight evidence summaries. Streaming raw sensor data over a constrained backhaul is rarely feasible in low-coverage contexts, so beacons and vehicles must instead exchange compact descriptors, such as event flags, statistical summaries, model updates, that preserve the relevant information while respecting bandwidth budgets at the edge (Gong et al., 2023); model-update exchanges further protect privacy by keeping raw data on the vehicle and the beacon (Du et al., 2020).

2.2. Edge AI models for immediate inference

For immediate decisions under limited connectivity, models must be both data-efficient and hardware-efficient. Two complementary families dominate: compact time-series predictors for energy/risk signals and lightweight vision backbones for scene cues, with both ultimately compressed and distilled to run on in-vehicle Electronic Control Units (ECUs) or roadside beacons (H.-I. Liu et al., 2024).

On the energy side, recent EV studies have moved beyond purely physics-based estimators towards data-driven predictors that ingest short windows of speed/acceleration, grade, temperature, and auxiliary loads. Large comparative evaluations in 2025 Scientific Reports show that gradient-boosted trees and light neural regressors can outperform legacy baselines for EV energy prediction when trained on real telemetry, highlighting the competitiveness of tabular learners (XGBoost/LightGBM) under tight latency and memory budgets (Hussain, Ching, Uttraphan, Tay, and Noor, 2025a; Hussain, Ching, Uttraphan, Tay, Noor, and Memon, 2025b). In these works, accuracy is reported with Mean Absolute Error (MAE)/Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) over trips or segments, and practical gains come from feature engineering that makes road grade and micro-weather explicit, which maps well to the beacon setting considered here, where such context is available locally (He et al., 2024).

Neural time-series models remain attractive when temporal dynamics must be learned directly on the edge. Temporal Convolutional Networks (TCNs) provide long receptive fields through dilations with fewer parameters than recurrent models, making them strong candidates for embedded inference; current surveys emphasise their efficiency and stability for forecasting tasks, where evaluation typically couples MAE/MAPE with horizon-wise error profiles to expose degradation at longer look-ahead times (Kong et al., 2025; Torres et al., 2021). Emerging edge-sized transformers extend this line, and domain-specific architectures such as an Energy Consumption Transformer for EVs indicate that attention mechanisms can capture heterogeneous, non-IID operating regimes (e.g., climbs vs. descents) when modest model sizes are affordable (Huang et al., 2025); the corresponding ablations usually relate error reductions to parameter count and window length, of direct relevance when setting beacon memory budgets (Ngo et al., 2025).

When perception cues matter, such as detecting low-friction surfaces, pooled water, or lane-level context, compact Convolutional Neural Networks (CNNs) remain the workhorse. Comparative studies through 2025 find MobileNetV3 and EfficientNet-Lite/V2 to offer the best accuracy-throughput trade-offs on edge Systems-on-Chip (SoCs), while SqueezeNet-class models excel when footprint and cold-start are dominant (H.-I. Liu et al., 2024). Inference quality is commonly measured with mean Average Precision (mAP)/F1, but for edge deployment the decisive metrics are end-to-end latency and energy per inference (reported in milliseconds and joules); both improve further with early-exit networks that terminate computation once confidence thresholds are met (Rahmath P et al., 2024). These multi-exit designs are particularly useful on beacons that must broadcast within strict deadlines while running from limited power.

Across modalities, the deployment path converges on TinyML techniques: post-training quantisation to int8 or lower, structured/unstructured pruning, and knowledge distillation from a richer cloud model into a compact student that lives on the vehicle or beacon (Somvanshi et al., 2025). The edge-inference literature documents sizeable latency and energy gains from the combination of quantisation-aware training, pruning, and hardware-aware neural architecture search, with

negligible accuracy loss when distillation is used (Chitty-Venkata et al., 2023); reported speedups are accompanied by retained accuracy deltas (Δ% top-1 or ΔMAE) and energy savings on representative Neural Processing Units (NPUs)/ Central Processing Units (CPUs) (Ngo et al., 2025). This forms the basis for a practical toolchain in the present setting, training centrally on pooled data when available, distilling to small students, then continuing on-device fine-tuning through federated rounds (Chellapandi et al., 2024).

Finally, running models where the data originates is now standard in vehicular edge architectures. Contemporary V2X frameworks and Roadside Unit (RSU)/roadside-edge studies argue for placing perception and forecasting models at the network edge to meet tight latency budgets (Gong et al., 2023); the underlying edge Deep Neural Network (DNN) inference techniques that make such deadlines achievable on resource-limited hardware are catalogued in recent reviews (Ngo et al., 2025). Reported evaluations often pair task metrics with system-level measures such as time-to-alert and packet-loss-robust latency. These metrics are relevant to the full deployment, but the present experimental batch reports only route-level energy and range-risk indicators. In short, the state of the art supports a pipeline in which compact predictors estimate energy/risk on local signals and are then updated through federated mechanisms, which are detailed next (McMahan et al., 2017).

2.3. Federated learning for vehicular edge

In vehicular edge settings with intermittent connectivity, vehicles and beacons are treated as FL clients, with beacons also acting as mini-aggregators when backhaul is scarce. The canonical baseline remains FedAvg, averaging local updates in rounds and establishing the modern FL paradigm of training where data stays local, its convergence and communication behaviour are taken as a reference point for the variants discussed next (McMahan et al., 2017). Building on FedAvg, FedProx adds a proximal term that improves stability under non-IID and system heterogeneity, both unavoidable when different road segments and vehicle types produce skewed data (Li et al., 2020), while SCAFFOLD controls client drift via control variates, often yielding faster convergence for the same communication budget (Karimireddy et al., 2020). In practice, these choices are evaluated not only by final accuracy but by convergence per round and per-round bytes, since these determine how often a beacon must seize scarce spectrum to ship updates. Server-side adaptive optimisers (FedAdam/FedYogi) frequently reduce the number of rounds needed to reach a target error, a property critical when clients are only sporadically reachable (Reddi et al., 2021).

Strict synchrony is fragile on the road, so asynchronous and hierarchical FL are natural design choices for the target deployment. Recent schemes down-weight stale updates, buffer versions, or correct version skew to preserve accuracy when clients connect at random times, which is a natural fit for cars and beacons that communicate only at segment entry or exit points (Lu et al., 2024; Zang et al., 2024). To further localise traffic and reduce backbone pressure, Hierarchical FL (HFL) can be implemented with vehicles pushing updates to a nearby beacon, beacons aggregating locally, and regional/ cloud resources synchronising only when available (Gao et al., 2025; Herabad, 2023). Because vehicles and road segments differ, personalised FL is also relevant: parameter-efficient fine-tuning or small client-side adapters can retain a compact global backbone while allowing each vehicle to specialise the model with private state such as payload, degradation or driving style (Yang et al., 2024, 2025).

Communication efficiency is pivotal in this setting. Quantisation schemes such as QSGD reduce payload sizes by orders of magnitude (Alistarh et al., 2017), while sign-based updates such as signSGD push the idea to one-bit gradients (Bernstein et al., 2018); combined with error-feedback, these compressors can approach the convergence of uncompressed Stochastic Gradient Descent (SGD), making accuracy per transmitted byte a meaningful design criterion at the beacon. Over-the-air aggregation (AirComp) and quality-aware client selection are also relevant in wireless links shared with safety traffic (Choi et al., 2025; Xiao et al., 2024). These mechanisms are not evaluated in the present digital-twin batch; they are deployment options for reducing airtime once the range-risk learning task has been validated.

Trust and safety complete the picture. Byzantine-robust aggregation via coordinate-wise median or trimmed-mean mitigates poisoned or faulty clients with provable statistical guarantees (Yin et al., 2018), and trust-bootstrapped rules such as FLTrust extend this protection to more aggressive adversaries (Cao et al., 2021). Secure Aggregation protects client updates in transit such that only their sum is revealed (Bonawitz et al., 2017), and Differentially Private SGD (DP-SGD) bounds information leakage from any single participant at the cost of utility loss (Abadi et al., 2016). In the present paper, these mechanisms are considered deployment requirements for a future field implementation rather than evaluated components. The experimental focus is restricted to range-risk accuracy, false-safe decisions and the feasibility of the vehicle-as-client, beacon-as-aggregator learning workflow. Finally, V2X deployments must align with Public Key Infrastructure (PKI)-based credentialing (IEEE 1609.2/ETSI TS 103 097), including pseudonymous certificates and replay protection for beacon-vehicle messages (ETSI Technical Committee ITS, 2017).

2.4. V2X communication substrate

The communication substrate of a beacon-centric V2X deployment must reconcile two competing needs: ultra-low-latency dissemination of safety alerts and best-effort delivery of bulkier model updates and telemetry. Three families of technologies dominate the design space. IEEE 802.11p/Dedicated Short-Range Communications (DSRC) and its European counterpart ITS-G5 provide ad hoc, decentralised vehicle-to-vehicle and vehicle-to-infrastructure links with deterministic medium access and low latency at short range, while cellular V2X (C-V2X) over the PC5 sidelink and over the Uu interface offers higher capacity, scheduled transmissions, and integration with operator networks. Comprehensive surveys of wireless V2X access conclude that no single technology dominates in all regimes (Clancy et al., 2024): experimental deployments on real highways have shown that C-V2X PC5 generally extends communication range, whereas IEEE 802.11p tends to deliver lower latency in low-density scenarios (Maglogiannis et al., 2022). The advent of 5G New Radio (NR) V2X and the longer-term 6G roadmap promise sub-10 ms latencies, flexible numerologies, and tight integration with edge computing, but require either dense base-station coverage or sophisticated sidelink resource allocation, precisely what is missing on low-coverage segments (Noor-A-Rahim et al., 2022).

The communication patterns relevant to the beacon-centric architecture proposed here follow naturally from these technologies. Beacons broadcast safety-critical summaries and alerts locally, exploit opportunistic uplinks to aggregators when backhaul is available, and rely on store-carry-forward and gossip-style dissemination among neighbouring beacons to bridge connectivity gaps. Recent work on beacon-oriented emergency message dissemination demonstrates that cluster-based broadcast schemes can substantially improve coverage and packet delivery on sparse highway scenarios (Bilal et al., 2023). Quality-of-Service (QoS) management is equally critical: safety messages must be prioritised over background federated-learning traffic, and channel access must respect duty-cycle and congestion-control constraints to avoid degrading the cooperative perception channel during peak load. Edge-intelligence surveys further argue that placing inference and aggregation at the network edge is necessary to satisfy these QoS budgets, since round-trips to a distant cloud cannot meet the deadlines of cooperative warnings (Gong et al., 2023).

2.5. Edge/Embedded hardware

The hardware substrate of a beacon-centric V2X deployment spans two heterogeneous tiers. In-vehicle compute is increasingly dominated by automotive SoC families that integrate Graphics Processing Units (GPUs) and dedicated neural processing units alongside heterogeneous ARM and x86 cores, with platforms in the NVIDIA Drive, Qualcomm Snapdragon Ride, NXP, Renesas, and TI families competing on Tera Operations Per Second (TOPS) per watt and on certifiable real-time behaviour. Recent surveys of edge intelligence in intelligent transportation systems (Gong et al., 2023), complemented by reviews of on-device deep-learning inference in resource-limited environments (Ngo et al., 2025), document that NPU-class accelerators are now able to sustain perception and forecasting workloads with sub-50 ms latency while remaining within strict thermal envelopes, provided that models are appropriately compressed. At the roadside tier, beacons and RSUs typically rely on rugged ARM or x86 edge boxes augmented with lightweight NPUs or Tensor Processing Units (TPUs), such as Coral, Hailo, ARM Ethos-U class, that allow continuous inference on power budgets compatible with solar or grid-tied installation (Somvanshi et al., 2025).

The deployment toolchain converges around standard inference runtimes. ONNX Runtime, NVIDIA TensorRT with INT8 quantisation, Apache TVM, and Intel OpenVINO are routinely used to compile models trained in PyTorch or TensorFlow into hardware-specific kernels, where transformer-oriented inference toolchains report further gains through coordinated quantisation, pruning, and operator fusion (Chitty-Venkata et al., 2023); comparative studies confirm substantial latency and throughput improvements on embedded targets across DNN families (Ngo et al., 2025). Above the runtime layer, edge containerisation frameworks such as K3s and CRI-O simplify the lifecycle management of perception and federated-learning workloads on beacons, supporting rolling updates and policy-driven deployment. Across both vehicle and beacon tiers, design targets are dictated by the cooperative-perception use case: end-to-end inference within tens of milliseconds, thermal envelopes compatible with passive or limited active cooling, and ruggedisation against the vibration, dust, and temperature ranges typical of unattended roadside installations (Gong et al., 2023).

2.6. Terrain, weather and environmental context for EV energy

Electric-vehicle range cannot be modelled adequately without explicit terrain and weather context. Grade-aware modelling exploits DEM-derived slope and curvature profiles, elevation accumulation per segment, and lane-level geometry to capture the asymmetric energy footprint of climbs and descents, including the partial recovery offered by regenerative braking. Empirical studies on real road networks confirm that road type and grade strongly modulate the energy demand of battery-electric vehicles, so range estimators that ignore them systematically under- or over-predict consumption on mountain passes and undulating rural roads (Jonas et al., 2022; Weiss et al., 2024). Micro-weather adds a second dimension of variability: ambient temperature, wind, precipitation, and pavement condition jointly determine

rolling and aerodynamic resistance and the auxiliary load drawn by cabin heating or air conditioning. Recent multi-perspective analyses of EV energy consumption based on large field data sets identify temperature, driving conditions, and route grade among the most influential factors and quantify their causal contribution to per-trip energy use (He et al., 2024; Mao et al., 2025).

These dependencies also explain why naive range estimators, such as linear functions of state of charge, simple averages over recent trips, or values inherited from type-approval cycles, including WLTP (United Nations Economic Commission for Europe, 2022) are insufficient for safety-critical planning. This is not to suggest that production on-board estimators are limited to these strategies; contemporary implementations combine several of them and may additionally exploit navigation-derived route information when a destination is set. The point is that all of these inputs are generic with respect to the segment, and that on a low-coverage segment the route-related information that would specialise them is missing. Empirical evaluations of EV consumption report systematic divergences between WLTP-declared values and real-world energy use, with average gaps of 15–20% that can grow substantially on demanding orography or in cold weather (Weiss et al., 2024; Yuan et al., 2024). A vehicle that enters a steep, cold, low-coverage segment with an over-optimistic range prediction may exhaust its battery in a location where neither cellular coverage nor charging infrastructure is available. Recent reviews of data-driven and hybrid range models argue for explicit incorporation of grade, weather, and driving-style features, and for the use of machine-learning pipelines that can be retrained as new traversals accumulate evidence (Mao et al., 2025; Prasad and Gudipalli, 2025).

The beacon-centric architecture proposed here is designed to exploit precisely this need by providing the local context and the federated-learning rounds required to keep such models aligned with the actual conditions of each segment.

2.7. Datasets, simulators and benchmarks

Reproducible evaluation of vehicular FL and V2X architectures relies on a layered ecosystem of simulators, datasets, and benchmarks. Microscopic traffic simulation is dominated by SUMO, whose recent releases support large-scale inter-modal scenarios, signal control, and direct co-simulation with vehicle dynamics and energy models (Lopez et al., 2018). For sensor-rich autonomous-driving research, CARLA remains the de-facto open simulator, providing flexible sensor suites, weather and lighting control, and hooks for cooperative perception experiments (Dosovitskiy et al., 2017). Synthetic V2V and V2X cooperative-perception datasets such as those built on top of CARLA, together with real-world counterparts including DAIR (Yu et al., 2022)-V2X and V2V4Real (Xu et al., 2023), have become standard testbeds for cooperative 3D detection and tracking under vehicle-infrastructure constraints. EV-specific co-simulation usually combines a traffic simulator with a physics-based or simplified powertrain model that ingests speed, acceleration, grade, and ambient temperature traces to compute energy use along a trajectory (He et al., 2024).

On the federated-learning side, scalable benchmarking platforms such as FedScale provide realistic client distributions, large-scale simulation runtimes, and reference implementations of canonical FL algorithms, and have been extended with vehicular-flavoured partitions that split data by client geography, time, or season to approximate the non-IID conditions encountered on the road (Lai et al., 2022). Surveys of FL in connected and automated vehicles note, however, that grade-aware splits and explicit V2X communication constraints (intermittent links, prioritised safety traffic, bounded duty cycle) are still under-represented in public benchmarks, leaving a clear gap for evaluations that jointly stress federated optimisation, embedded inference latency, and energy-aware range prediction (Chellapandi et al., 2024; Zhang et al., 2024).

Closing this gap is one of the motivations of the present work. The evaluation combines a real mountain-pass scenario with digital-twin EV traversals and simulated federated learning in order to study the energy and range-risk dimension of a deployment regime that is currently under-represented in standard benchmarks.

3. Methods and Materials

The proposed beacon-centric V2X/FL architecture is evaluated through a digital-twin simulation of a representative low-coverage mountain segment. This section describes, in order, the case-study scenario and its justification (Section 3.1), the beacon-centric architecture and its operational protocol (Section 3.2), the digital-twin pipeline that generates the simulation environment (Section 3.3), the simulated fleet and data acquisition workflow (Section 3.4), the energy-estimation methods compared in the experimental evaluation (Section 3.5), and the evaluation protocol with the Key Performance Indicators (KPIs) actually reported in the experimental results (Section 3.6).

The methodology is deliberately scoped to range-risk estimation rather than to the full V2X stack. Communication-layer KPIs (such as event-to-adaptation latency or message overhead) and perception-layer KPIs (such as risk-detection F1-score) are not produced by the present experimental batch; they would require dedicated perception and event-

propagation data, an instrumented V2X stack, and embedded hardware measurements, none of which are part of the current digital-twin configuration. These remain as future evaluation targets and are not included in the KPI set of Section 3.6.

3.1. Case study and scenario

Real-world applicability is an essential validation criterion for this type of architecture. The selected scenario is the *Puerto de la Quesera* mountain pass, located at 1712 m above sea level in the eastern Spanish Central System, within the Ayllón mountain range. The pass uniquely connects two villages separated by approximately 35 km, with sections featuring slopes exceeding 10% and no continuous mobile coverage along the route (Figure 1, left). The surrounding region has a population density below 0.5/km^2 over an area of 269 km^2 and is characterised by harsh winter conditions. The combination of complex orography, sparse traffic, micro-climatic variability and absent cellular coverage matches exactly the definition of *challenging segment* introduced in Section 1.

For tractable simulation, a 10.6 km fragment of the southern flank of the pass was selected (Figure 1, right; Figure 2). The fragment is contained within a Digital Elevation Model (DEM) bounding box of approximately 3.86 km $\times$ 2.35 km (9.08 km^2), covering elevations between 1203 m and 1798 m above sea level (a total relief of 595 m). This 595 m figure is the relief of the DEM bounding box, which covers the whole surrounding terrain; the elevation band spanned by the road itself is narrower, 1242.4–1493.5 m, as detailed in Section 3.3. The fragment was chosen for two reasons: it offers potential line-of-sight between its endpoints, enabling a direct beacon-to-beacon link without intermediate point-to-point repeaters, and it concentrates the orographic and climatic features that make the segment energetically demanding for an Electric Vehicle (EV).

3.2. Beacon-centric architecture and operational protocol

Two physical beacons are deployed at the entry and exit of the segment. Both beacons maintain reliable backhaul to a regional or cloud backend when available, and act as local hubs for V2X exchanges. When direct line-of-sight is feasible between the two beacons, as is the case for the selected fragment, no additional infrastructure is required to keep them synchronised. Otherwise, a small chain of intermediate point-to-point repeaters closes the backbone link without acting as V2X stations themselves; this keeps the certification scope, energy budget, and deployment cost of the intermediate nodes to a minimum.

The operational protocol of the architecture is defined by three interactions between vehicles and beacons.

1. **Pre-entry model delivery.** A vehicle approaching the entry beacon receives a compact segment-specific model descriptor before crossing the cellular shadow. This descriptor may include the current residual-model parameters, model version, feature schema, normalisation metadata, recent environmental summaries and any safety-critical alerts contributed by previous traversals. The beacon does not need to compute the final range value for the vehicle; instead, the vehicle evaluates the received model locally using private information such as usable battery capacity, entry SOC, battery state, payload estimate and driver-specific consumption profile.

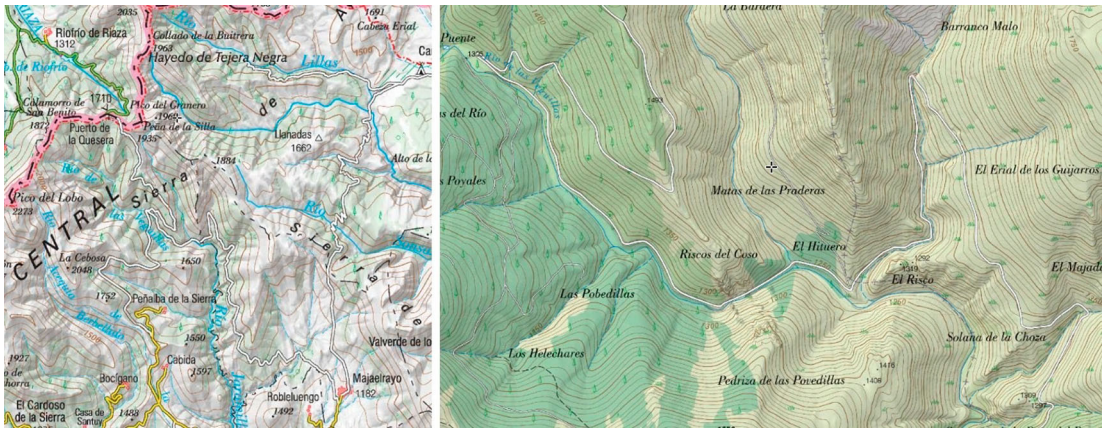

Figure 1. Road sections under study. Left: Puerto de la Quesera mountain pass road, 35 km from Riofrío de Riaza to Majaelrayo. Right: detail of the 10.6 km segment under study, on the southern side of the pass. Maps extracted and adapted from https://sigpac.mapama.gob.es/fega/visor/.

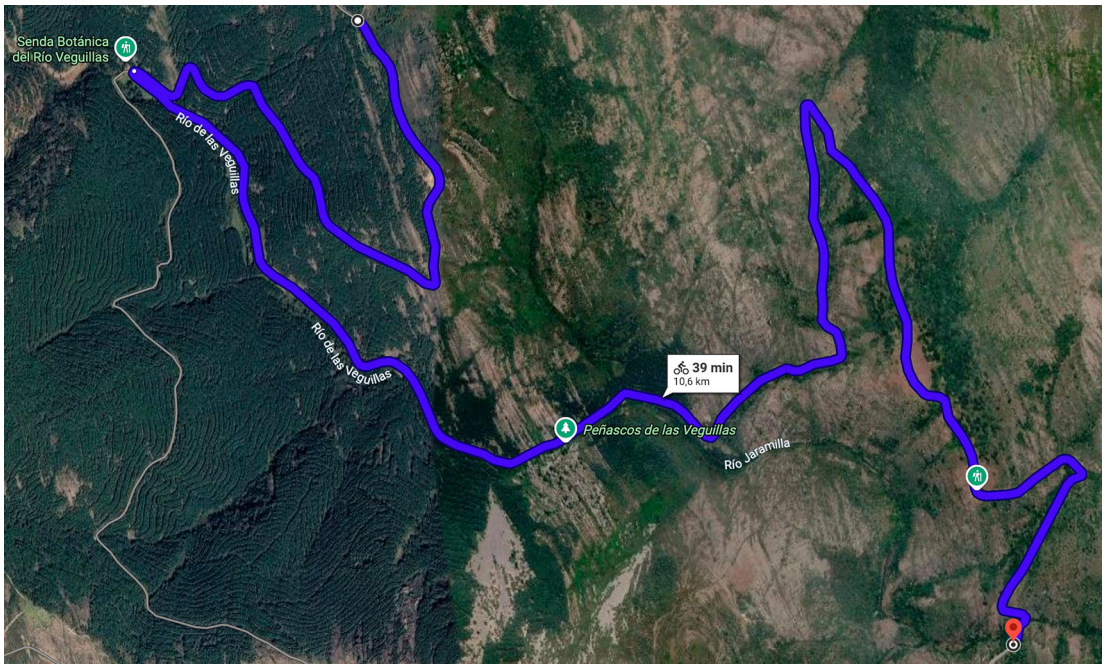

Figure 2. Satellite view of the 10.6 km fragment selected for digital-twin construction. The serpentine layout, multiple direction reversals and sustained altitude variation make this fragment representative of the most demanding sections of the Quesera pass.

2. **In-segment local intelligence.** Inside the segment, vehicles rely on on-board sensing and on opportunistic Vehicle-to-Vehicle (V2V) exchanges when in range. The architecture does not assume continuous cellular coverage inside the segment.

3. **Post-traversal evidence contribution.** When a vehicle reaches the exit beacon, it contributes traversal evidence in the form of a residual model update or compact summary statistics, rather than raw telemetry. The exit beacon aggregates these contributions, synchronises with the entry beacon through the direct or repeater-mediated link, and updates the descriptor that will be served to subsequent vehicles.

This protocol shapes the experimental evaluation in two ways. First, the unit of analysis is the *traversal* as a whole, because the operational decision to estimate range must be taken before the vehicle enters the segment. Second, the temporal logic of the beacon learning model is *prequential*: each vehicle is first evaluated with the current beacon model, and only afterwards is its evidence allowed to update the model used for subsequent vehicles. The exact implementation of this protocol in the experimental evaluation is described in Section 3.5.

3.3. Digital-twin pipeline

A digital twin of the selected fragment was constructed to enable controlled, reproducible EV traversals under heterogeneous vehicle and weather conditions. The end-to-end pipeline is summarised in Figure 3 and described in the following two subsections.

3.3.1. Geospatial preprocessing and HD Map generation

The publicly documented path to bring a real-world road segment into a high-fidelity driving simulator combines a Digital Elevation Model raster and a road centreline vector. The challenging step is to ensure that the lane geometry itself carries the real terrain elevation, rather than running on a flat plane covered by a textured mesh. A widely reported failure mode in the CARLA community is that the exported ASAM OpenDRIVE network (.xodr) ends up with $z = 0$ in all lanes, which silently invalidates any energy or perception study that depends on real road grade. To address this, a domain-specific desktop application, mapperV3, was developed in the MATLAB App Designer environment (MATLAB R2024a). The application ingests two public inputs:

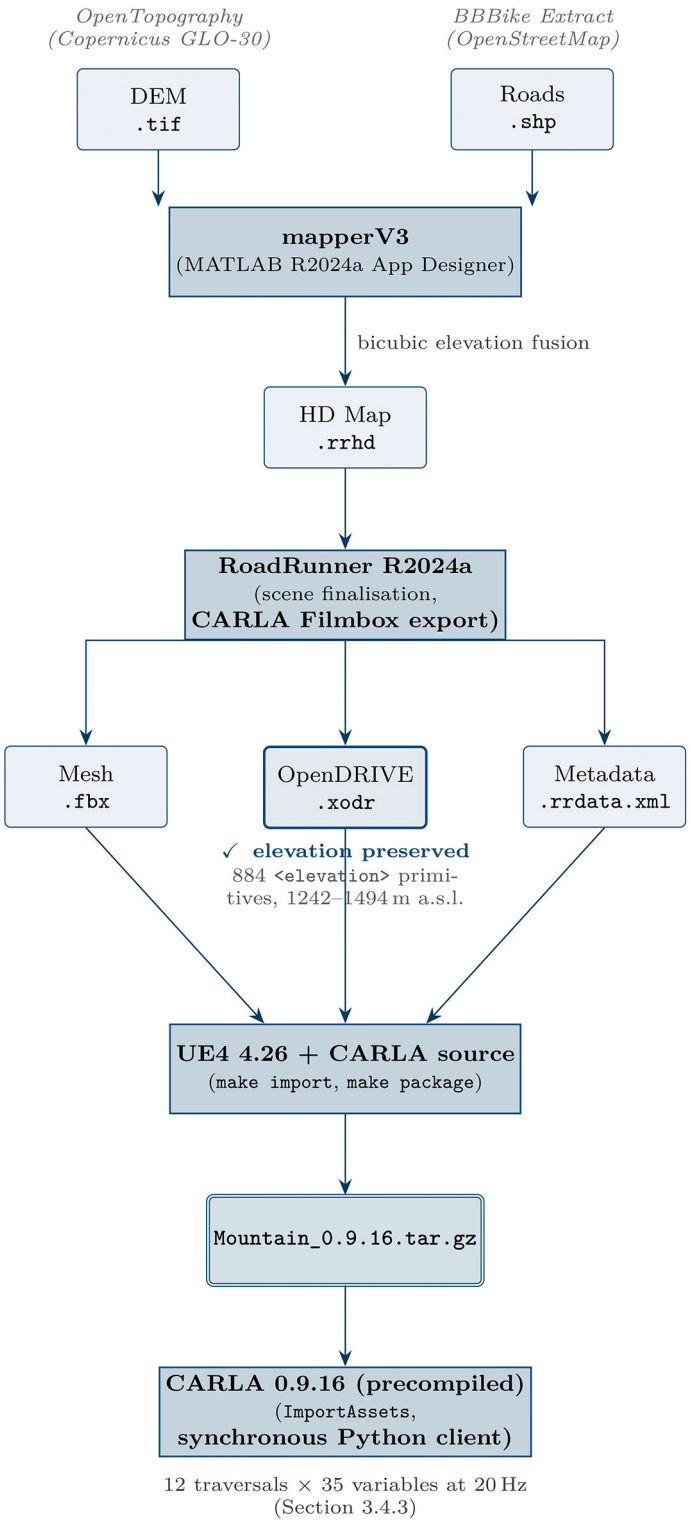

Figure 3. End-to-end digital-twin pipeline used to generate the simulation environment of Section 3.3. Public Digital Elevation Model and OpenStreetMap road data are fused by the mapperV3 application into a RoadRunner HD Map; the HD Map is finalised in RoadRunner R2024a and exported through the CARLA Filmbox path; the three resulting artefacts are imported into a source build of CARLA on Unreal Engine 4.26 and packaged into the deployable Mountain_0.9.16.tar.gz bundle ingested by the precompiled CARLA simulator at run time. The highlighted.xodr preserves 884 elevation primitives, confirming that the elevation pipeline is not affected by the widespread flat-OpenDRIVE failure mode.

- GeoTIFF digital elevation model (DEM) tile retrieved from OpenTopography (OpenTopography, n.d.), derived from the Copernicus GLO-30 dataset, with a horizontal resolution of approximately 30 m in the EPSG:4326 coordinate reference system.

- Road-network Shapefile retrieved from the BBBike extraction service from OpenStreetMap data (OpenStreetMap contributors, n.d.), covering the same bounding box as the DEM tile.

The application then executes the algorithmic pipeline detailed below.

Raster cleaning and reference frame. The GeoTIFF is read with `readgeoraster`, and no-data sentinel values (entries below −10,000) are replaced with `NaN`. The local East-North-Up (ENU) origin used by the downstream HD Map is set at the centroid of the DEM bounding box, computed as the mean of the latitude and longitude limits of the raster. This origin is stored as the `GeoReference` of the generated HD Map and ensures that the local coordinate system is consistent with the geodetic coverage of the input data.

Vector-on-raster elevation fusion. For each polyline in the road Shapefile, the algorithm extracts its (lat, lon) vertices, removes line-separating `NaN`s introduced by `shaperead`, and samples the DEM at each vertex using bicubic interpolation through `geointerp(..., 'Cubic')`. Vertices whose interpolated elevation is `NaN` (i.e., outside the DEM coverage) are discarded; polylines that retain fewer than two valid vertices are skipped. Surviving vertices are projected to local metric coordinates using a flat-earth approximation around the ENU origin, and a Gaussian smoother of window length five is applied to the (x, y) traces to suppress OpenStreetMap digitisation noise.

Lane and boundary construction. The smoothed centreline of each road is treated as the lane axis. Forward and backward driving lanes are constructed by offsetting the centreline by ±w_lane/2 along the perpendicular unit normal, with a fixed lane half-width w_lane/2 = 1.75 m. For each road, `two roadrunner.hdmap`. Lane objects are instantiated (forward and backward) together with three `roadrunner.hdmap.LaneBoundary` objects (the centre boundary shared by both lanes, plus the two outer boundaries). All geometric entities are populated with full three-dimensional geometry (x, y, z), so that elevation is preserved natively in the HD Map structure. The complete set of lanes and boundaries is assigned to a `roadrunnerHDMap` object and serialised to a `.rrhd` file.

The output of this stage is an HD Map that opens in RoadRunner with the roads already conformed to the terrain mesh, which is the prerequisite for the elevation-aware OpenDRIVE export of the next stage.

3.3.2. Scene finalisation, CARLA export, and packaging

The `.rrhd` file is opened in MathWorks RoadRunner R2024a. No additional manual editing is required beyond closing the terrain mesh outside the road envelope. The scene is then exported through the CARLA Filmbox export option (File → Export → CARLA Filmbox), producing three artefacts: a `.fbx` mesh, a `.xodr` ASAM OpenDRIVE network, and a `.rrdata.xml` metadata file.

The Filmbox export options used were: Split by Segmentation, Embed Textures, Export Only Highest LODs, and Apply Material Color To Texture. On the ASAM OpenDRIVE tab, the export was configured with OpenDRIVE 1.4, right-hand driving, Enforce connected road continuity, Export signals, Export objects, Export OpenCRG Options with Road Data Format = LRFI. This combination, identified after iterative testing, is the one that preserves the elevation profile of every road inside the `.xodr` network. Inspection of the exported `.xodr` file confirms that the elevation profile is non-trivial: the network contains 33 roads with a total of 884 elevation primitives spanning from 1242.4 m to 1493.5 m above sea level. This is the elevation band actually traversed by the road, which lies within the wider 1203–1798 m relief of the DEM tile: the DEM covers the whole surrounding terrain, whereas the `.xodr` carries only the road centreline elevations.

The three exported artefacts are then dropped into a source build of CARLA 0.9.16 and registered with the CARLA asset pipeline via `make import`. The map is packaged with `make package ARGS = "--packages = Mountain"`, producing a deployable bundle that any precompiled CARLA 0.9.16 installation can ingest through the `ImportAssets` script. The packaged map is referenced internally as `Mountain` and loaded at run time with the standard CARLA configuration utility (`config.py -m Mountain --delta-seconds 0.05`). The Unreal Engine version is 4.26.2; no manual edits were performed in the Unreal editor.

Road-grade observation. The exported HD Map preserves elevation in the visual mesh and physics colliders, in the lane geometry of the `.rrhd`, and in the elevation blocks of the `.xodr`. However, the elevation field of the navigable

waypoint network served by CARLA's Python API returned z = 0 in the present configuration. To guarantee a physically consistent road grade signal regardless of this limitation, the road grade was derived in real time from the pitch component of the vehicle transform (`vehicle.get_transform().rotation.pitch`), normalised to the physical range [−90°, +90°]. No filtering, smoothing or state estimation is applied to this signal: the value consumed by the energy model is the raw per-tick pitch at the 20 Hz acquisition rate, passed only through the aforementioned wrapping function. The pitch-based grade was validated against literature-based predictions of grade-induced energy consumption on equivalent slopes, and quantitatively against the geometry of the digital twin itself. Each traversal log records both the pitch-derived grade consumed by the energy model and the vehicle position in the world frame, whose vertical component is the elevation of the digital-twin mesh; the geometric road grade can be recovered from the trajectory and compared sample by sample with the value actually used. Compared at matched spatial resolution over the twelve traversals (119,760 one-metre samples, grades spanning −19.79° to +21.90°, 20 m secant window) the mean absolute difference is 0.039°, with a bias of +0.009°, R^2 = 0.9998, a regression slope of 1.009 and 95% limits of agreement of [−0.133°, +0.151°]. The agreement is stable with respect to the window length (0.048° at 10 m, 0.030° at 50 m) and holds across the whole grade range, with per-bin discrepancies below 0.2° in every 2° interval. Compared at one-metre resolution against the same 20 m geometric secant, the raw signal has a mean absolute difference of 0.552° with a standard deviation of 1.024° and a bias of +0.009°: the high-frequency component is substantial but essentially zero-mean, and is therefore attenuated by spatial averaging. Because vehicle pitch also responds to longitudinal acceleration, braking and suspension dynamics, the residual was decomposed against the logged longitudinal acceleration. Its sensitivity is +0.268°/(m/s^2) and orders with vehicle mass as suspension pitch should, +0.163°/(m/s^2) for the 1611 kg Tesla Model 3 against +0.451°/(m/s^2) for the 2490 kg Audi e-tron, but it accounts for only 0.45% of the residual variance in this campaign, in which the Traffic Manager holds 92.0% of samples within ±0.2 m/s^2 and applies the brake in 0.06% of them; in a campaign with aggressive braking or stop-and-go traffic this term would not be negligible. Two integral checks confirm the consequence for the energy model: the elevation gain implied by the unfiltered pitch signal departs from the mesh elevation gain by 0.69 m for the Tesla Model 3 and 2.90 m for the Audi e-tron over a net climb of approximately 152 m, and recomputing the grade-resistance energy with the geometric grade instead of the pitch-derived grade changes route-level energy by 0.003 kWh and 0.020 kWh respectively, that is 0.36% of net route energy on average and 0.74% in the worst case, which is an order of magnitude below the 0.220 kWh leave-one-route-out error of the segment-specific model. The same mass ordering appears in the matched-bandwidth error (0.023° for the Tesla against 0.054° for the Audi), consistent with static suspension deflection under load being the dominant residual source. The check establishes consistency between the pitch signal and the mesh elevation from which the scene was built; validation against real-terrain telemetry remains future work. Both analyses use only variables already recorded in the archived traversal logs.

3.4. Simulated fleet and data acquisition

The digital twin was traversed by a heterogeneous simulated fleet under a factorial scenario design. This section describes the EV library, the longitudinal energy model used to derive battery-level variables, the factorial design and the synchronous client-server acquisition.

3.4.1. Electric-vehicle library and longitudinal energy model

Two production EV models were instrumented in the simulator: a Tesla Model 3 RWD (rear-wheel drive) and an Audi e-tron 55 quattro (all-wheel drive Sport Utility Vehicle). Their main parameters, summarised in Table 2, were taken from manufacturer specifications, EPA test data, the public ev-database, and ADAC tests. Both vehicles were instantiated in CARLA through their native blueprint identifiers (`vehicle.tesla.model3` and `vehicle.audi.etron`, respectively). Both vehicles are represented by the same longitudinal energy-model structure, instantiated with vehicle-specific parameters taken from public sources; no manufacturer control strategy, thermal-management logic or proprietary energy-management behaviour is reproduced. The two configurations should therefore be read as two contrasting parameter sets (a light rear-wheel-drive saloon and a heavy all-wheel-drive SUV) rather than as digital replicas of the corresponding production vehicles, and the vehicle-specific results reported below characterise the methodology within the digital twin rather than the behaviour of those vehicles on the road.

Table 2. Parameters of the two EV models used in the simulation campaign. HVAC stands for Heating, Ventilation and Air Conditioning.

Parameter	Tesla Model 3 RWD	Audi e-tron 55 quattro
Curb mass (kg)	1611	2490
Drag coefficient Cd	0.23	0.27

Table 2. *Continued*

Parameter	Tesla Model 3 RWD	Audi e-tron 55 quattro
Frontal area A (m^2)	2.22	2.65
Nominal rolling resistance Crr	0.011	0.012
Nominal pack voltage (V)	350	396
Usable battery capacity (kWh)	72.0	86.5
Motor efficiency η_motor	0.90	0.88
Regen efficiency η_regen	0.85	0.82
Max regen power (kW)	60	220
Max discharge power (kW)	210	300
Internal pack resistance (Ω)	0.07	0.06
HVAC base power (W)	300	400

CARLA 0.9.16 does not provide a battery-level electrochemical model. A longitudinal energy model was therefore implemented in Python on top of the kinematic stream produced by the simulator. At each acquisition tick, the instantaneous traction power P_trac required to sustain the observed kinematic state is computed as the sum of four resistive terms plus an auxiliary load, as shown in Equation (1). For the computation of this equation, the four terms are defined as rolling resistance force F_rr (Equation 2), aerodynamic drag force F_aero (Equation 3), grade or slope resistance force F_grade (Equation 4), and inertial force F_inertia (Equation 5).

$$P_{\text{trac}}(t) = \left[F_{\text{rr}}(t) + F_{\text{aero}}(t) + F_{\text{grade}}(t) + F_{\text{inertia}}(t)\right] v(t) + P_{\text{aux}}(t) \tag{1}$$

$$F_{\text{rr}}(t) = C_{rr}^{\text{eff}}\, mg\cos\theta(t) \tag{2}$$

$$F_{\text{aero}}(t) = 1/2\rho_{\text{air}} C_d A v(t)^2 \tag{3}$$

$$F_{\text{grade}}(t) = mg\sin\theta(t) \tag{4}$$

$$F_{\text{inertia}}(t) = ma_{\text{long}}(t) \tag{5}$$

From Equations (1)–(5), the variables are defined as follows: v(t) is the longitudinal speed, a_long(t) is the longitudinal acceleration, and θ(t) the road grade inferred from the vehicle pitch angle as a proxy (Section 3.3.2), noting that dynamic effects such as acceleration-induced pitch and suspension dynamics may introduce estimation errors. In addition, m is the curb mass, g is the gravitational acceleration, ρ_air = 1.225 kg/m^3 is the air density, and Crr_eff is the effective rolling resistance coefficient after the climatic correction described below (see Section 3.4.2). In addition, Cd is the aerodynamic drag coefficient, A is the vehicle frontal area, and P_aux(t) represents the auxiliary power demand associated with onboard systems such as HVAC, electronics, and other non-propulsion loads.

The instantaneous battery power P_batt(t) (Equation 6) accounts for motor and regenerative braking efficiencies, as well as regenerative and discharge power limits specific to each vehicle, where η_motor is the drivetrain efficiency in traction mode, and η_regen is the efficiency in regenerative braking mode. It is clipped to the interval [−P_regen_max, P_disch_max], where P_regen_max and P_disch_max represent the maximum regenerative and discharge power limits of the battery system, respectively, imposed by the battery and power electronics constraints.

$$P_{\text{batt}}(t) = \begin{cases} \dfrac{P_{\text{trac}}(t)}{\eta_{\text{motor}}} & \text{if } P_{\text{trac}}(t) \geq 0 \\ P_{\text{trac}}(t)\,\eta_{\text{regen}} & \text{if } P_{\text{trac}}(t) < 0 \end{cases} \tag{6}$$

The state of charge is updated by numerically integrating P_batt(t) over time and normalising by the usable battery capacity. This is an energy-based (watt-hour) integration rather than Coulomb counting in the strict sense of integrating battery current; the quantity tracked is therefore a state of energy, and it is reported as SOC only because that is the conventional term in range-estimation practice. Pack voltage and current are recorded as derived diagnostic variables: terminal voltage is modelled as an affine function of SOC, while current is calculated from battery power. Both variables are used exclusively by the thermal sub-model and do not affect the state update. Replacing energy-based integration with

current integration, while retaining the same pack model, would result in a maximum deviation of 0.58 percentage points in the end-of-traversal state across the twelve traversals, with a mean deviation of 0.27 percentage points. These differences are small relative to the average SOC decrease of 4.06 percentage points per traversal. In addition, cumulative consumed and regenerated energies are tracked separately, ensuring that the route-level net energy demand defined in Equation 7 is well defined and used in Section 4.

$$E_{\text{net}} = E_{\text{used}} - E_{\text{regen}} \tag{7}$$

3.4.2. Climate parameterisation and factorial design

Three road-weather conditions were considered, denoted `dry24`, `wet10` and `snow_neg5`. Each condition is defined consistently across the simulator and the Python energy model.

In CARLA, the conditions are realised by configuring `WeatherParameters` with the precipitation, surface wetness, wind intensity and sun-altitude values listed in Table 3. CARLA 0.9.16 does not natively render snow particles; the `snow_neg5` condition is therefore not visually distinguishable from a cold overcast scene, and is instead characterised through its physical effects on the EV model, as described next.

On the model side, three climate-dependent multipliers are applied:

1. the auxiliary cabin power P_aux(T_amb) follows a piecewise-linear profile that grows with the deviation from the thermal comfort band (500 W at 24 °C, 1500 W at 10 °C, 3750 W at −5 °C);

2. the usable battery capacity is scaled by a temperature-dependent factor η_batt(T_amb) for Li-ion NMC chemistry: 1.00 at 24 °C, 0.95 at 10 °C, 0.86 at –5 ºC (K. Liu et al., 2018);

3. the effective rolling resistance Crr_eff is scaled by a road-surface multiplier: 1.0 for dry, 1.2 for wet, 2.0 for snow (Sandberg et al., 2011).

The full experimental design is a 2 × 2 × 3 factorial covering two vehicle types, two initial states of charge (SOC) (90% and 40%) and the three climate conditions, for a total of twelve scenarios labelled S01 through S12. The initial SOC factor was included to expose any saturation effect of the regenerative-braking power limit in the prolonged descents of the segment; its effect on the per-traversal net energy is reported in Section 4 and is found to be small in the current simulator configuration.

3.4.3. Synchronous client-server acquisition

Each traversal was executed in CARLA's synchronous mode with a fixed simulation timestep Δt = 50 ms, corresponding to an acquisition rate of 20 Hz. The simulator acted as a server (Unreal Engine 4.26.2, CARLA 0.9.16) and a Python client coordinated the run: it configured the map, applied the weather parameters of the active scenario, spawned the configured EV at a fixed entry waypoint of the segment, enabled the CARLA Traffic Manager as the autonomous driver, and advanced the simulation one tick at a time. After every tick, the client queried the simulator for the vehicle transform, velocity, acceleration, control inputs and weather state; it then evaluated the longitudinal energy model of Section 3.4.1 and the road-grade observation of Section 3.3.2, and appended a row to the per-traversal Comma-Separated Values (CSV) dataset.

Table 3. Climate conditions used in the factorial scenario design. CARLA-side parameters configure `WeatherParameters`; model-side multipliers act on the Python energy model.

Tag	Precip. (%)	Wetness (%)	Wind (%)	Sun alt. (°)	T_amb (°C)	P_aux (W)	Crr_eff/ Crr	η_batt
dry24	0	0	0	30	+24	500	1.0	1.00
wet10	60	70	30	10	+10	1500	1.2	0.95
snow_neg5	80	40	50	5	−5	3750	2.0	0.86

Each row of the resulting dataset contains 35 variables organised in five functional groups: identification and temporality (run identifier, vehicle identifier, simulation timestamp, relative time, time step, OpenDRIVE road and lane identifiers); positioning and topology (`pos_x`, `pos_y`, `pos_z`, road grade in degrees, road curvature, road friction); vehicle dynamics (scalar speed, longitudinal acceleration, lateral acceleration, throttle, brake, steering); battery-level energy variables derived from the longitudinal model (instantaneous battery power, state of charge, pack voltage and current, cumulative regenerated energy, battery temperature); and meteorological state (precipitation intensity, wind speed, road condition category). Each complete traversal of the 10.6 km segment took approximately seven minutes of wall-clock time at a sustained simulation-to-real ratio of approximately 3.2 $\times$. The segment of interest is 10.6 km long; the spawn and end waypoints of the simulated traversals lie slightly inside its endpoints, so each recorded log spans approximately 10.0 km of it. All reported energies refer to the recorded traversal.

A representative scenario (S11, Audi e-tron at 40% SOC under wet10) was executed twice under identical configuration to assess the residual stochasticity of the CARLA Traffic Manager. The coefficient of variation of the per-traversal net energy across the two repetitions was below 0.8%, indicating that variability between scenarios reported in Section 4 is dominated by the controlled experimental factors rather than by autopilot noise. The complete dataset, the EV model code and the acquisition script are made available as described in Section 6.

3.5. Energy-estimation methods compared

The evaluation in Section 4 compares generic estimators, a conventional non-local physics estimator, and segment-specific learning methods. The purpose is not to isolate the marginal contribution of a single input variable, but to evaluate whether a model associated with a difficult road segment can reduce route-level autonomy-estimation errors and unsafe range-risk decisions relative to estimators that are not adapted to that segment. The main comparison is therefore between (i) estimators that a vehicle could use without local traversal evidence and (ii) residual models maintained by the beacon from previous traversals.

3.5.1. Generic and non-local baselines

All baselines predict the route-level energy demand Ê_route that a vehicle of a given class will require to complete the 10.6 km segment before entering it. The following baseline methods are taken into account:

- **B0 — Generic nominal estimator.** The route energy is estimated as travelled distance multiplied by a vehicle-specific nominal consumption coefficient (Wh/km). This estimator has no access to local traversal evidence and represents a simple range extrapolation.

- **B1 — Conventional non-local physics estimator.** The road-load equation (Equation 1) is integrated along the segment under a generic entry assumption: the vehicle knows its class, nominal parameters and an expected speed profile, but the estimate is not corrected using previous traversals of the target segment. In the implementation used for the reported experiments, this baseline corresponds to the `baseline_entry_flat_wh` column and is the baseline used to initialise the residual model.

Additional physics-only ablations were computed during development to verify the consistency of the simulation pipeline, but they are not reported as main experimental results because the objective of the paper is to evaluate the operational value of segment-specific learning and alert generation.

3.5.2. Centralised segment-specific residual neural model

A residual neural model **M1** was trained to predict the residual r(x) between the observed route-level net energy and the conventional non-local baseline estimate, Equation 8; where E_route is the observed (ground-truth) route-level energy and Ê_base is the baseline route-energy estimate available before local correction (the `baseline_entry_flat_wh` output of B1). The network f_θ(x) approximates this residual, with x containing vehicle descriptors, segment descriptors, environmental descriptors and baseline-estimator outputs.

$$r(\boldsymbol{x}) = E_{\text{route}} - \widehat{E}_{\text{base}}, \quad f_\theta(\boldsymbol{x}) \approx r(\boldsymbol{x}) \tag{8}$$

The corrected route-energy prediction is obtained by adding the learned residual back to the baseline estimate, Equation 9, so that Ê_route ≈ E_route when the residual is predicted accurately.

$$\widehat{E}_{\text{route}} = \widehat{E}_{\text{base}} + f_\theta(\boldsymbol{x}) \tag{9}$$

The model is trained centrally on pooled traversal data and serves as a reference for the federated and prequential variants. Numeric features are standardised using statistics fitted on the training clients only, categorical variables (vehicle type and road condition) are one-hot encoded, and the target is the segment-level residual with respect to the baseline. The final route prediction is obtained by summing segment-level predictions along the full route.

The residual model is a lightweight Multi-Layer Perceptron (MLP) with three hidden layers of 64, 64 and 32 units, respectively, ReLU activations, no dropout, and a linear output. The final layer is initialised to zero so that an untrained model predicts a zero residual and therefore behaves exactly as the baseline. This avoids leaking the average residual of future vehicles into the cold-start beacon evaluation. Training uses mean-squared error loss and the Adam optimiser.

3.5.3. Federated residual neural model

The federated variant **M2** preserves the residual architecture of M1 but trains it through federated aggregation across the twelve traversals, treating each traversal as one client. Two aggregation rules are evaluated, FedAvg (McMahan et al., 2017) and FedProx (Li et al., 2020), with the latter expected to be more stable under non-Independent and Identically Distributed (non-IID) client data. This non-IID setting is natural here because traversals differ in vehicle class, climate condition and initial SOC. The federated aggregation underlying the reported prequential results directly implements FedAvg and FedProx: client parameters are combined through a weighted average, while FedProx additionally incorporates a proximal term into the local optimisation step. The process is simulated sequentially to remain consistent with the beacon protocol. The Flower framework (Beutel et al., 2020), used through the `flwr [simulation]` package in the 1.x API range ($\geq$1.8,<2.0), provides the explicit federated runtime experiments distributed with the code.

3.5.4. Prequential beacon learning

The prequential variant **M3** implements the operational protocol of Section 3.2. Traversals are presented to the beacon model in a random arrival order. For each arriving traversal, the current beacon model is first used to estimate $\hat{E}$_route before the vehicle enters the segment; the prediction error is then measured against the observed E_net; and only afterwards is the traversal allowed to update the beacon model. The update is applied either immediately (aggregation batch size of one) or after a fixed number of arrivals have accumulated (aggregation batch sizes of two and four were also evaluated). The protocol therefore enforces the same temporal causality that would apply in the field: a vehicle benefits only from the experience of previously crossed vehicles. Fifty independent random arrival permutations were generated, and the per-step error was averaged across permutations. The resulting learning curve is the central outcome of the beacon evaluation.

Two aggregation rules are evaluated within M3, mirroring M2: FedProx and FedAvg. The remaining neural and optimisation hyperparameters match those reported in Table 4. Flower-based replays were used to verify that the vehicle-as-client and beacon-as-aggregator workflow can be implemented with standard FL tooling; the primary statistical evidence reported in Section 4 is provided by the route-level cross-validation results and by the prequential beacon learning curve. The model broadcast by the beacon consists of two components with distinct lifecycles. The normalisation package, comprising the input-feature statistics and target scaling parameters, is configured during beacon commissioning and transmitted unchanged. By contrast, the residual network weights are updated using the traversal evidence contributed by each vehicle after it exits the segment.

Table 4. Main neural and federated-learning hyperparameters used in the reported experiments.

Component	Value
Residual model	MLP with hidden layers [64, 64, 32], ReLU activations, linear output, dropout 0.0, zero-initialised final layer
Feature set	Segment-entry and local descriptors: distance/position summaries, elevation and grade summaries, curvature, ambient temperature, precipitation, wind, SOC, usable capacity, battery temperature, expected speed, baseline estimates, vehicle type and road condition
Target	Segment-level residual with respect to `baseline_entry_flat_wh`; route-level energy obtained by summing segment predictions
Optimiser and loss	Adam, mean-squared error loss
Batch size	32

Table 4. *Continued*

Component	Value
Learning rate and weight decay	10^{-3} and 10^{-5}
Centralised training	200 epochs for the default split; 160 epochs in cross-validation; early stopping patience 30 and 25, respectively
Local FL training	5 local epochs in the main federated/prequential runs; 3 local epochs in cross-validation/checkpoint runs
Federated aggregation	FedAvg and FedProx; all available clients participate in each simulated round; FedProx proximal coefficient μ = 0.01
Flower configuration	12 rounds in the default run; 8 rounds in Flower checkpoint experiments; fraction_fit = 1.0, fraction_evaluate = 1.0, minimum fit clients 2
Prequential evaluation	50 random arrival permutations; aggregation batch sizes 1, 2 and 4; main tables report immediate updating after each vehicle
Random seed	42

3.6. Evaluation protocol and Key Performance Indicators

This subsection defines how the simulation output is turned into quantitative evidence: the unit of evaluation and data splits, the regression and risk-oriented Key Performance Indicators (KPIs) actually reported in this work, and the broader architectural KPIs deferred to future evaluation.

3.6.1. Unit of evaluation and data splits

The unit of evaluation is the *complete traversal* of the segment, not an individual simulator tick or sub-segment. This choice matches the operational use case described in Section 3.2: a vehicle approaching the entry beacon must estimate, before entering the segment, how much energy will be required to complete the section. Although each traversal also produces sub-segment information (the twelve traversals decompose into 1,212 road sub-segments covering 119.94 km of cumulative simulated driving), no claim is made about per-sub-segment prediction accuracy.

Five cross-validation protocols are used to characterise the residual model under different distribution-shift assumptions:

- **Leave-one-route-out (LORO):** each traversal is held out in turn while the remaining eleven are used for training. This is the standard within-distribution evaluation.
- **Leave-one-scenario-group-out:** traversals sharing the same combination of vehicle and climate are held out together. This evaluates generalisation to a previously unseen *operating regime* of the same fleet.
- **Repeated random client split:** repeated random partitions of the twelve traversals into train and test subsets, averaged over multiple seeds.
- **Leave-one-weather-out:** all traversals under a given climate condition are held out simultaneously. This is a deliberate out-of-distribution stress test.
- **Leave-one-vehicle-out:** all traversals of one of the two vehicle classes are held out simultaneously. This is a second deliberate out-of-distribution stress test.

The prequential evaluation of M3 uses its own temporally-ordered protocol, as described in Section 3.5.4, and is therefore not part of the cross-validation table.

3.6.2. Regression KPIs

For every method and every traversal, four regression KPIs are computed and aggregated across traversals.

- **MAE (kWh).** Mean absolute error between predicted and observed route-level net energy. Reported as the mean of per-traversal absolute errors, with the standard deviation across traversals as the dispersion measure.

- **Bias (kWh).** Signed mean error. A negative bias indicates systematic underestimation of the required energy, which is the safety-relevant failure mode in the present setting.

- **MAPE (%).** Mean absolute percentage error. Useful to compare methods on a relative scale across vehicle classes that differ in absolute energy demand.

- **Final-SOC MAE (% SOC).** Mean absolute error of the predicted final state of charge against the observed final state of charge. This KPI translates the energy error into the SOC margin that is actually displayed to drivers and used by range-anxiety mitigation logic.

3.6.3. Risk-oriented KPI and alert-generation logic

Average regression error is not sufficient to evaluate the proposed system, because the operational cost of an error is asymmetric. In a low-connectivity mountain segment, an over-optimistic prediction is more critical than a conservative one: a vehicle that is wrongly told that it can safely cross the section may reach the exit with insufficient reserve, whereas a conservative warning mainly produces inconvenience or a detour recommendation. For this reason, the evaluation includes an alert-oriented KPI set in addition to MAE, MAPE and bias.

For each route-level prediction, the estimator outputs the expected energy required to complete the segment, $\hat{E}_route$. Given a simulated entry state of charge SOC_in and the usable battery capacity C_use, the predicted final state of charge is computed following Equation 10. The observed final state of charge is computed analogously from the measured route-level net energy E_net as stated by Equation 11. Both quantities are energy balances against the same usable capacity: the energy available at entry less the net route energy demand, compared with the reserve. The risk label is therefore defined on the same energy basis as the simulator state of Section 3.4.1, and is not derived from the logged state trajectory.

$$\widehat{SOC}_{\text{out}} = SOC_{\text{in}} - 100 \cdot \frac{\hat{E}_{\text{route}}}{C_{\text{use}}} \tag{10}$$

$$SOC_{\text{out}} = SOC_{\text{in}} - 100 \cdot \frac{E_{\text{net}}}{C_{\text{use}}} \tag{11}$$

A minimum reserve threshold $\sigma_rsv = 5\%$ SOC is fixed. The predicted alert decision is then defined by Equation 12, where $\hat{A} = 1$ means that the system issues a low-reserve alert before the vehicle enters the segment. The true risk state could be computed by Equation 13.

$$\hat{A} = \begin{cases} 1 & \widehat{SOC}_{\text{out}} < \sigma_{\text{rsv}} \\ 0 & \widehat{SOC}_{\text{out}} \geq \sigma_{\text{rsv}} \end{cases} \tag{12}$$

$$A = \begin{cases} 1 & SOC_{\text{out}} < \sigma_{\text{rsv}} \\ 0 & SOC_{\text{out}} \geq \sigma_{\text{rsv}} \end{cases} \tag{13}$$

The alert evaluation is therefore a binary decision problem derived from the energy-estimation task. A false-safe decision occurs when $\hat{A} = 0$ and $A = 1$: the system does not issue an alert, but the vehicle actually finishes the segment below the reserve threshold. This is the primary safety-relevant failure mode. A false-warning decision occurs when $\hat{A} = 1$ and $A = 0$: the system issues an alert, but the vehicle would have completed the segment above the reserve threshold. False warnings are undesirable because they may reduce user trust, but they are less critical than false-safe decisions.

The following risk-oriented metrics are reported:

- **False-safe rate:** fraction of route-SOC decisions in which the estimator fails to warn despite the actual final SOC being below the reserve threshold.

- **False-warning rate:** fraction of route-SOC decisions in which the estimator warns despite the actual final SOC being above the reserve threshold.

- **Risk recall:** fraction of truly risky cases correctly detected by the estimator.
- **Alert precision:** fraction of emitted alerts that correspond to truly risky cases.

The false-safe rate is treated as the primary risk KPI. A model is considered operationally better when it reduces false-safe decisions without producing an excessive number of false warnings. This criterion reflects the intended role of the beacon: not to provide an exact battery forecast in all conditions, but to prevent unsafe over-confidence before entering an isolated and energetically demanding segment.

3.6.4. Architectural KPIs as future evaluation targets

The KPI set above is restricted to energy estimation and range-risk decisions, which are the outcomes that the present experimental campaign actually produces. Architectural KPIs that would characterise the deployed beacon-V2X-FL system as a whole, including event-to-adaptation latency, on-vehicle adaptation time, risk-detection F1-score, risk-mitigation rate, no-update gap, stale-context fraction, and power-consumption overhead, require dedicated perception data, an instrumented V2X communication stack, and measurements on representative embedded hardware. They are therefore not reported in Section 4 and are explicitly stated as future evaluation targets. Their definition and the corresponding instrumentation are sketched in the discussion (Section 5).

4. Results and Analysis

This section evaluates the proposed beacon-centric V2X/FL architecture as a local range-risk estimation mechanism for a low-connectivity mountain-road segment. The purpose of the evaluation is not to isolate the contribution of any single physical input, but to determine whether a segment-specific model, learned from previous traversals and distributed through roadside beacons, can reduce over-optimistic range decisions before an EV enters a demanding section.

The evaluation is organised around four research questions:

- **RQ1:** How variable is the energy demand of the selected segment across vehicle and road-weather conditions?
- **RQ2:** Does a segment-specific model reduce route-level energy-estimation error and unsafe over-optimistic (false-safe) decisions relative to generic or non-local estimators?
- **RQ3:** How robust is the residual model under route-level cross-validation and distribution-shift protocols?
- **RQ4:** Does the prequential beacon protocol reduce unsafe over-optimistic decisions as more vehicles traverse the segment?

The results reported here are restricted to energy estimation and range-risk assessment. Communication-layer and perception-layer KPIs, such as event-to-adaptation latency, risk detection F1-score and risk mitigation rate, require dedicated perception and V2X instrumentation and are therefore left for future work.

4.1. Route-level energy variability

As it was previously stated, the evaluation was conducted on 12 simulated traversals of the selected 10.6 km mountain segment. Each traversal is treated as one vehicle-client in the federated-learning setting. The simulated fleet comprises two EV models, a Tesla Model 3 and an Audi e-tron, under three road-weather conditions: dry pavement at 24 °C, wet pavement at 10 °C, and snow/low-temperature conditions at −5 °C. Two initial SOC levels were considered, 90% and 40%, although the current simulator configuration gives limited influence to initial SOC in the final net energy.

The 12 traversals were decomposed into 1,212 road sub-segments, covering 119.94 km of cumulative simulated driving. The evaluation metric, however, is computed at traversal level. This matches the operational use case: a vehicle approaching the entry beacon must decide, before entering the segment, whether the expected energy requirement leaves enough reserve at the exit.

Table 5 summarises the observed net energy demand. The same road segment requires substantially different energy depending on vehicle class and road-weather condition. For the Tesla Model 3, the route energy increases from approximately 1.68 kWh in dry conditions to approximately 3.35 kWh in snow. For the Audi e-tron, the corresponding range is approximately 2.82 kWh to 4.63–4.67 kWh. This variability supports the need for a local segment-specific

Table 5. Observed net energy demand for the simulated mountain-segment traversals. Values are grouped by vehicle and road-weather condition.

Vehicle	Condition	Observed net energy demand
Tesla Model 3	Dry, 24 °C	≈ 1.68 kWh
Tesla Model 3	Wet, 10 °C	≈ 2.09–2.21 kWh
Tesla Model 3	Snow, −5 °C	≈ 3.35 kWh
Audi e-tron	Dry, 24 °C	≈ 2.82 kWh
Audi e-tron	Wet, 10 °C	≈ 3.14–3.27 kWh
Audi e-tron	Snow, −5 °C	≈ 4.63–4.67 kWh

model: the relevant quantity is not merely the distance to be travelled, but the expected energy demand of this particular section under its current operating conditions. These vehicle-specific figures characterise the two parameter sets within the digital twin; the energy-management behaviour of the corresponding production vehicles is not modelled here and is not the object of the comparison.

4.2. Generic versus segment-specific estimation

The first comparison evaluates whether generic estimators provide sufficient reliability for the selected segment. Table 6 reports the main route-level results. The nominal-consumption baseline obtains an MAE of 0.777 kWh and a negative bias of −0.777 kWh. The conventional non-local physics baseline obtains a larger MAE of 1.975 kWh and a similarly negative bias. In both cases, the relevant observation is not only the magnitude of the error but its sign: the generic estimators are systematically optimistic and tend to underestimate the energy required to complete the segment.

By contrast, the centralised segment-specific residual model obtains an MAE of 0.220 kWh in leave-one-route-out evaluation. This corresponds to an approximate 71.7% reduction relative to the nominal-consumption baseline and an approximate 88.9% reduction relative to the conventional non-local physics baseline. The result indicates that traversal evidence from the same segment contains useful information for correcting route-level energy estimates.

The largest errors of the generic estimators occur in adverse conditions. Under snow and low temperature, the nominal estimator underestimates the energy required by approximately 1.8 kWh on average. Such a deviation is operationally relevant because, for a vehicle entering the segment with low SOC, an error of this magnitude can determine whether the predicted final SOC remains above or below the reserve threshold. The safety-relevant consequence of this systematic underestimation, i.e. its effect on false-safe range decisions, is quantified in the alert-oriented evaluation of Section 4.5.

4.3. Cross-validation of the segment-specific residual model

The residual neural model was evaluated with route-level cross-validation. Splits were always performed by complete traversal, never by simulation tick or road sub-segment, to avoid leakage between consecutive samples from the same route. Table 7 summarises the results.

The model performs well when the test traversal belongs to a regime represented in the training set, as shown by the leave-one-route-out, leave-one-scenario-group-out and repeated random split protocols. However, performance deteriorates when an entire weather condition or an entire vehicle type is withheld from training. This behaviour is expected given the small dataset size: 12 traversals, two vehicle models and three weather regimes are sufficient for a proof of concept, but not for training a universal estimator for unseen domains.

Table 6. Generic versus segment-specific route-level energy estimation. Negative bias indicates systematic underestimation of the required energy.

Method	MAE (kWh)	Bias (kWh)	MAPE (%)	Final-SOC MAE (% SOC)
Generic nominal estimator	0.777	−0.777	22.67	1.12
Conventional non-local physics estimator	1.975	−1.975	64.48	2.71
Segment-specific residual model, LORO	0.220	+0.078	7.75	0.30

Table 7. Cross-validation results of the segment-specific residual model. Values are MAE mean ± standard deviation in kWh.

Protocol	Residual model MAE (kWh)
Leave-one-route-out	0.220 ± 0.164
Leave-one-scenario-group-out	0.167 ± 0.145
Repeated random client split	0.231 ± 0.141
Leave-one-weather-out	1.228 ± 1.583
Leave-one-vehicle-out	1.568 ± 1.150

This result supports the intended deployment logic. The proposed system should not be interpreted as a static global model trained once and applied everywhere. Its value lies in maintaining a local model for a specific segment and updating it as vehicles traverse that segment under the conditions actually encountered there.

4.4. Prequential beacon learning

The beacon-based setting was evaluated with a prequential protocol. Vehicles arrive sequentially. For each vehicle, the current beacon model is first used to estimate the route energy before entering the segment. The vehicle then completes the traversal, the prediction error is measured, and only afterwards is the traversal allowed to update the beacon model. This protocol enforces the same temporal causality as the field deployment: a vehicle can benefit only from the evidence contributed by previous vehicles.

The learning curve was computed over 50 random vehicle-arrival permutations and different aggregation batch sizes. Table 8 reports the results for immediate updating after each traversal. With no prior vehicles, the beacon starts from the conventional non-local baseline. As evidence accumulates, FedProx progressively reduces its route-energy error, from 1.975 kWh with no prior vehicles to 0.617 kWh after eight. The trajectory is not strictly monotonic, a transient rise appears at four vehicles and, in average MAE terms, FedProx only overtakes the simple nominal estimator after about eight traversals; with fewer vehicles the nominal estimator remains competitive. The operationally decisive advantage of the beacon appears earlier in the alert-oriented evaluation (Section 4.5), where the false-safe rate is reduced well before the MAE crossover.

The learning curve indicates that the beacon model is most useful after it has accumulated evidence from several previous traversals. This is consistent with the proposed architecture: the first vehicles provide the initial local evidence, and subsequent vehicles receive progressively better segment-specific estimates. The choice of aggregation batch size affects this process. Updating after every vehicle provides the fastest adaptation, which is attractive for low-traffic roads. Aggregating every two vehicles offers a compromise between responsiveness and client aggregation, while larger batches delay learning and may be less suitable for sparse mountain passes.

Table 9 compares the central segment model and the prequential beacon on common metrics. Against the estimators available without local learning, the beacon lowers the false-safe rate from 14.58% and 8.33% to 3.75% (FedAvg) and 3.38% (FedProx) after eight vehicles, and the route-level error from 1.975 and 0.777 kWh to 0.682 and 0.617 kWh, without raw telemetry leaving the vehicles. This is the contribution of federated learning in the proposed architecture. M1 is an upper bound rather than a competitor: it is trained on every other traversal, including the operating regime of the test route, whereas the beacon has only seen the vehicles that have already crossed and never centralises raw telemetry. With

Table 8. Prequential beacon learning curve with immediate updating after each traversal. Values report MAE mean ± standard deviation in kWh over vehicles not yet seen by the beacon.

Vehicles previously seen	FedProx	FedAvg	Nominal estimator
0	1.975 ± 0.000	1.975 ± 0.000	0.777 ± 0.000
1	1.602 ± 0.166	1.611 ± 0.185	0.778 ± 0.067
2	0.860 ± 0.230	0.931 ± 0.269	0.785 ± 0.103
4	0.951 ± 0.391	0.918 ± 0.338	0.764 ± 0.132
8	0.617 ± 0.339	0.682 ± 0.305	0.768 ± 0.315

Table 9. Central segment model and prequential beacon on common metrics. Route-level MAE and false-safe rate; the evaluation set of each row is given in the second column.

Model	Evaluation	Route-level MAE (kWh)	False-safe rate
Conventional non-local estimator (B1)	12 routes	1.975	14.58%
Generic nominal estimator (B0)	12 routes	0.777	8.33%
M1 — central residual model	leave-one-route-out, 12 routes	0.220	1.04%
M3 — prequential beacon, FedAvg, 8 vehicles seen	remaining unseen, 50 permutations	0.682 ± 0.305	3.75%
M3 — prequential beacon, FedProx, 8 vehicles seen	remaining unseen, 50 permutations	0.617 ± 0.339	3.38%

twelve traversals the difference between the two aggregation rules is within the variability across arrival orders and is not interpreted here as evidence that one rule outperforms the other. The reduction in false-safe decisions is also not free: the beacon's false-warning rate is non-zero (2.13%, 3.66% and 1.56% after two, four and eight vehicles for FedProx; 2.18%, 3.34% and 1.75% for FedAvg), where that of the central model is 0.00%.

4.5. Alert-oriented range-risk evaluation

The operational purpose of the beacon is not only to reduce average error but to prevent unsafe over-confidence. For this reason, the energy prediction was converted into an alert decision using a 5% SOC reserve threshold. For each traversal, eight entry-SOC values were simulated, producing 96 route-SOC decisions. These 96 decisions derive from twelve independent traversals: the number of independent observations is twelve, and the 1.04% false-safe rate of the segment-specific model corresponds to a single case out of 96. The percentages should be read accordingly. A false-safe case occurs when the estimator predicts that the vehicle will finish above the reserve threshold, but the observed final SOC falls below it.

Table 10 reports the main alert-oriented results. The conventional non-local physics estimator produces 14 false-safe cases, corresponding to a false-safe rate of 14.58%. The nominal estimator reduces this to 8 cases, but still fails to warn in several low-SOC scenarios. The segment-specific residual model reduces the false-safe rate to 1.04% in leave-one-route-out evaluation. The risk recall of 0.00 for the conventional non-local estimator follows from its systematic underestimation of the required energy: every risky case is predicted to finish above the reserve threshold, so none is detected, which is also why its false-warning rate is 0.00%. The same 0.00% has the opposite meaning for the segment-specific model, which detects 93% of the risky cases without issuing any false warning. The null recall is specific to the 5% reserve: at 7%, the non-local estimator detects 12 of the 26 risky cases. This is the failure mode that motivates using false-safe cost rather than mean error as the primary KPI.

The prequential beacon model also improves the alert behaviour as more vehicles contribute evidence. Table 11 shows the evolution for FedProx with immediate updating. With no previous vehicles, the beacon inherits the 14.58% false-safe rate of the initial non-local model. After two previous vehicles, the false-safe rate falls to 5.43%. After eight previous vehicles, it reaches 3.38%. This represents an approximate 60% reduction relative to the nominal estimator and an approximate 76% reduction relative to the conventional non-local estimator.

Table 10. Alert-oriented evaluation using a 5% SOC reserve threshold. Values are computed over 96 route-SOC decisions.

Method	False-safe rate	False-safe cases	False-warning rate	Risk recall
Conventional non-local physics estimator	14.58%	14	0.00%	0.00
Generic nominal estimator	8.33%	8	0.00%	0.43
Segment-specific residual model, LORO	1.04%	1	0.00%	0.93

Table 11. Evolution of false-safe rate as the beacon observes more vehicles. FedProx and FedAvg with immediate updating are shown.

Vehicles previously seen	FedProx false-safe	FedAvg false-safe	Nominal false-safe	Conventional non-local false-safe
0	14.58%	14.58%	8.33%	14.58%
2	5.43%	6.08%	8.35%	14.70%
4	3.63%	4.22%	8.28%	14.50%
8	3.38%	3.75%	8.44%	14.12%

Table 12. Sensitivity of the alert results to the reserve threshold: false-safe cases and, in parentheses, the fraction of genuinely risky cases left undetected. Values are computed over 96 route-SOC decisions.

Reserve threshold	Truly risky cases (of 96)	Conventional non-local	Generic nominal	Segment-specific (LORO)
3%	4	4 (100%)	4 (100%)	0 (0%)
5%	14	14 (100%)	8 (57%)	1 (7%)
7%	26	14 (54%)	8 (31%)	1 (4%)
10%	40	16 (40%)	4 (10%)	0 (0%)

This alert-oriented result is the strongest operational evidence in the experiment. Even when the beacon model is not always the best method in average regression error, it reduces the most safety-critical failure mode: telling a driver that the segment can be completed safely when the final SOC actually falls below the reserve threshold.

Table 12 reports the alert results recomputed from the existing route-level predictions at reserve thresholds of 3%, 5%, 7% and 10% SOC. Raising the reserve threshold increases the number of situations that are genuinely risky, from four to forty of the ninety-six route-SOC decisions, so the absolute number of false-safe cases is not a monotone function of the threshold: a case is false-safe only while the threshold lies between its true and its predicted final state of charge. Expressed as the fraction of genuinely risky cases that go undetected, the behaviour is unambiguous. The conventional non-local estimator misses every risky case up to a 5% reserve and still misses 40% of them at 10%; the generic nominal estimator improves from 100% to 10%; and the segment-specific residual model misses at most one case out of ninety-six at any threshold, and none at 3% or 10%. The ordering of the three estimators is preserved throughout, so the conclusion of this section does not depend on the particular choice of a 5% reserve.

4.6. Summary of results

The results support three main observations. First, the selected mountain segment produces highly variable energy demand depending on vehicle type and road-weather condition, so generic range extrapolation is insufficient. Second, a segment-specific residual model substantially reduces route-level energy-estimation error when evaluated within the observed distribution. Third, the prequential beacon protocol progressively reduces false-safe decisions as more vehicles traverse the segment, which directly supports the intended use case of range-risk alerting in low-connectivity mountain roads.

5. Discussion and Conclusions

This work proposed and evaluated a beacon-centric V2X/FL architecture for range-risk estimation in low-connectivity mountain-road segments. The central premise is that an EV entering such a segment should not rely exclusively on generic consumption factors or static range extrapolation. The energy required to complete the section depends on the segment itself, the current road-weather condition, the vehicle class, and the recent behaviour observed from previous traversals. The proposed beacon acts as a local holder of this segment-specific knowledge.

The experimental results support this premise. The observed energy demand of the same 10.6 km segment varies substantially across vehicle and road-weather conditions, from approximately 1.68 kWh for the Tesla Model 3 in dry conditions to more than 4.6 kWh for the Audi e-tron under snow and low temperature. Generic estimators do not fully capture this variability. The nominal estimator underestimates the required energy by 0.777 kWh on average, while the

conventional non-local physics estimator underestimates it by 1.975 kWh. The sign of the bias is important: the problem is not merely random error, but systematic optimism.

This systematic optimism has direct operational consequences. A route-energy error becomes safety-relevant when it changes the predicted final state of charge from below to above the minimum reserve threshold. For this reason, the study evaluated not only regression metrics but also alert-oriented metrics. The false-safe rate measures the proportion of cases in which the system fails to warn although the actual final SOC falls below the reserve threshold. This is the most relevant failure mode for the proposed application, because a false-safe decision may encourage a driver to enter an isolated segment with insufficient energy margin.

The alert-oriented evaluation shows the practical value of a segment-specific model. The conventional non-local estimator produces a false-safe rate of 14.58%, and the nominal estimator produces 8.33%. The segment-specific residual model reduces this value to 1.04% in leave-one-route-out evaluation. More importantly, the prequential beacon protocol shows progressive improvement as local evidence accumulates. With no previous vehicles, the beacon inherits the initial model and its false-safe rate is 14.58%. After two previous vehicles, the false-safe rate falls to 5.43%; after eight previous vehicles, it reaches 3.38%. Thus, even with a small simulated fleet, the beacon reduces the most safety-critical type of error. This reduction comes at a minor cost: the beacon produces some false warnings, unlike the central model, but this is an expected trade-off of a more sensitive alert policy.

The prequential design is essential to interpreting this result. Unlike a conventional offline evaluation, the beacon model is not allowed to use the current vehicle before predicting for it. Each vehicle benefits only from previous traversals, and only contributes after leaving the segment. This reproduces the causal structure of the intended deployment. The result therefore supports the idea that roadside infrastructure can accumulate local segment knowledge and make it available to subsequent vehicles without requiring continuous cloud connectivity.

The cross-validation results also clarify the scope of the learned model. The residual model performs well when tested on routes drawn from operating regimes represented in the training set, but its error increases under leave-one-weather-out and leave-one-vehicle-out protocols. This limitation is expected given the current dataset size: the evaluation includes only 12 traversals, two EV models and three weather regimes. Rather than invalidating the proposal, this finding reinforces the deployment logic. The model should not be interpreted as a universal estimator trained once and applied everywhere; it should be interpreted as a local, adaptive segment model that improves as vehicles actually cross the section under the relevant conditions.

From an architectural standpoint, the proposed system has three implications. First, the beacon should not simply broadcast a fixed energy value for the segment. Such a value would ignore private vehicle state such as payload, battery degradation, cabin-heating strategy or driver-specific consumption profile. Instead, the beacon should distribute a compact segment-specific model or model update, which the vehicle evaluates locally using its private information. Second, federated learning is justified by the ownership and locality of traversal data: the raw trace remains on the vehicle, while only model updates or compact statistics are shared. Third, the aggregation policy must be adapted to traffic volume. Immediate updating provides fast adaptation for low-traffic roads, whereas larger aggregation batches may provide stronger aggregation but delay usefulness. The federated component of this work should therefore be read as a demonstration that the beacon protocol is realisable under the communication constraints of a low-coverage segment using the standard semantics of FedAvg and FedProx, rather than as a comparison of aggregation rules.

Several limitations must be acknowledged. First, the evaluation is based on a digital twin and a longitudinal EV energy model rather than on real vehicle telemetry. Both vehicle configurations share that single model structure and differ only in parameters, so the vehicle-specific results evaluate the range-risk methodology within the digital-twin model and are not a validation of the energy-management behaviour of the corresponding production vehicles. Real deployments would introduce additional factors such as tyre pressure, battery ageing, HVAC control strategy, payload, traffic interaction, driver behaviour and manufacturer-specific powertrain control. Second, the number of traversals is small, which limits the statistical strength of the federated-learning comparison. Third, the current experiment evaluates range-risk alerts, not the full V2X/perception stack. Communication latency, packet delivery, power-consumption overhead, perception-risk detection and event-to-adaptation latency require dedicated instrumentation and are left for future work. Fourth, the risk threshold was fixed at 5% SOC; production deployments may require dynamic thresholds depending on distance to charger, outside temperature, vehicle health and driver preference.

To extend the evaluation framework beyond route-level energy estimation and range-risk alerting, Table 13 defines system-level KPIs for the deployment of the NOUS reference architecture. These KPIs cover communication continuity,

Table 13. System-level KPIs defined as future evaluation targets for the complete NOUS/V2X deployment.

KPI	Definition	Required future instrumentation
Section length	Length of the road segment covered by the beacon-based service, including low-coverage areas and any instrumented entry, exit or relay points.	Road-segment survey, beacon placement plan and communication-coverage characterisation.
No-update gap	Fraction of the segment in which a vehicle cannot receive a fresh update from a beacon or neighbouring vehicle.	Instrumented V2X communication stack, packet-delivery logs and position-synchronised coverage traces.
Stale-context fraction	Fraction of the segment in which the vehicle operates with context information older than the most recent relevant event affecting the section.	Event timestamping, context-version tracking and synchronised beacon/vehicle logs.
Event-to-adaptation latency	Time elapsed between the detection of a relevant road or risk event and the moment at which a following vehicle has incorporated the corresponding update into its operational policy.	End-to-end event propagation traces, V2X message logs and on-vehicle adaptation timestamps.
Risk-detection accuracy	F1-score or equivalent detection metric for identifying hazardous road-context conditions, such as low friction, obstacles, visibility loss or other perception-level risks.	Labelled perception dataset, validated risk annotations and perception models deployed on vehicles or roadside units.
Risk-mitigation rate	Fraction of detected hazardous situations for which the system adapts the vehicle or beacon policy before the affected point is reached.	Joint perception, communication and decision logs linking event detection, policy adaptation and vehicle position.
Range-estimation accuracy	Accuracy of the predicted energy demand or final state of charge for completing the segment.	Already evaluated in this paper through route-level energy-error and final-SOC metrics; future work should validate it with real telemetry.
Power-consumption overhead	Additional power or energy cost introduced by running the AI, V2X and FL components on the vehicle or roadside edge hardware.	Measurements on representative embedded platforms, including inference, communication and model-update workloads.
On-vehicle adaptation time	Time required by the vehicle to apply a received model, context or policy update after reception.	Embedded implementation on an in-vehicle or representative edge device, with update-loading and inference-readiness timestamps.

context freshness, event-driven adaptation, perception-level risk detection, embedded execution cost and on-vehicle update time. They provide a structured validation roadmap for future deployments of the beacon–V2X–FL architecture.

Future work should therefore extend the evaluation in four directions. First, the simulation campaign should include additional vehicle classes, payload levels, battery-health states, driving styles and traffic conditions. Second, real or semi-real telemetry should be collected on the selected mountain segment to validate the digital-twin conclusions. Third, the FL protocol should incorporate asynchronous updates, secure aggregation, robust aggregation and update compression, all of which are relevant in sparse vehicular networks. Fourth, the system-level KPIs defined in Table 13 should be evaluated in an instrumented deployment that includes perception data, V2X communication traces and representative embedded hardware. Finally, the binary alert policy should be extended into graded warnings, for example normal, low-margin and high-risk states, so that the driver receives more interpretable guidance than a single safe/unsafe decision.

In conclusion, the study shows that range-risk estimation on isolated mountain segments benefits from a model that is specific to the segment and updated from recent traversals. Generic or non-local estimators can be systematically optimistic, producing false-safe decisions at low entry SOC. A beacon-centric V2X/FL architecture provides a technically plausible mechanism to reduce this risk: vehicles receive a local model before entering the section, evaluate it with their private state, and contribute updates after completing the traversal. The most important result is not only a reduction in average energy-estimation error, but the progressive reduction of false-safe alerts as the beacon observes more vehicles. This supports the proposed architecture as a privacy-preserving and locally adaptive range-risk mitigation layer for challenging road segments.

Ethics and consent
Ethical approval and consent were not required.

Software Availability
The software supporting this study comprises the Python pipeline for preprocessing, baseline computation, residual-model training, Flower-based federated training and prequential beacon evaluation (NumPy, pandas, scikit-learn, PyTorch, Flower, Matplotlib and PyYAML; federated experiments use `flwr [simulation]` in the >= 1.8,<2.0 range), together with the mapperV3 digital-twin middleware, which is a MATLAB App Designer tool, also distributed as a standalone Windows executable requiring no MATLAB licence, that fuses a GeoTIFF DEM and an OpenStreetMap road network into an elevation-aware RoadRunner HD Map.

- Source code available from: https://github.com/AIRInstitute/nous_mountain
- Archived software available from: https://doi.org/10.5281/zenodo.21159398
- License: MIT

Data Availability
All data underlying the results are openly available so that the findings in Section 4 can be reproduced independently. The complete dataset (Vidal Cuevas et al., 2026) is archived on Zenodo under the persistent identifier https://doi.org/10.5281/zenodo.21159398 and comprises:

- **Underlying data**. Twelve simulated EV traversal logs of the mountain-road segment (2 EV models × 2 initial SOC levels × 3 road-weather conditions), generated from the digital-twin pipeline of Section 3. Each CSV file stores the time-stamped variables recorded along the traversal: elapsed time (s), position in the simulator world frame (pos_x, pos_y, pos_z, in metres, with pos_z the mesh elevation), vehicle speed (speed_kmh, km/h), road grade derived from vehicle pitch (road_grade_deg, degrees), instantaneous battery power (power_watts, W) and cumulative used and regenerated energy (energy_used_cum_wh, energy_regen_cum_wh, Wh), battery state of charge (soc_percent, %), and the active road-weather condition (road_condition). The record contains the twelve raw traversal logs, which are the underlying data of the study, together with the complete analysis pipeline and the configuration files used, so that the processed segment-level table, the route-level predictions, the risk-decision tables and the prequential learning-curve summaries behind Tables 5 to 12 can be regenerated in full by running the archived pipeline on the archived logs.
- **Extended data**. The digital twin of the Puerto de la Quesera pass for CARLA Simulator 0.9.16 (33 roads, 884 elevation primitives, spanning the 1242–1494 m band above sea level): the diagnostic scripts, which are included in the repository, and the map source (mesh, OpenDRIVE network and metadata), which is distributed within the external import package. The full deployable import package (288 MB) exceeds the repository file-size limit and is linked externally from the record.
- **Configuration**. The configuration files used for cross-validation and prequential beacon evaluation, required to reproduce the reported results.

The twelve traversal logs are entirely original simulation output and are dedicated to the public domain under the Creative Commons CC0 1.0 Universal dedication. The CARLA digital-twin map assets are released under the Creative Commons Attribution 4.0 International license (CC-BY 4.0) rather than CC0, because they derive from OpenStreetMap road data (© OpenStreetMap contributors, Open Database License) and from the Copernicus GLO-30 Digital Elevation Model (© European Space Agency/European Union, Copernicus Programme), both of which impose non-waivable downstream attribution obligations.

Acknowledgements
This research has been supported by the project "A catalyst for EuropeaN ClOUd Services in the era of data spaces, high-performance and edge computing (NOUS)", Grant Agreement Number 101135927. Funded by the European Union, views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

Abadi M, Chu A, Goodfellow I, *et al.*: **Deep learning with differential privacy.** *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.* 2016; 308–318.
Publisher Full Text

Alistarh D, Grubic D, Li J, *et al.*: **QSGD: Communication-efficient SGD via gradient quantization and encoding.** *Adv. Neural Inf. Proces. Syst.* 2017; **30**.
Reference Source

Bao Z, Hossain S, Lang H, *et al.*: **A review of high-definition map creation methods for autonomous driving.** *Eng. Appl. Artif. Intell.* 2023; **122**: 106125.
Publisher Full Text

Bernstein J, Wang Y-X, Azizzadenesheli K, *et al.*: **signSGD: Compressed optimisation for non-convex problems.** *Proceedings of the 35th International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research.* 2018; **80**: 560–569.
Reference Source

Beutel DJ, Topal T, Mathur A, *et al.*: **Flower: A friendly federated learning research framework.** *arXiv Preprint arXiv:2007.14390.* 2020.

Bilal M, Munir EU, Alarood AA: **BEMD: Beacon-oriented emergency message dissemination scheme for highways.** *Ad Hoc Netw.* 2023; **142**: 103095.
Publisher Full Text

Bonawitz K, Ivanov V, Kreuter B, *et al.*: **Practical secure aggregation for privacy-preserving machine learning.** *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS).* 2017; 1175–1191.
Publisher Full Text

Cao X, Fang M, Liu J, *et al.*: **FLTrust: Byzantine-robust federated learning via trust bootstrapping.** *Proceedings of the 28th Annual Network and Distributed System Security Symposium (NDSS).* 2021.
Publisher Full Text

Chellapandi VP, Yuan L, Brinton CG, *et al.*: **Federated learning for connected and automated vehicles: A survey of existing approaches and challenges.** *IEEE Transactions on Intelligent Vehicles.* 2024; **9**(1): 119–137.
Publisher Full Text

Chitty-Venkata KT, Mittal S, Emani M, *et al.*: **A survey of techniques for optimizing transformer inference.** *J. Syst. Archit.* 2023; **144**: 102990.
Publisher Full Text

Choi S, Kim M, Park D: **Optimal beamforming in over-the-air federated learning for efficient model aggregation.** *ICT Express.* 2025; **12**: 136–141.
Publisher Full Text

Clancy J, Mullins D, Deegan B, *et al.*: **Wireless access for V2X communications: Research, challenges and opportunities.** *IEEE Commun Surv Tutor.* 2024; **26**(3): 2082–2119.
Publisher Full Text

Dosovitskiy A, Ros G, Codevilla F, *et al.*: **CARLA: An open urban driving simulator.** *Proceedings of the 1st Annual Conference on Robot Learning (CoRL), Proceedings of Machine Learning Research.* 2017; **78**: 1–16.
Reference Source

Du Z, Wu C, Yoshinaga T, *et al.*: **Federated learning for vehicular Internet of Things: Recent advances and open issues.** *IEEE Open Journal of the Computer Society.* 2020; **1**: 45–61.
PubMed Abstract | **Publisher Full Text**

ETSI Technical Committee ITS: *Intelligent transport systems (ITS); security; security header and certificate formats ({ETSI Technical Specification} ETSI TS 103 097 V1.3.1).* European Telecommunications Standards Institute; 2017.
Reference Source

Gao W, Tavallaie O, Chen S, *et al.*: **Federated learning as a service for hierarchical edge networks with heterogeneous models.** *Service-Oriented Computing – ICSOC 2024, Lecture Notes in Computer Science.* 2025; **15404**: 85–99.
Publisher Full Text

Ghazali AK, AbAziz NAAb, Hassan MK: **Advanced algorithms in battery management systems for electric vehicles: A comprehensive review.** *Symmetry.* 2025; **17**(3).
Publisher Full Text

Gong T, Zhu L, Yu FR, *et al.*: **Edge intelligence in intelligent transportation systems: A survey.** *IEEE Trans. Intell. Transp. Syst.* 2023; **24**(9): 8919–8944.
Publisher Full Text

He H, Wang Z, Jia M, *et al.*: **Analysis of factors influencing energy consumption of electric vehicles: Statistical, predictive, and causal perspectives.** *Appl. Energy.* 2024; **375**: 124110.
Publisher Full Text

Herabad MG: **Communication-efficient semi-synchronous hierarchical federated learning with balanced training in heterogeneous IoT edge environments.** *Internet of Things.* 2023; **21**: 100642.
Publisher Full Text

Huang H, He H, Wang Y, *et al.*: **Energy consumption prediction of electric vehicles for data-scarce scenarios using pre-trained model.** *Transp. Res. Part D: Transp. Environ.* 2025; **146**: 104830.
Publisher Full Text

Hussain I, Ching KB, Uttraphan C, *et al.*: **Evaluating machine learning algorithms for energy consumption prediction in electric vehicles: A comparative study.** *Sci. Rep.* 2025a; **15**(1): 16124.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Hussain I, Ching KB, Uttraphan C, *et al.*: **Optimizing electric vehicle energy consumption prediction through machine learning and ensemble approaches.** *Sci. Rep.* 2025b; **15**(1): 29065.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Jarke M, Otto B, Ram S: **Data sovereignty and data space ecosystems.** *Bus. Inf. Syst. Eng.* 2019; **61**(5): 549–550.
Publisher Full Text

Jonas T, Wilde T, Hunter CD, *et al.*: **The impact of road types on the energy consumption of electric vehicles.** *J. Adv. Transp.* 2022; **2022**(1): 1–14.
Publisher Full Text

Jung MAT, Schröder F: **Exponential adoption of battery electric cars.** *PLoS One.* 2023; **18**(12): e0295692.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Karimireddy SP, Kale S, Mohri M, *et al.*: **SCAFFOLD: Stochastic controlled averaging for federated learning.** *Proceedings of the 37th International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research.* 2020; **119**: 5132–5143.
Reference Source

Kong X, Chen Z, Liu W, *et al.*: **Deep learning for time series forecasting: A survey.** *Int. J. Mach. Learn. Cybern.* 2025; **16**: 5079–5112.
Publisher Full Text

Lai F, Dai Y, Singapuram SS, *et al.*: **FedScale: Benchmarking model and system performance of federated learning at scale.** *Proceedings of the 39th International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research.* 2022; **162**: 11814–11827.
Reference Source

Li T, Sahu AK, Zaheer M, *et al.*: **Federated optimization in heterogeneous networks.** *Proceedings of the 3rd Conference on Machine Learning and Systems (MLSys).* 2020; **2**: 429–450.
Reference Source

Liu H-I, Galindo M, Xie H, *et al.*: **Lightweight deep learning for resource-constrained environments: A survey.** *ACM Comput. Surv.* 2024; **56**(10): Article 267.
Publisher Full Text

Liu K, Li K, Peng Q, *et al.*: **A brief review on key technologies in the battery management system of electric vehicles.** *Front. Mech. Eng.* 2018; **14**(1): 47–64.
Publisher Full Text

Lopez PA, Behrisch M, Bieker-Walz L, *et al.*: **Microscopic traffic simulation using SUMO.** *2018 21st International Conference on Intelligent Transportation Systems (ITSC).* 2018; 2575–2582.
Publisher Full Text

Lu R, Zhang W, Li Q, *et al.*: **Adaptive asynchronous federated learning.** *Futur. Gener. Comput. Syst.* 2024; **152**: 193–206.
Publisher Full Text

Maglogiannis V, Naudts D, Hadiwardoyo S, *et al.*: **Experimental V2X evaluation for C-V2X and ITS-G5 technologies in a real-life highway environment.** *IEEE Trans. Netw. Serv. Manag.* 2022; **19**(2): 1521–1538.
Publisher Full Text

Mao R, Xu W, Qian Y, *et al.*: **Understanding the determinants of electric vehicle range: A multi-dimensional survey.** *Sustainability.* 2025; **17**(10).
Publisher Full Text

McMahan HB, Moore E, Ramage D, *et al.*: **Communication-efficient learning of deep networks from decentralized data.** *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Proceedings of Machine Learning Research.* 2017; **54**: 1273–1282.
Reference Source

Mulpuri SK, Sah B, Kumar P: **An intelligent battery management system (BMS) with end-edge-cloud connectivity – a perspective.** *Sustainable Energy Fuels.* 2025; **9**: 1142–1159.
Publisher Full Text

Ngo D, Park H-C, Kang B: **Edge intelligence: A review of deep neural network inference in resource-limited environments.** *Electronics.* 2025; **14**(12): 2495.
Publisher Full Text

Noor-A-Rahim M, Liu Z, Lee H, *et al.*: **6G for vehicle-to-everything (V2X) communications: Enabling technologies, challenges, and**

opportunities. *Proc. IEEE.* 2022; **110**(6): 712–734.
Publisher Full Text

OpenStreetMap contributors: **OpenStreetMap.** n.d.
Reference Source

OpenTopography: **OpenTopography: High-resolution topography data and tools.** n.d.
Reference Source

Otto B: **The evolution of data spaces.** Otto B, ten Hompel M, Wrobel S, editors. *Designing data spaces: The ecosystem approach to competitive advantage.* Springer; 2022; pp. 3–15.
Publisher Full Text

Prasad SL, Gudipalli A: **An effective range estimation and state-of-charge to mitigate range anxiety in electric vehicles.** *Heliyon.* 2025; **11**(1): e41494.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rahmath PH, Srivastava V, Chaurasia K, *et al.*: **Early-exit deep neural network—a comprehensive survey.** *ACM Comput. Surv.* 2024; **57**(3): Article 75.
Publisher Full Text

Reddi SJ, Charles Z, Zaheer M, *et al.*: **Adaptive federated optimization.** *Proceedings of the 9th International Conference on Learning Representations (ICLR).* 2021.
Reference Source

Sandberg U, Haider M, Conter M, *et al.*; **Rolling resistance – basic information and state-of-the-art on measurement methods.** *MIRIAM (Models for rolling resistance in Road Infrastructure Asset Management Systems), Deliverable #1 in MIRIAM SP1.* 2011.
Reference Source

Somvanshi S, Islam MM, Chhetri G, *et al.*: **From tiny machine learning to tiny deep learning: A survey.** *ACM Comput. Surv.* 2025; **58**(7): Article 168.
Publisher Full Text

Torres JF, Hadjout D, Sebaa A, *et al.*: **Deep learning for time series forecasting: A survey.** *Big Data.* 2021; **9**(1): 3–21.
Publisher Full Text

United Nations Economic Commission for Europe: *Global technical regulation no. 15: Worldwide-harmonized light vehicles test procedure (WLTP) ({UN Global Technical Regulation} ECE/TRANS/180/Add.15/Amend.6).* UNECE, World Forum for Harmonization of Vehicle Regulations (WP.29); 2022.
Reference Source

Vidal Cuevas JE, Mezquita Y, Valdeolmillos D, *et al.*: Mountain-Pass EV Energy: Digital-Twin Dataset and Federated Beacon Experiments. [Dataset]. *Zenodo.* 2026.
Publisher Full Text

Wan L, Zhao J, Wiedholz A, *et al.*: **A systematic literature review on vehicular collaborative perception—a computer vision perspective.** *IEEE Transactions on Intelligent Transportation Systems.* 2025; PP 1–38.
Publisher Full Text

Wang J, Topilin I, Feofilova A, *et al.*: **Cooperative intelligent transport systems: The impact of c-V2X communication technologies on road safety and traffic efficiency.** *Sensors.* 2025; **25**(7).
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Weiss M, Cloos KC, Helmers E: **Energy consumption of electric vehicles in Europe.** *Sustainability.* 2024; **16**(17): 7529.
Publisher Full Text

Xiao B, Yu X, Ni W, *et al.*: **Over-the-air federated learning: Status quo, open challenges, and future directions.** *Fundam. Res.* 2024; **5**: 1710–1724.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Xu R, Xia X, Li J, *et al.*: **V2V4Real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception.** *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2023; 13712–13722.
Publisher Full Text

Yang Y, Long G, Lu Q, *et al.*: **Federated low-rank adaptation for foundation models: A survey.** *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI).* 2025; 10779–10787.
Publisher Full Text

Yang Y, Long G, Shen T, *et al.*: **Dual-personalizing adapter for federated foundation models.** *Advances in Neural Information Processing Systems (NeurIPS).* 2024; **37**.
Reference Source

Yin D, Chen Y, Kannan R, *et al.*: **Byzantine-robust distributed learning: Towards optimal statistical rates.** *Proceedings of the 35th International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research.* 2018; **80**: 5650–5659.
Reference Source

Yu H, Luo Y, Shu M, *et al.*: **DAIR-V2X: A large-scale dataset for vehicle-infrastructure cooperative 3D object detection.** *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2022; 21361–21370.
Publisher Full Text

Yuan X, He J, Li Y, *et al.*: **Data-driven evaluation of electric vehicle energy consumption for generalizing standard testing to real-world driving.** *Patterns.* 2024; **5**(4): 100950.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Zang Y, Xue Z, Ou S, *et al.*: **Efficient asynchronous federated learning with prospective momentum aggregation and fine-grained correction.** *Proceedings of the AAAI Conference on Artificial Intelligence.* 2024; **38**: 16642–16650.
Publisher Full Text

Zhang S, Li J, Shi L, *et al.*: **Federated learning in intelligent transportation systems: Recent applications and open problems.** *IEEE Trans. Intell. Transp. Syst.* 2024; **25**(5): 3259–3285.
Publisher Full Text

Zhou H, Zheng Y, Huang H, *et al.*: **Toward robust hierarchical federated learning in Internet of Vehicles.** *IEEE Trans. Intell. Transp. Syst.* 2023; **24**(5): 5600–5614.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 03 September 2026

https://doi.org/10.21956/openreseurope.26595.r79537

© **2026 Mądziel M.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Maksymilian Mądziel
Rzeszów University of Technology, Rzeszów, Poland

The manuscript addresses an interesting and practically relevant problem: estimating EV range risk on connectivity-poor mountain road segments by combining a segment-specific residual energy model with a beacon-centric federated-learning architecture. The study is clearly motivated, the digital-twin workflow is described in considerable detail, and the use of false-safe decisions as a safety-oriented KPI is particularly appropriate. However, I believe that several technical aspects require clarification before the manuscript can be considered fully convincing.

1. Clarification of the relationship between WLTP/rated range and on-board range estimation.

The statement in the Introduction (page 3) and Section 2.6 that range information presented to drivers is “most often anchored in type-approval cycles such as WLTP” appears too broad and potentially misleading from an automotive engineering perspective. A distinction should be made between (i) homologated/rated range derived from a standardized certification procedure and (ii) dynamic on-board estimation of remaining range or route-specific energy consumption. Modern production EVs may use different estimation strategies depending on the displayed quantity and operating mode, including rated-consumption baselines, recent energy consumption, predicted speed, environmental conditions, and, when navigation is active, route-related information. The authors should therefore avoid implying that the on-board range estimator is generally a direct implementation of WLTP data and should clarify precisely which type of range information is meant in this statement.

2. The validation of the pitch-derived road-grade signal should be strengthened.

The road grade is derived from vehicle.get_transform().rotation.pitch because the CARLA waypoint API returns $z = 0$ in the present configuration. The manuscript states that this pitch-based grade was validated against literature-based predictions of grade-induced energy consumption. However, this is an indirect validation of the grade signal itself. Since vehicle pitch may contain contributions from longitudinal acceleration, braking, and suspension dynamics, agreement with an energy-consumption relationship does not demonstrate that the estimated pitch angle accurately represents the geometric road slope. The authors should provide a direct quantitative validation of the pitch-derived grade against a reference grade obtained from the 3-D road/elevation geometry, or otherwise report the magnitude of the discrepancy between the two

signals. Any filtering or smoothing applied to the pitch signal should also be explicitly described.
3. The description of SOC updating as "Coulomb counting" should be corrected or clarified.
In Section 3.4.1, the authors state that SOC is updated via Coulomb counting, but the subsequent description indicates that battery power (P_batt(t)) is numerically integrated over time and normalized by usable battery capacity. This is an energy-based integration rather than conventional Coulomb counting, which is based on the integration of battery current over time. Please clarify the actual SOC-update formulation and revise the terminology accordingly. This distinction is relevant because SOC is subsequently used to define the ground-truth risk state and the final-SOC evaluation metrics.
4. The manuscript should distinguish more clearly between an energy-model study and validation of production-EV behaviour.
Although two production EV models are used, both are simulated using the same simplified longitudinal energy-model structure, with differences introduced through vehicle-specific parameters. The resulting predictions should therefore be interpreted primarily as an evaluation of the proposed range-risk methodology within the digital-twin model, rather than as a validation of the actual energy-management behaviour of Tesla and Audi production vehicles. The manuscript partially acknowledges this limitation, but I recommend making this distinction explicit when presenting the vehicle-specific results and discussing the broader implications of the study.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Transport Engineering; Microscopic Traffic Simulation; Electric Vehicles (EV) Energy Modeling; Vehicle Dynamics and Emissions Modeling

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 10 August 2026

https://doi.org/10.21956/openreseurope.26595.r78623

© 2026 Simsek A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Ahmed Ihsan Simsek
Fırat Üniversitesi, Elazığ, Turkey

Overall Assessment
This is a carefully conducted and enjoyable study. It focuses on a real problem, is honest and well-documented. The code, data, and map are shared under open licenses, and the digital-twin pipeline represents significant engineering effort. The most valuable contribution of the study is its selection of asymmetric false-safe cost as the primary KPI instead of average error. It correctly identifies that misleading a vehicle into "passing" is far more critical than giving unnecessary warnings, and this distinction is carefully formulated. The design of the prequential protocol to reflect the causal structure in the field (each vehicle is first evaluated with the existing model, but then the model is updated) is also a sound choice. These two choices form the backbone of the study and are well-executed.
My overall impression is positive, and I find the article suitable for publication. The following points are primarily suggestions to strengthen the presentation and more closely align claims with available evidence. They do not require a new experiment. Actually they are largely improvements at the framing and explanation level.
Improvement Suggestions
1. The proof-of-concept nature of the study could be emphasized.
The results are based on 12 drives, and as the authors state, the effect of SOC is limited, and the snow condition is modeled with fixed physical multipliers. Therefore, the number of conditions that produce significant independent variation is effectively around 2 vehicles × 3 weather. This scale is quite suitable for demonstrating the approach. Positioning the study explicitly as a proof-of-concept in the title and abstract would better align the readers expectations with the result.
2. The contribution of federated learning could be made a little more visible.
FL is central to the approach, and the article presents both central (M1) and federated/prequential (M2/M3) results. Presenting the central and federated models side-by-side on a common axis would clarify what FL contributes to the reader. Given that the difference between FedProx and FedAvg is small at this sample size, presenting this comparison as a "demonstration of applicability with standard tools" would also be a reasonable alternative.
3. The source of the slope signal could be mentioned in the abstract.
The key aspect of the pipeline is to transfer the actual elevation data to the energy model. Since the CARLA API returns z=0 in the waypoint network, the slope is derived in terms of vehicle pitch. This is a practical and creative solution. However since pitch can be affected by acceleration and suspension, briefly mentioning this derivation in the abstract and supporting it with a simple quantitative validation (e.g., comparison at known slopes) would strike a good balance with the emphasis on "elevation preserved".
4. The threshold and number of observations could be explained a little more.
Warning decisions are generated with a fixed 5% threshold, and 96 decisions come from

replicating 12 drives with 8 synthetic input-SOC values. Briefly stating that these are not 96 completely independent samples and, if possible, including a single threshold-sensitivity analysis would help interpret decimals such as 1.04%. Mentioning the actual number of independent observations (12) would also suffice.

5. The inter-vehicle generalization limit could also be made visible in the abstract.

The authors present leave-one-vehicle-out as a deliberate out-of-distribution stress test and honestly acknowledge the high error (1.568 kWh) in the text and this is a commendable approach. The brief mention of this limit at the abstract level and the statement that both vehicles were simulated with the same energy model (with only the parameters changed) ensures the reader's accurate understanding of the scope.

I also have some minor and mostly editorial points. I provided below.

Minor points

• The tool name appears as "mapperV2" in the text, but as "mapperV3" in the figure and software sections. I think merging them would be beneficial.

• "595 m relief" refers to the DEM; the band the vehicle passed through is approximately 250 m (1242–1494 m). Briefly mentioning this distinction would be helpful.

• The risk recall = 0.00 value of the non-local estimator is noteworthy. Maybe a one-sentence comment would be helpful for the reader.

Conclusion

This study is an honest, reproducible, and valuable proof of concept. Its fundamental finding is valid, and its operational framing is strong. The above suggestions are largely at the presentation and framing level and do not require new experiments; bearing these in mind, I would gladly support the paper for full approval.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and does the work have academic merit?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I work on decision support systems, MCDM, disaster logistics, machine learning and deep reinforcement learning. I am less focused on the V2X communications and embedded hardware layers.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.