Check for updates

RESEARCH ARTICLE

REVISED Web and social event signals for AI-driven mobile network demand forecasting

[version 2; peer review: 2 approved with reservations]

Marcello Pietri, Marco Mamei

DISMI, Universita degli Studi di Modena e Reggio Emilia Dipartimento di Scienze e Metodi dell'Ingegneria, Reggio Emilia, Emilia-Romagna, Italy

v2 **First published:** 11 Apr 2026, **6**:106
https://doi.org/10.12688/openreseurope.23356.1
Latest published: 07 Oct 2026, **6**:106
https://doi.org/10.12688/openreseurope.23356.2

Abstract

Background

Mobile network demand is increasingly volatile due to collective human activities such as concerts, sports events, and public gatherings. Traditional capacity planning methods, largely based on historical network indicators, struggle to anticipate these transient and localized demand surges. Recent advances in social sensing and artificial intelligence suggest that web and social signals can provide early indicators of real-world events, enabling proactive, event-aware network management in future mobile network infrastructures.

Methods

This paper presents a combined analysis and empirical study on AI-driven event-aware demand forecasting for mobile networks. We first review methods for extracting event signals from web and social data and analyze prior evidence linking such signals to cellular traffic variations. We then introduce a forecasting-driven orchestration pipeline and evaluate it through a case study using the NetMob'23 dataset, which provides high-resolution, service-level mobile traffic traces from multiple urban areas. The empirical evaluation focuses on the city of Nantes. Several forecasting models—ranging from naïve baselines and linear regression to Random Forests and LSTM neural networks, and a PatchTST-style Transformer baseline —are compared. We further investigate the impact of event-related features based on scheduled-event contextual information and temporally aligned event indicators and introduce an asymmetric loss function designed to penalize traffic underestimation in proactive orchestration scenarios.

Open Peer Review

Approval Status ? ?

	1	2
version 2 (revision) 07 Oct 2026	? view	
version 1 11 Apr 2026	? view	? view

1. **Ashok P**, Symbiosis Institute of Digital and Telecom Management (SIDTM), Pune, India

2. **Christos Ntanos**, National Technical University of Athens, Zografou, Greece
Afroditi Blika, National Technical University of Athens, Zografou, Greece

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

Results

Among the evaluated models, LSTM opt_std achieves the lowest RMSE (1.94×10^7) and the highest R2 (0.936), whereas the asymmetric-loss LSTM opt_cst achieves the largest relative overload reduction, reaching 71.3%, 69.1%, and 61.2% at the 70th, 80th, and 90th percentile capacity thresholds, respectively, with RMSE =2.24×10^7 and MAPE =24.01%. For opt_cst, removing event-related inputs increases RMSE from 2.24×10^7 to 2.59×10^7 and MAPE from 24.01% to 28.9%, corresponding to reductions of 13.4% and 16.9% when event-aware inputs are used. Event-input perturbations produce modest and non-monotonic sensitivity, with RMSE changes between −1.90% and + 0.21% and OR@80 changes between −1.37 and + 1.19 percentage points relative to the clean condition.

Conclusions

The study confirms that event-aware contextual features into AI-based forecasting pipelines provides a measurable anticipatory advantage for proactive mobile network orchestration. Event-aware forecasting emerges as a promising component for predictive, self-optimizing future mobile network infrastructures, bridging social sensing and automated network management.

Keywords
5G networks, demand forecasting, social media

HE This article is included in the Horizon Europe gateway.

Corresponding author: Marco Mamei (marco.mamei@unimore.it)

Author roles: Pietri M: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Mamei M**: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101135927 (A catalyst for EuropeaN ClOUd Services in the era of data spaces, high-performance and edge computing - NOUS) This research has also been supported by Fondo di Ateneo per la Ricerca 2025, Ricerca Diffusa, Piano di Sviluppo Dipartimentale, UniMORE.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Pietri M and Mamei M. 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: Pietri M and Mamei M. **Web and social event signals for AI-driven mobile network demand forecasting [version 2; peer review: 2 approved with reservations]** Open Research Europe 2026, **6**:106
https://doi.org/10.12688/openreseurope.23356.2

First published: 11 Apr 2026, **6**:106 https://doi.org/10.12688/openreseurope.23356.1

REVISED **Amendments from Version 1**

The abstract was revised with following Among the evaluated models, LSTM opt_std achieves the lowest RMSE (1.94×10^7) and the highest R2 (0.936), whereas the asymmetric-loss LSTM opt_cst achieves the largest relative overload reduction, reaching 71.3%, 69.1%, and 61.2% at the 70th, 80th, and 90th percentile capacity thresholds, respectively, with RMSE $=2.24\times10^7$ and MAPE =24.01%. For opt_cst, removing event-related inputs increases RMSE from 2.24×10^7 to 2.59×10^7 and MAPE from 24.01% to 28.9%, corresponding to reductions of 13.4% and 16.9% when event-aware inputs are used. Event-input perturbations produce modest and non-monotonic sensitivity, with RMSE changes between -1.90% and +0.21% and OR@80 changes between -1.37 and +1.19 percentage points relative to the clean condition. Introduction part updated.

Experimental section updated with The experimental workflow couples one-hour-ahead traffic forecasts with a simplified retrospective capacity benchmark. The purpose of this benchmark is to quantify how forecast-informed provisioning can reduce traffic exceeding a fixed reference capacity. It is not intended to reproduce a complete RAN, core-network, or operator resource-allocation process.

Each spatial node is evaluated through its service-cluster traffic series. We assume a simplified monotonic relationship between observed traffic volume and required capacity and abstract away detailed radio scheduling, handover dynamics, compute placement, and nonlinear resource-consumption effects. This abstraction allows forecasting methods to be compared under a common capacity-constrained criterion without introducing network-topology assumptions that are not available in NetMob'23 dataset. Figures 4,5,6,7 revised and updated.

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

Introduction

The proliferation of data-intensive and latency-sensitive applications has made mobile network demand increasingly volatile and spatially uneven. Collective human activities—such as concerts, demonstrations, sport events, or seasonal migrations—can abruptly overload localized portions of the Radio Access Network (RAN), producing service degradation, Quality-of-Experience (QoE) drops, and energy inefficiencies. In 5G deployments, where resource utilization is tightly coupled with user mobility and service heterogeneity, such fluctuations pose severe challenges to network operators that must ensure consistent performance across distributed edge infrastructures.

Traditional traffic engineering and capacity planning methods, mostly based on historical Key Performance Indicators (KPIs) and periodic reconfigurations, lack the ability to anticipate transient surges and to react at the temporal and spatial granularity demanded by modern applications. Reactive scaling and over-provisioning are no longer sustainable as networks evolve toward massive device connectivity and strict Service-Level Agreements (SLAs) for latency-critical services such as augmented reality or connected mobility. This motivates the shift toward predictive orchestration, where resource allocation decisions are guided by demand forecasts and situational awareness rather than post hoc monitoring.

Recent progress in Artificial Intelligence (AI) and social sensing provides new avenues to anticipate network demand through external, event-related signals. Web search queries, trending hashtags, and digital event listings often precede real-world gatherings, offering a valuable proxy for human mobility and network usage dynamics. For instance, a surge in X (ex. Twitter) activity about a major football match or a spike in Google Trends related to a music festival often translates into localized bandwidth pressure on surrounding cells. Evidence from cellular-network studies shows that social activity and crowded events can correlate with mobile demand and RAN performance (B. Yang et al. 2016; Shafiq et al. 2013). Related work in road traffic and urban mobility further shows that social-media and web signals can provide useful contextual predictors in adjacent domains (Essien et al. 2021; Yao, Zhang, and Zhang 2021; X. Yang, Bekoulis, and Deligiannis 2023). These findings motivate the investigation of external contextual signals for proactive mobile-network management. At the network management level, dynamic resource sharing and slicing have been extensively investigated as enablers of flexible capacity allocation. Márquez et al. (Márquez et al. 2018, 2019) demonstrated the efficiency gains of adaptive network slicing and motivated the need for anticipatory orchestration mechanisms that can predict demand heterogeneity across multiple services and slices. Building on this perspective, our previous work (Pietri et al. 2025) explored multi-layer edge orchestration in 5G environments, where early detection of demand peaks was leveraged to inform edge workload placement and traffic redirection between far-edge and MEC tiers. That study already evidenced that event-aware forecasting can substantially reduce overload periods and improve the fairness of resource distribution.

Despite these advances, the end-to-end integration of event detection, demand forecasting, and automated orchestration remains an open research frontier. Existing approaches often address isolated components—such as social event detection (Hu et al. 2022; Mondal et al. 2025) or network resource control—but rarely connect them into a unified, predictive management loop. Key open questions include: *(i)* the reliability, geo-temporal alignment, and transferability

of web/social signals with respect to network KPIs; *(ii)* the quantification of forecast lead time required for operational decisions; and *(iii)* the design of learning-based controllers capable of embedding such forecasts into closed-loop orchestration frameworks compliant with 5G/6G standards.

This paper contributes a concise yet comprehensive analysis on AI-driven event-aware demand forecasting for mobile and edge networks. It synthesizes the state of the art across three complementary layers: *(i)* event and social-signal detection, *(ii)* empirical correlation between web activity and mobile demand, and *(iii)* AI-assisted orchestration for adaptive network slicing and workload allocation. To ground the discussion in practice, we further complement the analysis with a case study on real-world mobile traffic, comparing classical and AI-based forecasting pipelines under realistic operational constraints. The conceptual flow of these components is summarized in Figure 1, which depicts the end-to-end pipeline from web and social signal acquisition to network-level orchestration actions. By consolidating evidence from both communication systems and data-driven social analytics, this study outlines a forward-looking research agenda toward proactive, event-aware management in future mobile network infrastructures.

State of the art in event-aware forecasting and network orchestration

The interplay between social signals, event detection, and mobile network dynamics has attracted increasing research interest, bridging the areas of social sensing, human mobility analysis, traffic forecasting, and AI-based network automation. This section reviews three complementary domains forming the foundation of event-aware orchestration: *(i)* social sensing and event detection, *(ii)* empirical correlation between social/web activity and mobile network demand, and *(iii)* AI-driven network management and orchestration frameworks.

Social sensing and event detection

Social sensing exploits large-scale user-generated content to infer real-world phenomena. Early approaches such as that from Walther and Kaisser (Walther and Kaisser 2013) proposed geo-spatial event detection from Twitter streams, while Figure 1: End-to-end pipeline for event-aware AI-driven demand forecasting and proactive network managementGu et al. (Gu, Qian, and Chen 2016) and Zhang et al. (Zhang et al. 2018) applied machine learning and deep models to detect traffic incidents from tweets in real time. Subsequent studies expanded the temporal and semantic scope of event extraction: Essien and Petrovic (Essien et al. 2021) incorporated contextual embeddings to improve traffic-flow prediction, and Yang et al. (X. Yang, Bekoulis, and Deligiannis 2023) reformulated event detection as a slot-filling task under neural architectures. More recent surveys by Karimiziarani et al. (Karimiziarani, Foroumandi, and Moradkhani 2022), Hu et al. (Hu et al. 2022), and Mondal et al. (Mondal et al. 2025) provide comprehensive taxonomies of methodologies, ranging from burst and change-point detection to multimodal fusion of text, image, and geolocation data. Mredula et al. (Mredula, Behera, et al. 2022) further emphasized adaptive models and streaming architectures capable of handling concept drift and evolving language. These works collectively highlight the maturity of event detection but also reveal the lack of systematic integration with downstream network-control applications.

From social signals to network demand

Parallel research lines have investigated the correlation between online social activity and mobile data consumption. Yang et al. (B. Yang et al. 2016) provided one of the first empirical evidences that Twitter post volume can explain up to 70% of variance in cellular demand. Yao et al. (Yao, Zhang, and Zhang 2021) demonstrated that augmenting traffic models with social-media features improves next-day forecasts of urban mobility patterns. Similarly, Shafiq et al. (Shafiq et al. 2013) analyzed network performance during large gatherings, showing consistent throughput degradation and signaling overload in crowded scenarios. Complementary studies exploiting Google Trends data (Mavragani, Ochoa, and Tsagarakis 2018; Jun, Yoo, and Choi 2018) confirmed that search query volumes are predictive of human activity and can serve as external regressors for traffic forecasting.

Industrial reports reinforce these findings. The Ericsson Mobility Report ("Enhancing Event Experience — Mobility Report (Case Studies)" 2016) documented case studies of temporary network congestion during mass events such as concerts and sport matches, leading to the notion of "event-aware" capacity management. More recent analyses (Zanella

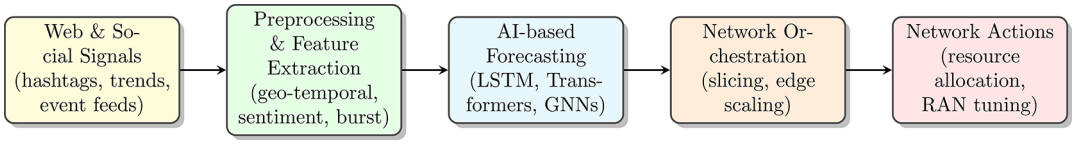

Figure 1. End-to-end pipeline for event-aware AI-driven demand forecasting and proactive network management.

et al. 2024) investigated how social-driven collective movements (e.g., protests) propagate spatial load shifts across the RAN, reinforcing the operational relevance of event anticipation.

AI-Driven forecasting and network orchestration

Machine learning for network prediction has evolved from conventional statistical models to fully adaptive, data-driven architectures. Early forecasting approaches in mobile networks relied on autoregressive or regression-based models such as ARIMA, Ridge, or ensemble trees, which remain competitive for short-term prediction but lack robustness to nonstationary traffic conditions. The introduction of recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) architectures (Hochreiter and Schmidhuber 1997) marked a decisive step forward, enabling the modeling of long-range temporal dependencies and complex seasonality patterns that characterize mobile traffic across heterogeneous services. These deep architectures, when combined with contextual or mobility features, significantly improve the ability to anticipate bursty events and local congestion, as confirmed by multiple studies and by our previous work on multi-layer orchestration (Pietri et al. 2025). More recently, Transformer-based architectures have become increasingly relevant in time-series forecasting. PatchTST, for example, represents time series as sequences of temporal patches and applies Transformer encoders to capture longer-range dependencies more efficiently than point-wise attention mechanisms (Nie et al., 2023). In parallel, graph-based spatio-temporal models such as STGCN and STAEformer explicitly model spatial dependencies among traffic sensors or network entities, making them important baselines when explicit spatial topology is available (Yu et al., 2018; Liu et al., 2023).

A key research evolution concerns the transition from static to online and adaptive forecasting. Traditional offline-trained models suffer from concept drift—the gradual change of data distributions due to evolving user behavior, application popularity, or seasonal patterns. To overcome this limitation, Mehri et al. (Mehri, Chen, and Mehrpouyan 2024) proposed an online learning framework for cellular traffic prediction that updates model parameters continuously in real time, providing higher responsiveness to distributional shifts. Similarly, Ma et al. (Ma et al. 2020) presented a comprehensive survey on proactive 5G optimization, emphasizing closed-loop learning and the integration of machine intelligence within network management functions. Musumeci et al. (Musumeci, Rottondi, and Nag 2019) also reviewed the broad spectrum of machine learning techniques applicable to 5G systems, including supervised, reinforcement, and deep learning approaches for radio resource management, mobility prediction, and slice orchestration.

At the management and orchestration layer, AI-assisted control constitutes a fundamental paradigm shift from reactive optimization toward predictive and autonomous decision-making. Márquez et al. (Márquez et al. 2018, 2019) empirically demonstrated the efficiency gains of dynamic network slicing, showing how adaptive resource sharing across services can increase utilization without violating isolation constraints. Our previous work (Pietri et al. 2025) extended this principle to multi-layer edge orchestration, leveraging forecasting models to anticipate service load variations and guide slice reallocation between edge and core nodes. Complementary efforts (Abu-gellban 2020) investigated the use of streaming analytics and drift-adaptive models for real-time traffic detection, while Mehmood et al. (Mehmood et al. 2024) introduced a novel edge architecture capable of detecting and mitigating concept drift directly at the network periphery, enabling continuous model retraining and localized self-adaptation in smart environments. Cost-aware capacity forecasting has also been addressed directly in network slicing. DeepCog (Bega et al. 2019, 2020) combines traffic forecasting with resource-provisioning decisions and explicitly considers the asymmetric operational consequences of under- and over-provisioning. This provides a direct precedent for direction-aware forecasting objectives. The present study complements this line of work by focusing on exogenous event context and its interaction with short-horizon demand forecasting and overload mitigation.

In parallel, industrial initiatives and standardization bodies have progressively embraced AI-based orchestration as part of the next-generation management vision. The ETSI Zero-touch Network and Service Management (ZSM) framework (Zero-Touch Network and Service Management (ZSM); Requirements Based on Documented Scenarios 2019; "ETSI ZSM: Zero-Touch Network & Service Management" 2025) defines a closed-loop automation paradigm that integrates prediction, intent-based policies, and self-optimization. Similarly, the O-RAN Alliance ("O-RAN ALLIANCE Vertical Industry White Paper" 2023) promotes the embedding of AI/ML functions within near-real-time RAN controllers to support proactive optimization and cross-domain coordination. These developments pave the way toward fully autonomous, self-optimizing 6G infrastructures, where forecasting and orchestration are tightly coupled through continuous feedback loops. Recent telecom-industry analyses similarly identify 5G connectivity, AI/ML, IoT, and edge computing as converging drivers of network automation and service innovation, reinforcing the relevance of predictive and context-aware network management (Ashok and Hallur, 2025).

Despite these advances, several open challenges remain. Effective deployment of AI-driven orchestration requires reliable data fusion across domains (social, network, mobility), explainability of learned models to ensure operational

trust, and robust cross-layer coordination among the RAN, transport, and edge computing tiers. The integration of forecasting, drift detection, and zero-touch orchestration thus represents a converging research frontier toward sustainable and proactive management of large-scale intelligent networks.

Synthesis and research gaps

Despite substantial progress in social sensing, forecasting, and AI-assisted orchestration, the fusion layer connecting event-awareness to automated network control remains immature and fragmented. Most existing studies operate in isolation, focusing either on event detection from social media, demand forecasting from historical KPIs, or resource management at the edge, with limited integration across these components. As a result, a unified pipeline capable of translating social signals into actionable orchestration policies is still largely missing.

Four main research gaps can be identified. First, few works explicitly quantify the lead time between web or social indicators and actual traffic surges, which is essential to evaluate how early prediction can enable proactive scaling in realistic operational settings. Second, the propagation of uncertainty from forecasting models to orchestration decisions is seldom analyzed, even though it critically affects the stability of closed-loop controllers and the reliability of slice reallocation. Third, cost-aware and asymmetric forecasting objectives have been investigated for resource provisioning, notably in DeepCog (Bega et al. 2019, 2020) and in our earlier experiments (Pietri et al. 2025). However, their integration with exogenous event context and reliability-aware event inputs remains comparatively underexplored. Fourth, the evaluation of modern Transformer-based and graph-based spatio-temporal forecasting models remains only partially addressed in mobile-network demand forecasting. While such models are increasingly relevant in general traffic forecasting, their application to mobile network datasets requires careful consideration of data availability, temporal depth, and explicit spatial topology.

In particular, graph-based architectures require an explicit representation of spatial or operational adjacency, such as radio topology, handover relations, or network-level connectivity. When such information is not exposed by the dataset, constructing a graph requires additional modelling assumptions that may affect comparability across forecasting methods. In addition, adaptive learning methods—such as online retraining and concept drift detection (Mehmood et al. 2024; Mehri, Chen, and Mehrpouyan 2024)—are still in early experimental stages when applied to mobile edge environments, and their interoperability with management frameworks such as ETSI ZSM and O-RAN (Zero-Touch Network and Service Management (ZSM); Requirements Based on Documented Scenarios 2019; "O-RAN ALLIANCE Vertical Industry White Paper" 2023) remains an open challenge. Bridging these paradigms requires advances in data fusion, explainable learning, and cross-layer coordination between prediction, decision, and actuation loops.

Overall, the next frontier of research lies in event-aware AI-driven orchestration: an integrated framework that connects social sensing, adaptive forecasting, and zero-touch management through continuous feedback and uncertainty-aware optimization. This vision aligns with the ongoing transition toward self-sustaining, intent-based future mobile network, where predictive intelligence becomes an intrinsic part of network operation.

Methods: A case study on event-aware forecasting

To complement the analysis with an applied perspective, we present an extended case study inspired by our previous work (Pietri et al. 2025), illustrating how event-aware forecasting can support proactive orchestration and predictive capacity management in mobile edge networks. While our previous study explored traffic prediction with recurrent models and exogenous web signals, here we broaden the analysis to compare multiple forecasting strategies—from naïve persistence to classical machine learning to recurrent and PatchTST-style Transformer models—and assess their impact under capacity-constrained orchestration scenarios. Figure 2 summarizes the pipeline to forecast network demand in our case study.

Dataset and context

The experiment uses the NetMob'23 dataset (Martínez-Durive et al. 2023), obtained under the access and license conditions established for the original dataset. It provides anonymized spatio-temporal mobile-traffic traces collected by Orange France across twenty metropolitan areas during Spring 2019. The dataset includes both downlink and uplink

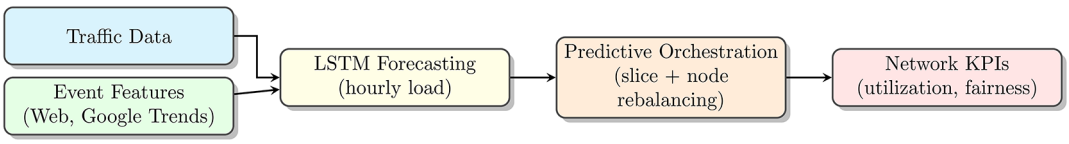

Figure 2. Pipeline of the case study: integrating event-aware forecasting into network orchestration.

volumes for 68 distinct mobile services (e.g., video streaming, social networks, instant messaging), sampled every 15 minutes and spatially aggregated on a 100 m × 100 m grid. These fine-grained measurements capture heterogeneous traffic patterns emerging from residential, commercial, and event-related activities, providing a realistic ground for evaluating adaptive forecasting and orchestration policies.

We focused on the city of Nantes, which exhibits diverse temporal dynamics driven by commuting flows, university areas, and large venues such as the La Beaujoire stadium and Parc des Expositions. During the observation period, several major events took place (e.g., Ligue 1 football matches, international fairs, and concerts), producing strong, localized demand surges. This setting allows us to investigate the sensitivity of forecasting and orchestration pipelines to the presence of abrupt, event-induced variations.

As in (Pietri et al. 2025), data were pre-aggregated hourly and normalized for privacy (i.e., divided by a fixed traffic amount). Although the resulting values are dimensionless, they remain consistent across time and geographic areas, ensuring comparability of relative variations and performance indicators – see Dataset in (Pietri and Mamei 2026).

Experimental design

The experimental workflow couples one-hour-ahead traffic forecasts with a simplified retrospective capacity benchmark. The purpose of this benchmark is to quantify how forecast-informed provisioning can reduce traffic exceeding a fixed reference capacity. It is not intended to reproduce a complete RAN, core-network, or operator resource-allocation process.

Each spatial node is evaluated through its service-cluster traffic series. We assume a simplified monotonic relationship between observed traffic volume and required capacity and abstract away detailed radio scheduling, handover dynamics, compute placement, and nonlinear resource-consumption effects. This abstraction allows forecasting methods to be compared under a common capacity-constrained criterion without introducing network-topology assumptions that are not available in NetMob'23 dataset.

For the operational evaluation, static reference capacities are defined from the 70th, 80th, and 90th percentiles of the observed demand in the evaluation block. Forecast-informed capacity can increase above the corresponding static reference when predicted demand requires it. Performance is then quantified through the relative reduction in overload with respect to the static-capacity benchmark, as formally defined below.

This operational benchmark should therefore be interpreted as a retrospective proxy for under-provisioning risk. It does not directly estimate operator capex, energy consumption, RAN scheduling efficiency, or end-to-end performance in an operational 5G/6G deployment. We assume that mobile traffic generated by base stations (BSs) is classified into multiple service categories. Network resources within each node are partitioned across several network slices, with each slice dedicated to a distinct service class. For example, a slice dedicated to video services handles all corresponding traffic flows. This model abstracts from specific network components, such as distinctions between radio access network (RAN) and 5G core functions (e.g., data plane versus user plane), given the lack of detailed configuration data and the broader conceptual focus of the study.

Although simplified, this architecture is sufficiently general to represent a wide range of modern network deployments. By focusing on traffic flows and resource distribution across slices and hierarchical levels, the model supports scalable simulations and theoretical evaluations—both of which are critical for network planning and performance optimization (see Figure 3).

Event-aware feature engineering. The forecasting models use endogenous traffic features and, when enabled, exogenous event-aware features. Endogenous inputs consist of lag1–lag24 together with hour-of-day and day-of-week encodings.

Event-aware inputs are derived primarily from scheduled-event records and include event_active, event_count, event_intensity, event_relevance, event_before, event_after, event_temporal_proximity, event-type indicators, and a public-holiday indicator. Event records are aligned with the hourly traffic index; events spanning multiple hours remain active over the corresponding intervals, while before/after and temporal-proximity variables represent the periods surrounding each event. Missing event-related values are encoded as zero.

The same event-aware feature construction is used across forecasting models. In the no-event ablation, only these exogenous variables are removed, while traffic lags and regular temporal encodings remain unchanged.

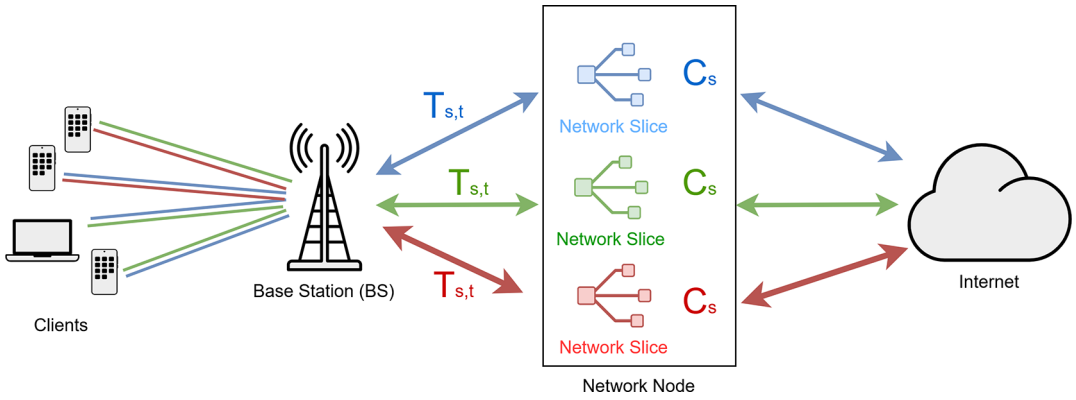

Figure 3. Model of the network. A network node serves the traffic from a given geographical area. Multiple network slides are instantiated to match requirements of different classes of services. Traffic for a given slice *s* at a given time *t* is represented by Ts,t. Slice capacity is represented by Cs. The network goal is to minimize the time and the magnitude of excess load (Ts,t > Cs) while minimizing the sum of Cs (as it entails capex for the telecom operator).

Google Trends and social-media streams are treated as optional contextual sources in the broader pipeline. For the historical 2019 case study, web/search information, where used, was obtained retrospectively; the experiments therefore do not establish that such signals would have been available with the same timing in a live forecasting system. The primary reproducible event-aware inputs are scheduled-event records and their derived temporal indicators.

For each model, the same event-aware feature set was used when exogenous variables were enabled. The ablation experiment reported in Figure 6 was obtained by removing all event-related variables while keeping the endogenous traffic lags and temporal encodings unchanged. This ensures that the comparison isolates the contribution of event-awareness from the contribution of regular temporal seasonality.

Models and training setup. The comparative evaluation involves five representative forecasting families, selected to cover a progressive spectrum of model complexity and learning capacity:

- Naïve: a persistence baseline assuming y(t) = y(t-1). This model has no trainable parameters and serves as a lower-bound reference.

- Ridge: a linear autoregressive model with Ridge regularization, using lagged traffic values, hour-of-day indicators, day-of-week indicators, and event-aware features when enabled. The regularization strength was selected on the training data.

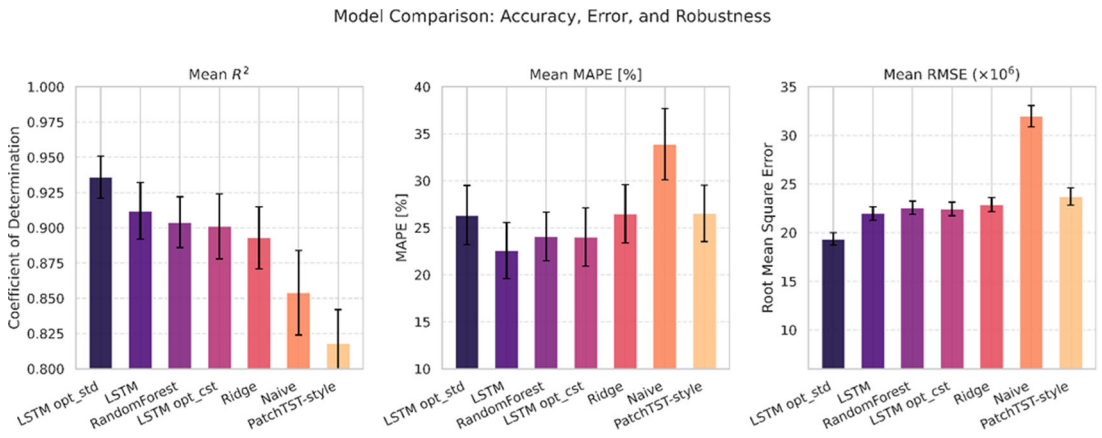

Figure 4. Forecasting comparison across the 39 spatial nodes × 5 service-cluster series. Bars report mean ± standard deviation for R^2, MAPE, and RMSE; higher R^2 and lower MAPE/RMSE indicate better symmetric forecasting performance.

Table 1. Forecasting accuracy and operational gains. MAE and RMSE are expressed in the dimensionless normalized traffic units used by the dataset; OR@70–90 are relative overload-reduction percentages with respect to the static percentile-capacity benchmark.

Model	MAE	RMSE	MAPE	OR [%]		
				@70	@80	@90
LSTM opt-cst	1.60×10^7	2.24×10^7	24.01%	71.3	69.1	61.2
LSTM opt-std	1.21×10^7	1.94×10^7	26.35%	65.5	61.8	59.2
LSTM	1.56×10^7	2.20×10^7	22.59%	59.6	50.1	35.5
Ridge	1.67×10^7	2.29×10^7	26.50%	61.7	51.3	34.6
RandomForest	1.62×10^7	2.26×10^7	24.08%	56.8	44.1	25.7
Naïve	2.29×10^7	3.20×10^7	33.89%	58.9	50.5	37.8
PatchTST-style	1.72×10^7	2.37×10^7	26.53%	53.8	42.5	26.8

- Random Forest: a non-parametric ensemble capturing nonlinear interactions between temporal, service-level, and event-aware features. The model was trained using the same input feature set as Ridge.

- LSTM: a recurrent neural network with memory cells (Hochreiter and Schmidhuber 1997), modeling temporal dependencies in service demand. The LSTM models used the same chronological training/test split as the other baselines and were trained with early stopping to prevent overfitting.

- PatchTST-style Transformer: a lightweight patch-based Transformer baseline inspired by recent time-series forecasting architectures. The model segments the input history into temporal patches and applies a Transformer encoder to capture longer-range temporal dependencies.

The split is chronological and the forecast horizon is one hour. The final 20% of each post-lag series is held out as the contiguous test block. Within the preceding 80%, the final 20% is reserved as validation for model selection and early stopping, corresponding to an effective 64%/16%/20% train/validation/test partition of the supervised samples. For a complete hourly series spanning 17 March-31 May 2019, the 24-hour lag warm-up yields approximately 18 March-4 May for training, 5–16 May for validation, and 17–31 May for testing. No test observation is used by Optuna. For the PatchTST-style baseline, we used the same node-level forecasting setup and the same event-aware input variables adopted for the other models. The model used a 24-hour input context, patch length 6, stride 3, d_model = 64, 4 attention heads, 2 Transformer encoder layers, and dropout = 0.10. Training was performed with AdamW, learning rate 10^−3, weight decay 10^−4, batch size 32, and early stopping on a validation split of the training set. The PatchTST-style model was evaluated using the same accuracy and orchestration metrics as the other forecasting models.

LSTM settings are selected globally on a deterministic subset of nodes and then reused across clusters.

- LSTM opt_std: one LSTM layer, MAE training loss and Adam optimization. Optuna varies the number of LSTM units (50–500), maximum epochs (50–500), batch size (2–128), and learning rate (10^-5–10^-1, logarithmic scale). Trials are ranked by validation RMSE; the full search comprises 1000 trials.

- LSTM opt_cst: the same recurrent search space, but trained with the asymmetric direction-aware provisioning loss defined below. The network is trained on per-node MinMax-scaled targets and model selection uses the validation partition; the final test block remains untouched until evaluation.

Asymmetric loss. Following (Bega et al. 2019, 2020), we employ a customized direction-aware, piecewise loss function. Let e = P −T denote the signed prediction error, where e < 0 corresponds to underprediction. The loss function is parameterized by $\alpha \in (0, 1)$, $\epsilon > 0$, $K \geq \alpha$, $\delta > 0$, $\eta \in (0, \alpha)$, and tail width $\Delta > 0$. We define the threshold parameter e_{end}, the slope m_{sh}, and the smooth transition function $\sigma(z)$ as:

$$e_{\text{end}} = \epsilon\alpha + \delta,$$
$$m_{\text{sh}} = \frac{\alpha - \eta}{e_{\text{end}}},$$
$$\sigma(z) = 0.5[1 - \cos(\pi \cdot \text{clip}(z, 0, 1))].$$

The asymmetric bounded loss L(e) is then defined piecewise as:

$$L(e) = \begin{cases} \min(K, \alpha - \epsilon e), & \text{if } e \le 0, \\ \max(\eta, \alpha - m_{\text{sh}} e), & \text{if } 0 < e \le e_{\text{end}}, \\ \eta + (K - \eta)\sigma\left(\dfrac{e - e_{\text{end}}}{\Delta}\right), & \text{if } e > e_{\text{end}}. \end{cases}$$

The construction is continuous and bounded in [0, K]. Underprediction receives a high penalty, moderate positive error enters a lower-cost shoulder, and large overprediction is penalized again through the smooth tail. Thus, the objective expresses the operational preference for avoiding service shortfall without making unbounded overprovisioning cost-free. The loss parameters are interpreted in normalized target space, and their defaults/search ranges are recorded by the released code.

The asymmetric objective is not expected to minimize RMSE. Its intended effect is to alter the residual distribution by making under-prediction more costly than moderate over-prediction. Accordingly, the analysis separates symmetric accuracy from direction-specific behavior. In addition to MAE, RMSE, MAPE, and R^2, the released evaluation code exports under-prediction rate and magnitude as well as corresponding over-prediction diagnostics on the original target scale. The main comparison below therefore interprets the asymmetric objective through both its conventional forecasting metrics and the operational overload benchmark, without assuming that it must improve every symmetric error measure.

Operational metrics. Forecast accuracy was measured through Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2). Operational performance is assessed through a retrospective overload-reduction benchmark. For each capacity level $q \in \{0.70, 0.80, 0.90\}$, the static capacity C_q is defined as the q-th percentile of the observed demand in the evaluation block. Given observed demand T_t and forecast P_t, forecast-informed capacity is defined as

$$\begin{aligned} C_{q,t}^{\text{dyn}} &= \max(C_q, P_t), \\ O_q^{\text{static}} &= \sum_t \max(0, T_t - C_q), \\ O_q^{\text{dyn}} &= \sum_t \max\left(0, T_t - C_{q,t}^{\text{dyn}}\right). \end{aligned}$$

respectively. Relative overload reduction is then

$$\text{OR}_q = 100 \frac{O_q^{\text{static}} - \text{O}_q^{\text{dyn}}}{O_q^{\text{static}}}.$$

We report OR@70, OR@80, and OR@90 as percentages relative to the corresponding static-capacity benchmark. This metric is a retrospective operational proxy and is not interpreted as a direct estimate of capex, energy consumption, or live-network performance.

Event-input robustness protocol. To isolate uncertainty in the exogenous event information, the observed traffic series, held-out targets, and trained model weights are kept fixed while only the event calendar is perturbed within the evaluation block. Because only three perturbable scheduled events occur in that block, percentage severities would collapse to identical integer counts; the experiment therefore uses absolute counts. The 27 factorial conditions combine 0, 1, or 2 cancelled event records; 0, 1, or 2 spurious event records inserted into originally event-free periods; and temporal displacements of 0, 1, or 2 hours applied to selected surviving events. Public holidays are kept deterministic. After every perturbation, all derived event features are regenerated before inference. Results are reported as changes relative to the clean C0/FP0/SHIFT0 condition, so the experiment measures event-input sensitivity rather than changes in realized demand.

Results and Insights

Experiments cover five service clusters over 39 spatial nodes, yielding 195 node-slice time series. Figure 4 summarizes descriptive results across those series and Table 1 reports all the results. The Naive persistence baseline produced RMSE

= 3.20 x 10^7 and MAPE = 33.9%; Ridge and Random Forest reduced RMSE to 2.29 x 10^7 and 2.26 x 10^7, respectively. Because the five service series at the same spatial node are correlated and the study preserves no repeated stochastic runs as independent experimental replicates, we report means and standard deviations descriptively and do not attach inferential significance claims to pairwise model differences.

The standard LSTM reached RMSE = 2.20 x 10^7 and MAPE = 22.6%. In this dataset and horizon it improves over the simple persistence baseline, while Ridge and Random Forest remain competitive.

The PatchTST-style Transformer baseline achieved MAE = 1.72 × 10^7, RMSE = 2.37 × 10^7, MAPE = 26.53%, and R^2 = 0.818. It improves over the naïve baseline in terms of forecasting error, but does not outperform the standard LSTM, Random Forest, or Ridge models in this setting. This suggests that, for the considered node-level hourly forecasting task, patch-based Transformer modelling does not automatically translate into better performance without further tuning or richer multivariate and spatial context.

Among the optimized recurrent models, LSTM opt_std provides the lowest RMSE, 1.94 × 10^7, and the highest R^2, 0.936. LSTM opt_cst yields MAE = 1.60×10^7, RMSE = 2.24×10^7, MAPE = 24.01%, and R^2 = 0.901. The two objectives therefore produce a metric-dependent trade-off: opt_cst has lower MAPE than opt_std (24.01% versus 26.35%), but higher RMSE and lower R^2, while the untuned LSTM retains the lowest MAPE among these three recurrent configurations (22.59%).

Operational Impact

We next evaluate how each forecast changes the retrospective overload benchmark. Table 1 and Figure 5 report OR@70, OR@80, and OR@90. These are relative percentage reductions in overload with respect to the corresponding static percentile-capacity reference, using the definition given in the Methods.

LSTM opt_cst achieves the largest overload-reduction values among the evaluated models: 71.3%, 69.1%, and 61.2% at the 70th, 80th, and 90th percentile capacity thresholds. LSTM opt_std reaches 65.5%, 61.8%, and 59.2%, respectively. The asymmetric objective therefore improves OR by 5.8, 7.3, and 2.0 percentage points relative to opt_std at the three thresholds, despite its higher RMSE. OR itself is a relative percentage with respect to the static overload benchmark and should not be interpreted as direct capex or energy savings.

Among the non-optimized models, Ridge reaches 61.7%, 51.3%, and 34.6%; the standard LSTM reaches 59.6%, 50.1%, and 35.5%; Random Forest reaches 56.8%, 44.1%, and 25.7%; Naive reaches 58.9%, 50.5%, and 37.8%; and PatchTST-

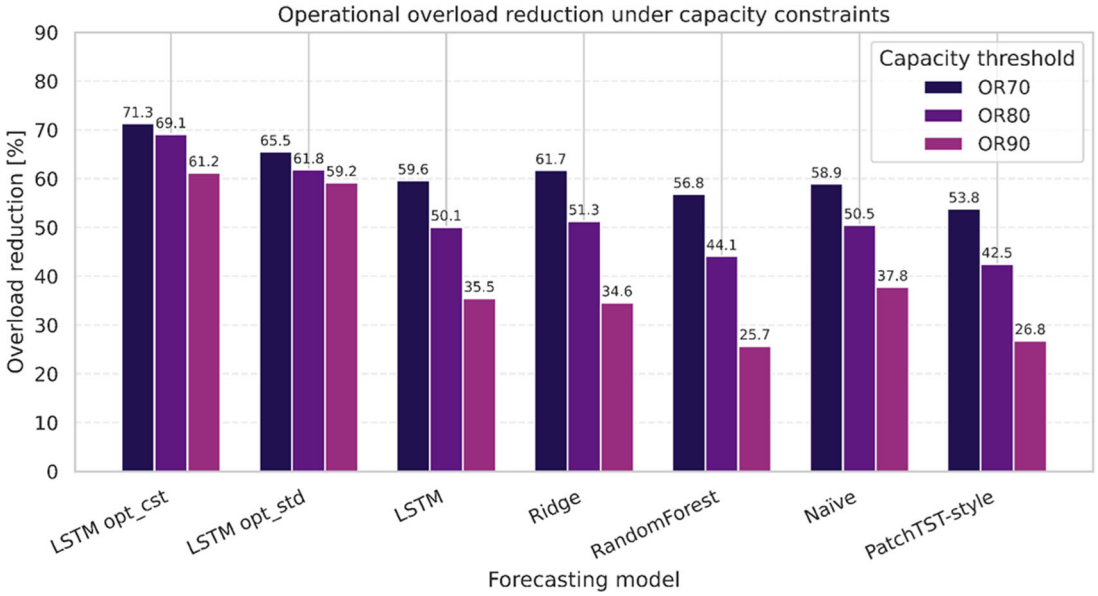

Figure 5. Relative overload reduction (OR) achieved by each forecasting model under static capacities set to the 70th, 80th, and 90th percentiles of observed evaluation demand. Values are percentages relative to the corresponding static overload reference.

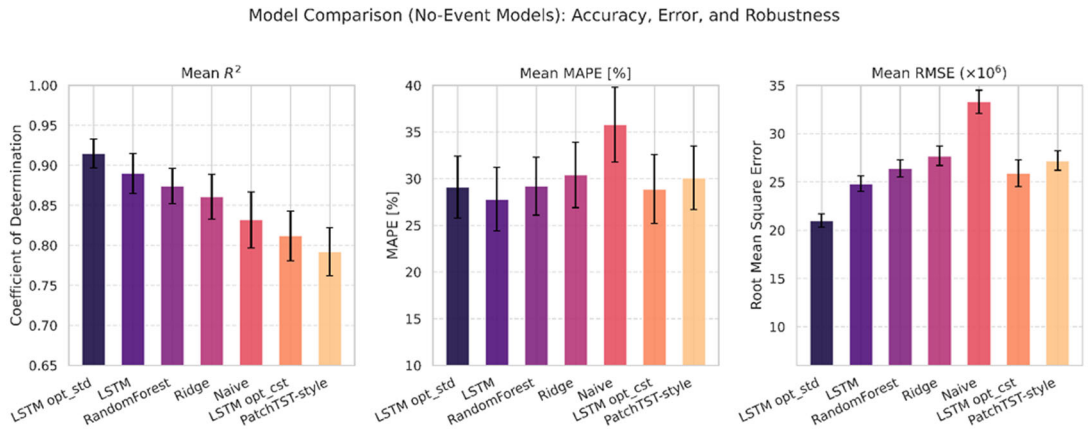

Figure 6. Comparison of forecasting models without event-related features. Bars report mean +/− standard deviation for R^2, MAPE, and RMSE across the node-slice series. For opt_cst, the no-event configuration yields R^2 = 0.812, MAPE = 28.9%, and RMSE = 2.59 x 10^7. Persistence is omitted because it does not consume event features and is therefore unchanged by the ablation.

style reaches 53.8%, 42.5%, and 26.8%. The ranking therefore depends on the capacity threshold, and simple models remain competitive in selected regimes.

Overall, the asymmetric-loss LSTM gives the strongest overload reduction in the reported operational benchmark, while opt_std remains preferable when RMSE/R2 are the primary objectives. The metric-dependent behavior is precisely the intended distinction: the asymmetric training objective is operationally motivated and should not be interpreted as a general-purpose mechanism for minimizing every conventional forecasting error.

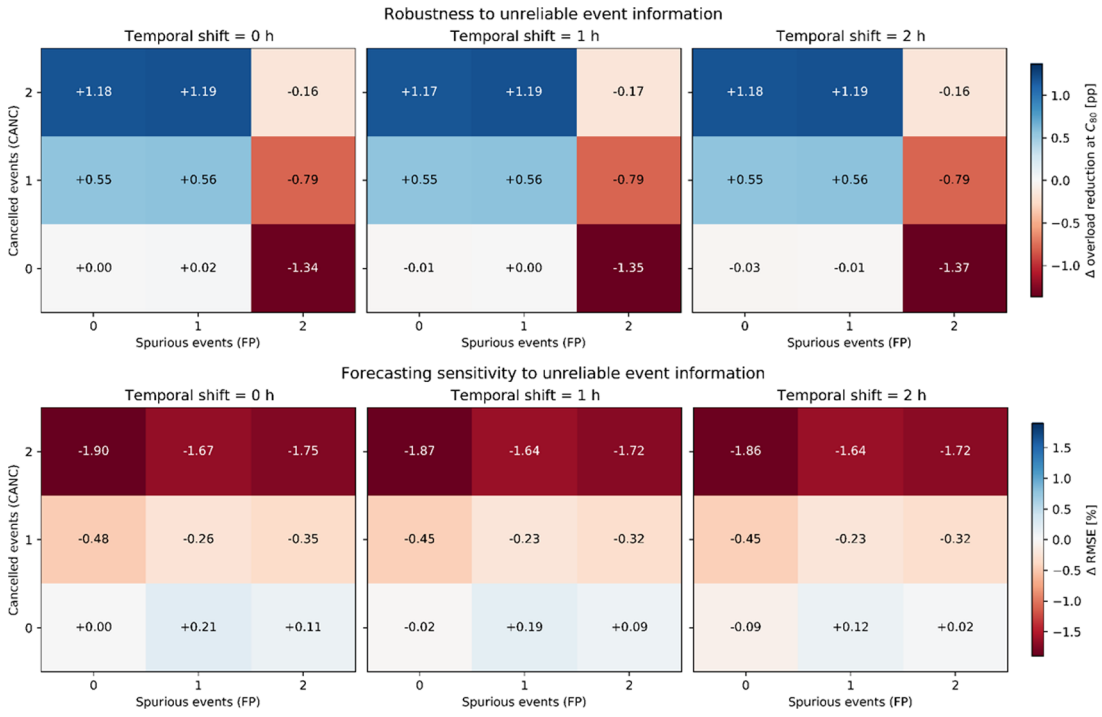

Figure 7. Sensitivity to event-input corruption relative to the clean robustness-run baseline. Top row: change in OR@80 in percentage points. Bottom row: relative change in RMSE. Columns show temporal shifts of 0, 1, and 2 hours; rows/columns within each heatmap vary the number of cancelled and spurious event records from 0 to 2. Observed traffic targets and trained model weights are fixed in every condition.

Optimized LSTM Results

The opt_std and opt_cst implementations use the Optuna protocol defined in the Methods. The released software includes the complete search spaces, objective logic, train/validation/test handling, direction-aware loss, and direction-specific evaluation metrics, allowing the optimization procedure to be independently reproduced.

LSTM opt_std reached RMSE = 1.94 x 10^7, MAPE = 26.35%, and R^2 = 0.936. The corrected full-search opt_cst run reached MAE = 1.60 x 10^7, RMSE = 2.24x10^7, MAPE = 24.01%, and R^2 = 0.901. Thus, opt_cst accepts higher RMSE and lower R^2 than opt_std while achieving higher overload reduction: 71.3%, 69.1%, and 61.2% compared with 65.5%, 61.8%, and 59.2%.

Impact of Event-Aware Features

To further quantify the contribution of contextual inputs, we repeated the forecasting workflow with all exogenous event/web variables removed while retaining lag1-lag24 and the hour/day temporal encodings. The train/validation/test partitions and model/evaluation definitions are otherwise unchanged.

For opt_cst, the event-aware and no-event configurations provide a direct controlled ablation. With event-aware inputs, opt_cst obtains RMSE = 2.24 $\times$ 10^7, MAPE = 24.01%, and R^2 = 0.901; without event-related features, the corresponding values are RMSE = 2.59 $\times$ 10^7, MAPE = 28.9%, and R^2 = 0.812. Using the no-event configuration as the denominator, event-aware inputs reduce RMSE by 13.4% and MAPE by 16.9%, while R^2 increases by 0.089.

The broader comparison in Figure 6 shows that the contribution of event-related inputs depends on the forecasting model and evaluation metric. The controlled ablation removes only the event/web variables while retaining traffic lags, temporal encodings, model family, and chronological evaluation logic.

Robustness to Event-Input Errors

As a further experiment, we evaluated the robustness of the proposed event-aware forecasting and orchestration framework against unreliable or imperfect event information. This analysis is motivated by the fact that, in real deployments, external event feeds are rarely fully accurate: events may be cancelled at short notice, public interest can be overestimated, and the temporal alignment between announced events and actual user presence is often noisy. Understanding how these imperfections affect both forecasting accuracy and operational performance is therefore essential – see Figure 7.

The robustness analysis uses 27 input-level perturbation conditions combining zero, one, or two cancelled scheduled-event records; zero, one, or two spurious event records inserted into originally event-free periods; and temporal shifts of zero, one, or two hours applied to selected surviving events. Observed traffic targets and trained model weights remain fixed in every condition, and all event-derived features are regenerated after each perturbation.

Across the 27 event-input perturbation conditions, RMSE changes range from −1.90% to +0.21%, while OR@80 changes range from −1.37 to +1.19 percentage points relative to the clean condition. Sensitivity is therefore modest and non-monotonic. False-positive event records produce the clearest adverse operational effect: with no cancellations and two spurious events, OR@80 decreases by approximately 1.34–1.37 percentage points across the evaluated temporal shifts. Temporal shifts alone have comparatively small effects. Event cancellations can occasionally improve symmetric error or overload reduction, indicating that individual scheduled events are not uniformly informative of realized mobile demand.

Discussion and Implications

The comparative results lead to several insights:

- **No single forecasting configuration dominates all metrics.** LSTM opt_std provides the lowest RMSE and highest R^2, the untuned LSTM provides the lowest MAPE among the three recurrent configurations, and LSTM opt_cst provides the largest overload reduction at all three evaluated capacity thresholds. The asymmetric objective should therefore be interpreted as an operational trade-off rather than as a general improvement in symmetric forecasting accuracy.

- **Sequential modelling remains important, but model choice matters.** Within the evaluated baselines, LSTM-based models provide the most favourable balance between forecasting accuracy and operational overload mitigation. The PatchTST-style Transformer baseline improves over the naïve predictor but does not outperform LSTM, Ridge, or Random Forest in this node-level hourly setting, suggesting that Transformer-based forecasting may require richer multivariate context, longer histories, or explicit spatial structure to fully express its advantages.

- **Spatially explicit deep models require richer topology.** Graph-based baselines such as STGCN and STAEformer are highly relevant for traffic forecasting, but their direct application would require an explicit graph describing radio topology, handover relations, or operational adjacency among network entities. Since the NetMob'23 data used in this case study provide spatially aggregated measurements but not the underlying operational network graph, introducing such models would require additional topology-inference assumptions.

- **Custom asymmetric loss functions connect prediction and provisioning objectives.** By penalizing underestimation more severely, opt_cst shifts the forecasting trade-off toward overload mitigation. In this case study it yields higher OR at all three capacity thresholds than opt_std, while accepting higher RMSE and lower R^2.

- **Event-aware context provides measurable predictive value in the controlled ablation.** For opt_cst, including event-related inputs reduces RMSE by 13.4% and MAPE by 16.9% relative to the no-event configuration, while R^2 increases from 0.812 to 0.901. The magnitude of this effect varies across forecasting models and metrics.

- **Robustness to imperfect event information should be evaluated explicitly.** Across the tested perturbation grid, changes in RMSE and OR@80 are modest but non-monotonic. False-positive event records produce the clearest adverse operational effect, while temporal shifts alone have comparatively small effects. Individual event cancellations can occasionally improve a metric when the corresponding event signal is weakly aligned with realized demand.

- **Simple baselines remain informative.** Ridge and Random Forest remain competitive at selected OR thresholds, showing that operational conclusions depend on both forecasting error and the chosen capacity regime.

Overall, the results show that event-aware contextual information can improve forecasting in this historical Nantes case study, while the asymmetric loss changes the balance between symmetric forecasting error and overload reduction. These operational gains are measured only within the retrospective percentile-capacity benchmark and should not be interpreted as direct estimates of energy consumption, cost savings, capex reduction, or live-network performance.

Limitations

The empirical study is based on a single city and a relatively short Spring 2019 observation period from a 4G-era dataset. The effective spatial sample consists of 39 nodes, with five correlated service-cluster series per node. Evaluation uses a single chronological split and descriptive aggregation rather than repeated-seed or spatially blocked inferential testing. Scheduled-event information is used retrospectively, and the live availability and latency of web/search signals are not established. Finally, OR is a simplified retrospective capacity benchmark and does not directly model RAN scheduling, handovers, slice isolation, operator energy consumption, capex, or end-to-end performance in current 5G/6G deployments.

Conclusions

This work presented an extended study on event-aware forecasting for proactive orchestration in mobile edge networks, combining a focused review of existing AI techniques with a quantitative case study on real-world mobile traffic data. Starting from our previous work (Pietri et al. 2025), we broadened the analysis by benchmarking multiple forecasting families—from naïve persistence and linear regression to tree ensembles, recurrent models, and a PatchTST-style Transformer baseline—and by evaluating their operational impact under capacity-constrained orchestration scenarios. The empirical results should be interpreted within the limits of a single-city case study based on Spring 2019 mobile traffic traces. Broader validation across additional cities, longer time windows, and more recent network deployments is required before generalizing the findings to operational 5G/6G systems.

Results obtained from the NetMob'23 show that no single forecasting configuration dominates all metrics. LSTM opt_std achieves the lowest RMSE and highest R2, whereas opt_cst achieves the largest overload reduction at all three evaluated capacity thresholds. In the opt_cst ablation, event-aware inputs reduce RMSE by 13.4% and MAPE by 16.9% relative to

the no-event configuration, while R2 increases from 0.812 to 0.901. These findings support the value of scheduled-event context for one-hour-ahead forecasting in the considered Nantes dataset, without implying statistical significance or generalization beyond this setting.

From a broader perspective, the integration of event-awareness into forecasting loops represents a key step toward situationally adaptive orchestration. As 6G networks evolve toward pervasive intelligence, the synergy between AI-based forecasting, federated edge learning, and digital-twin representations will enable continuous, context-driven reconfiguration of network resources with minimal human intervention.

Future research will extend this framework along (i) replication across additional cities, seasons, longer observation periods, and current network deployments; (ii) inclusion of seasonal-naïve references such as t-24 and t-168 and SARIMA-class models, together with repeated-seed and spatially blocked statistical comparisons; (iii) deeper evaluation of Transformer-based and graph-based spatio-temporal models when richer multivariate data and explicit topology are available; (iv) prospective integration of live web/event feeds with explicit publication-delay, missing-event, and reliability models; and (v) explainable and uncertainty-aware orchestration under richer operational constraints.

Ultimately, this study reinforces the view that event-aware AI forecasting is a promising component for predictive, self-optimizing future mobile network infrastructures, linking environmental awareness to dynamic, data-driven network automation.

Software availability

Source code available from: https://github.com/mmamei/Web_Social_Events_for_Mobile_Network_Demand_-Forecasting

Archived software available from: https://doi.org/10.5281/zenodo.22978033-License:CC-BY 4.0

Ethics and consent

Ethical approval and consent were not required.

Data availability

Underlying data available in this link https://doi.org/10.5281/zenodo.22978033 (Pietri, M., & Mamei, M. (2026)). - License: CC-BY 4.0

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References

Abu-gellban H: **A survey of real-time social-based traffic detection.** *2020 IEEE International Conference on Intelligence and Security Informatics (ISI).* 2020; 1–6.
Publisher Full Text

Akiba T, Sano S, Yanase T, *et al.*: **Optuna: A Next-Generation Hyperparameter Optimization Framework.** *ACM International Conference on Knowledge Discovery & Data Mining.* New York, NY, USA: 2019.

Ashok P, Hallur G: **Transformational Trends Shaping the Future of the Telecom Industry.** *In Recent Advances in Electrical & Electronic Engineering.* 2025; **18**(10).
Publisher Full Text

Bega D, Gramaglia M, Fiore M, *et al.*: **DeepCog: Cognitive Network Management in Sliced 5G Networks with Deep Learning.** *IEEE INFOCOM.* 2019; **2019**: 280–288.
Publisher Full Text

Bega D, Gramaglia M, Fiore M, *et al.*: **DeepCog: Optimizing Resource Provisioning in Network Slicing with AI-Based Capacity Forecasting.** *IEEE Journal on Selected Areas in Communications.* 2020; **38**(2): 361–376.
Publisher Full Text

Enhancing Event Experience — Mobility Report (Case Studies): **Ericsson Mobility Report.** 2016.
Reference Source

Essien A, Petrounias I, Sampaio P, *et al.*: **A Deep-Learning Model for Urban Traffic Flow Prediction with Traffic Events Mined from Twitter.** *World Wide Web.* 2021; **24**(4): 1345–1368.
Publisher Full Text

ETSI ZSM: Zero-Touch Network & Service Management: **ETSI official page.** 2025.
Reference Source

Gu Y, Qian Z(S), Chen F: **From Twitter to Detector: Real-Time Traffic Incident Detection Using Social Media Data.** *Transportation Research Part C: Emerging Technologies.* 2016; **67**: 321–342.
Publisher Full Text

Hochreiter S, Schmidhuber J: **Long Short-Term Memory.** *Neural Comput.* 1997; **9**(8): 1735–1780.
Publisher Full Text

Hu X, Ma W, Chen C, *et al.*: **Event Detection in Online Social Network: Methodologies, State-of-Art, and Evolution.** *Comput Sci Rev.* 2022; **46**: 100500.
Publisher Full Text

Jun S-P, Yoo HS, Choi S: **Ten Years of Research Change Using Google Trends: From the Perspective of Big Data Utilizations and Applications.** *Technol. Forecast. Soc. Chang.* 2018; **130**: 69–87.
Publisher Full Text

Karimiziarani M, Foroumandi E, Moradkhani H: **A Tutorial on Event Detection Using Social Media Data Analysis: Applications, Challenges, and Open Problems.** *arXiv preprint arXiv:2207.03997.* 2022.

Liu H, Dong Z, Jiang R, *et al.*: **Spatio-Temporal Adaptive Embedding Makes Vanilla Transformer SOTA for Traffic Forecasting.** In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM 2023). 2023: 4126–4130.
Publisher Full Text

Ma B, Zhao Z, Liu Y, *et al.*: **A Survey of Online Data-Driven Proactive 5g Network Optimisation Using Machine Learning.** *IEEE Access.* 2020; **8**: 35601–35625.
Publisher Full Text

Márquez C, Gramaglia M, Fiore M, *et al.*: **How Should i Slice My Network? A Multi-Service Empirical Evaluation of Resource Sharing Efficiency.** *ACM MobiCom.* 2018.
Publisher Full Text

Márquez C, Gramaglia M, Fiore M, *et al.*: **Resource Sharing Efficiency in Network Slicing.** *IEEE Trans. Netw. Serv. Manag.* 2019; **16**(3): 909–923.
Publisher Full Text

Martínez-Durive OE, Mishra S, Ziemlicki C, *et al.*: **The NetMob23 dataset: A high-resolution multi-region service-level mobile data traffic cartography.** 2023.
Reference Source

Mavragani A, Ochoa G, Tsagarakis KP: **Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review.** *J. Med. Internet Res.* 2018; **20**(11): e270.
Publisher Full Text

Mehmood H, Khalid A, Kostakos P, *et al.*: **A Novel Edge Architecture and Solution for Detecting Concept Drift in Smart Environments.** *Futur. Gener. Comput. Syst.* 2024; **150**: 127–143.
Publisher Full Text

Mehri H, Chen H, Mehrpouyan H: **Cellular Traffic Prediction Using Online Prediction Algorithms.**2024.
Reference Source

Mondal M, Khayati M, Hông S, *et al.*: **A Survey of Multimodal Event Detection Based on Data Fusion.** *VLDB J.* 2025; **34**: 1–25.
Publisher Full Text

Mredula MS, Behera RK, *et al.*: **A Review on the Trends in Event Detection by Analyzing Social Media Data.** *Sensors.* 2022; **22**(12): 4531.
Publisher Full Text

Musumeci F, Rottondi C, Nag A, *et al.*: **An Overview on Application of Machine Learning Techniques in 5g Networks.** *Comput. Netw.* 2019.

Nie Y, Nguyen NH, Sinthong P, *et al.*: **A Time Series is Worth 64 Words: Long-Term Forecasting with Transformers.** In Proceedings of the 11th International Conference on Learning Representations (ICLR 2023). 2023.
Publisher Full Text

O-RAN ALLIANCE Vertical Industry White Paper: **O-RAN Alliance White Paper.** 2023.
Reference Source

Pietri M, Hadjidimitriou NS, Mamei M, *et al.*: **Traffic analysis and resource adaptation in large-scale 5G multi-layer edge networks.** *Pervasive Mob Comput.* 2025.
Publisher Full Text

Pietri M, Mamei M: Web and Social Event Signals for AI-Driven Mobile Network Demand Forecasting. Dataset. 2026.
Publisher Full Text

Shafiq MZ, Ji L, Liu AX, *et al.*: **A First Look at Cellular Network Performance During Crowded Events.** *ACM SIGMETRICS.* 2013; 17–28.
Publisher Full Text

Walther M, Kaisser M: **Geo-Spatial Event Detection in the Twitter Stream.** *Advances in Information Retrieval (ECIR 2013), LNCS 7814.* Springer; 2013; 356–67.
Publisher Full Text

Yang B, Guo W, Chen B, *et al.*: **Estimating Mobile Traffic Demand Using Twitter.** *IEEE Wireless Communications Letters.* 2016; **5**(4): 380–383.
Publisher Full Text

Yang X, Bekoulis I, Deligiannis N: **Traffic Event Detection as a Slot Filling Problem.** *Eng. Appl. Artif. Intell.* 2023; **123**: 106202.
Publisher Full Text

Yao W, Zhang H, Zhang B: **Next-Day Morning Traffic Prediction Using Social Media Data.** *Transp Res C Emerg Technol.* 2021; **124**: 102768.
Publisher Full Text

Yu B, Yin H, Zhu Z: **Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting.** In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018). 2018: 3634–3640.
Publisher Full Text

Zanella A, *et al.*: **Impact of Public Protests on Mobile Networks.** *IEEE INFOCOM 2024 – Posters.* 2024.

Zero-Touch Network and Service Management (ZSM); Requirements Based on Documented Scenarios: **ETSI GS ZSM 001 V1.1.1.** 2019.
Reference Source

Zhang Z, He Q, Gao J, *et al.*: **A Deep Learning Approach for Detecting Traffic Accidents from Social Media Data.** *Transp Res Part C Emerg Technol.* 2018; **86**. 1801.01528.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ? ?

Version 2

Reviewer Report 10 October 2026

https://doi.org/10.21956/openreseurope.27491.r82703

© **2026 P 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.

?

Ashok P
Symbiosis Institute of Digital and Telecom Management (SIDTM), Pune, Maharashtra, India

Major 1: Event features underspecified
Largely addressed.
A dedicated "Event-aware feature engineering" subsection now lists the features (event_active, count, intensity, relevance, before/after, proximity, type, holiday), the hourly alignment, zero-filling and the ablation design. Remaining gaps: the scheduled-event calendar's **source and construction** are still not described, nor are the normalization and node-level mapping. Google Trends is now openly demoted to "optional/retrospective". The title ("web and social event signals") and Figure 2 ("Web, Google Trends") still overstate what was tested.

Major 2: Single city, single period; 5G/6G overreach.
Partly addressed.
The claims are toned down, and the 4G-era, single-city scope is stated in the Limitations and Conclusions. No second city was added, which is a legitimate choice. However, the abstract and keywords still say "5G networks" and "future mobile network infrastructures".

Major 3: No modern baselines.
Partly addressed.
A PatchTST-style baseline was added with full hyperparameters, and the STGCN/STAEformer omission is justified (no topology in NetMob'23). That justification is reasonable. The Transformer is "style" and lightly tuned, and it is the worst model on RMSE among the non-naïve models, so a stronger-tuned or official implementation would be fairer.

Major 4: Δpp normalization circular/misleading.
Addressed.
The normalized-to-best representation was removed. OR is redefined as a relative percentage (percentile capacity), Table 1 and Figure 5 were rebuilt on one scale, and the formulas are explicit.

Major 5: Unvalidated orchestration assumptions.
Partly addressed.

The assumptions are now stated clearly, and the claims are limited to a "retrospective proxy". The requested sensitivity analysis (non-linear load, capacity uncertainty) was not done. This is acceptable given the reframing, but the orchestration remains an abstract percentile rule.

Major 6: No inferential statistics.
Not addressed (justified in part).
No tests were added. The authors argue that a single split with correlated series does not support them, and they removed significance language. That is a defensible position. However, a Wilcoxon/Diebold–Mariano test on **39 nodes** (not 195) is feasible and would not need repeated runs.

Minor 1–770% variance, 30% claim, Figure 5, split dates, Optuna, "slides", keywords.
Addressed.
All fixed. The "up to 30%" claim is replaced by 13.4% (RMSE) and 16.9% (MAPE). Split dates and the 64/16/20 protocol are given.

2. Concerns arising from the revision

1. **The headline result was corrected, which is positive but raises a trust issue.** Following Reviewer 2, the opt_cst RMSE changed from 7.49×10^6 to 2.24×10^7 (about 3× worse). The original error shows that the earlier pipeline was unreliable. The corrected results now show opt_cst is *worse* than opt_std on RMSE/R^2 and the untuned LSTM on MAPE. I recommend independent verification that the deposited code regenerates Table 1 exactly.
2. **Best-performing configurations are not released.** The response states that the specific best configurations from the full search are not distributed and can only be regenerated by rerunning 1000 trials. That is weak for reproducibility, so the selected hyperparameters and seeds should be deposited.
3. **Table 1 inconsistencies.** Naïve OR@90 (37.8%) beats the untuned LSTM (35.5%) and Ridge (34.6%), and Naïve RMSE is quoted as 3.20×10^7 in the text. Figure 4 appears to show it near 3.2×10^7 on a $\times10^6$ axis, so the axis units should be checked. Opt_cst's OR advantage is also not accompanied by any measurement of over-provisioning, so the asymmetric loss's benefit is one-sided.
4. **Event ablation is limited to opt_cst.** The ablation is reported in full only for one model. Figure 6 shows no-event models, but a side-by-side event vs. no-event table for all models is needed to support "event-aware context provides measurable value".
5. **Ablation vs. robustness tension.** Event-aware inputs improve RMSE by 13.4% for opt_cst, but only **three** perturbable events exist in the test window, and cancelling events sometimes improves results. The claim that event features drive the gain therefore rests on very few episodes. Per-event results should be shown.
6. **Leakage control.** Scheduled events are known in advance, which is acceptable. But the statement that web/search data were "obtained retrospectively" means any Trends use cannot be called a forecast-time-valid input. This is properly disclosed.
7. **Missing baselines.** Seasonal-naïve (t-24, t-168) and SARIMA were again not added, though cheap to implement. Without them, "LSTM-based models offer the most favourable balance" is not well supported given that Ridge is competitive.
8. **Minor.** The Conclusions say the study "confirms" an anticipatory advantage, which is too strong. The Ashok & Hallur citation is only loosely relevant. The "far-edge/orchestration" framing in the abstract and Figure 2 still implies more than a percentile-capacity rule delivers.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Wireless Communication, Optical Communication, Free Space Optics, Telecom Networks, AI/ML in Telecom

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.

Version 1

Reviewer Report 15 July 2026

https://doi.org/10.21956/openreseurope.25267.r77085

© 2026 Ntanos C et al. 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.

?

Christos Ntanos
Decision Support Systems Laboratory, National Technical University of Athens, Zografou, Attica, Greece
Afroditi Blika
Decision Support Systems Laboratory, National Technical University of Athens, Zografou, Athens, Greece

Summary
The paper pairs a short literature review with a case study on the NetMob'23 dataset, restricted to Nantes, with 68 services grouped into five clusters. It compares persistence, Ridge, Random Forest and LSTM forecasters, plus two Optuna-tuned LSTM variants, one with an asymmetric loss that penalizes underestimation, and turns forecast quality into an overload-reduction metric against a static allocation. An ablation removes the event features, and a 27-scenario study tests robustness to unreliable event information. The headline claims are that the asymmetric-loss LSTM is best on all metrics, that event features cut errors by up to 30%, and that the pipeline degrades gracefully.

Overall
The problem is relevant and well-motivated, and the code and data are welcomed. The paper has three serious problems, though. The event features at the center of the study are never specified, and the deposited code does not contain them. Several results sections use definitions and numbers that do not match the stated design or each other. And the abstract, discussion and conclusions state quantities that the reported results do not support. In addition, several methodological aspects are introduced only in the Results rather than the Methods, making the experimental workflow difficult to follow and limiting the reproducibility of the study. All of this is fixable in a revision, so I recommend **Approved with Reservations**, but the revision needed is

substantial.

Major issues

1. The deposited code and data do not match the experiments the paper describes. We downloaded the deposit on 10/07/2026: both versions of the Zenodo record (27/12/2025 and 10/03/2026) and the linked GitHub repository. All three contain essentially the same material. Perhaps the deposit contains a stale version of the code. For example:
(a) The model-comparison script is explicitly event-agnostic. Its own comments read "Event-Agnostic Forecasting Benchmark", "Nessun evento/Google Trends" and "Nessuna feature evento"; the features are 24 hourly lags plus hour and day-of-week dummies. I could not find any Google Trends data, event calendar or web signal in the deposit.
(b) There is no Optuna, no asymmetric loss and no opt_std/opt_cst anywhere, and Optuna is missing from requirements.txt. Nothing is tuned either: Ridge and Random Forest use fixed settings, although the paper says hyperparameters "were tuned by grid or Bayesian optimization" for every model.
(c) The prediction files carry one "predicted" column from an unidentified model, covering the whole 17/03-31/05/2019 period rather than a test window, so Table 1 cannot be reproduced or checked.
(d) Capacity in the code is a percentile of the traffic distribution, which matches neither of the paper's two definitions.
Either the deposit is incomplete, or the paper describes experiments other than those performed. The authors should deposit the full pipeline behind Table 1 and Figures 4-7, or revise the paper to describe what the deposit actually does.

2. The event features are never specified. (Also Reviewer 1, point 1.) The paper does not say which Google Trends queries were used, how the event calendar was built, how signals were mapped to nodes and hours, or how they were scaled. Two more things are needed: evidence that each feature was available at forecast time, since Trends is published with a lag and using same-hour values would leak the future, and the full feature list of the base configuration, without which the ablation means little. *In particular, the paper does not clearly distinguish between the baseline forecasting features (e.g., traffic history and temporal encodings) and the event-related features that are subsequently removed in the ablation study. As a result, it is difficult to understand exactly which inputs are added by the event-aware framework and, consequently, to interpret and reproduce the reported impact of event-awareness.*

3. The headline result is surprising, and the paper gives no way to check it. The asymmetric loss is designed to overestimate, so it should pay a price on the standard error metrics. Instead, the paper reports that it beats the standard-loss model on those same metrics, by a factor of 2.6 in RMSE and by 8.6 points of MAPE, and does not explain how. Two details add to the doubt: the tuned opt_std has a worse MAPE than the untuned LSTM and the paper never mentions a validation set, while the early stopping in the deposited code watches the training loss. The authors should provide the code for these two models, the exact train/validation/test protocol with dates, and an explanation of why a loss that overestimates improves symmetric error. Until then the headline result cannot be verified. Moreover, the manuscript does not provide sufficient methodological details regarding the custom asymmetric loss itself. Since this loss function constitutes one of the main methodological contributions of the paper, its mathematical formulation, intuition, and expected impact on both forecasting and operational metrics should be described more thoroughly.

4. The operational metric cannot be reproduced, and capacity has three different

definitions. (Also Reviewer 1, point 5.) The Methods define C as a fraction of average hourly demand, the Results as a share of the observed peak, and the code uses a percentile of the distribution. The static allocation baseline is never defined, the orchestrator is never described, and Δpp is called percentage points while every deposited implementation computes a relative percentage. Since Δpp carries the paper's operational message, all of this must be clearly defined and made consistent. *Beyond reproducibility, the operational workflow itself remains difficult to follow. While the paper conceptually describes forecast-driven orchestration, it does not clearly explain how traffic forecasts are translated into capacity allocation decisions, how slice capacities are updated, or how the overload reduction metric (Δpp) is computed from these decisions. A more explicit description of the orchestration procedure would considerably improve the clarity of the experimental framework.*

5. The robustness experiment does not test what it claims. I differ here from Reviewer 1, who called this the strongest part; as written it cannot be interpreted. The paper says traffic was inflated and demand peaks removed, which means the ground truth was edited, not any event input, and indeed no event input exists to corrupt. The scenario parameters are also never reported; in the deposited files, values are simply multiplied by fixed factors (0.6 or 1.25). And the results do not behave as described: the paper calls 24.4 pp the most adverse scenario, yet a milder cell in Figure 7 shows 18.6 pp, and the lowest far-edge value (17.9 pp) occurs at a 1-hour shift, not 2. The authors should state exactly what was modified in each scenario and report the factors used. *Moreover, the robustness analysis introduces a new experimental setup that is not previously described in the Methods section. Since this constitutes a distinct experiment, the perturbation strategy (event cancellations, false positives, temporal shifts, and the generation of the 27 scenarios) should be described in the methodology rather than being introduced together with the results.*

6. A network hierarchy appears mid-paper, undefined, with numbers that do not match Table 1. The Methods define a flat model, but Figures 4 and 6 mention "far-edge nodes", and the robustness section reports far-edge, near-edge and core layers, along with a "Scenario B"; none of these terms is defined. At the same 80% threshold, the robustness section reports 29.45 pp where Table 1 reports 79.3, and the paper does not explain the difference. "195 nodes" also needs correcting: the deposited data hold 39 spatial nodes × 5 service clusters.

7. Several headline numbers have no support in the reported results. (a) The "20-25 pp higher overload mitigation" from event features appears nowhere; the ablation reports accuracy only. (b) "Reduces forecasting errors by up to 30%" reverses the arithmetic: removing the features raises RMSE by 31%, which is a 24% reduction in the other direction. (c) "Δpp = 100 pp at all capacity thresholds" mistakes the normalised score, which is 100 by definition, for percentage points; the absolute values are 82.4/79.3/73.1. (d) The "threefold decrease in RMSE" names no comparator (2.9× against the untuned LSTM, 4.3× against naïve, 2.6× against opt_std). (e) The 9-10% penalty in Δpp* is arbitrary; measure the over-provisioning instead or drop the metric. I agree with Reviewer 1 that normalising to the proposed model is circular, and the absolute numbers undercut the story anyway: at every threshold at least one trivial or linear baseline beats the untuned LSTM (Ridge at 70% and 80%, naïve at 90%).

8. Baselines and horizon. (Also Reviewer 1, point 3, on modern models.) The persistence baseline $y(t-1)$ is the weakest possible reference for daily-cyclical series; seasonal-naïve baselines ($t-24$, $t-168$) and a SARIMA-class model, which the text itself calls competitive, are standard and cheap to add. Figure 1 advertises Transformers and GNNs that are never evaluated. The forecast horizon is never stated, although the introduction lists lead-time quantification as an open gap, and lead time is exactly where event signals should earn their keep.

9. Related work. The asymmetric capacity-forecasting loss has a direct, uncited precedent:

DeepCog (Bega et al., IEEE INFOCOM 2019, pp. 280-288, DOI: 10.1109/INFOCOM.2019.8737488; IEEE JSAC 38(2):361-376, 2020, DOI: 10.1109/JSAC.2019.2959245), from the same group as the Márquez et al. papers already cited. It should be cited and contrasted, since the text implies the idea comes from the authors' own prior work. Separately, the introduction backs the claim that social traces explain variance in mobile traffic demand with four road-traffic papers (Abu-gellban 2020; X. Yang et al. 2023; Essien et al. 2021; Yao et al. 2021); only B. Yang et al. 2016 and Shafiq et al. 2013 concern cellular demand. I also agree with Reviewer 1 that the "70% of variance" figure needs its conditions stated. *More generally, although the literature review is comprehensive, the progression from prior work to the research gap could be strengthened. The manuscript reviews event detection, traffic forecasting, and network orchestration largely as independent topics, but does not clearly explain why their integration remains an open challenge before introducing the stated research gap. A clearer synthesis of how these research directions relate to one another would improve the motivation for the proposed study.*

10. Scope and statistics. (Also Reviewer 1, points 2 and 6.) Everything comes from one city over 77 days in spring 2019, 4G-era traffic, while the conclusions speak of 5G/6G infrastructures. The 80/20 split leaves about 15 days of test data, so the event effect rests on a handful of episodes; the authors should name the events in the test window, show per-event results, maybe add a second NetMob'23 city, and limit the 5G/6G claims to what a 4G-era study can support. Also, the Methods describe only a single 80/20 temporal train/test split, with no third (validation) partition defined. Yet the term "validation" is used substantively in the Results/Optimization discussion, implying a validation partition that is never defined anywhere in Methods. This is an internal inconsistency a reader can identify directly from the manuscript text.
Significance tests are needed for the model comparisons (Reviewer 1 already suggested suitable ones), noting that the effective spatial sample is 39 nodes, not 195 independent series.

Minor issues

1. Page 4: the Figure 1 caption is pasted mid-sentence ("...while Figure 1: End-to-end pipeline ... managementGu et al.").
2. "Essien and Petrovic" should be Petrounias, per the reference list.
3. Figure 3 caption: "network slides" should read "slices" (also Reviewer 1), and "Let T,s,t denote the traffic slice s, at time t" should read "Let Ts,t denote the traffic of slice s at time t".
4. Table 1 mixes decimal commas (5,62 × 10^6) with decimal points (0.1773); state that MAE and RMSE are in dimensionless normalised units.
5. "The apparent 100% performance gain ... (Figure 4)" points at the wrong figure; the 100% is in Table 1 and Figure 5.
6. Figure 5 mixes scales: the four baseline models plot at Table 1's absolute values, while opt_std and opt_cst sit at their normalised values. Ordering and legend legibility as per Reviewer 1.
7. Figure 6's caption calls bar charts "histograms" (histograms are for continuous values), and both the Figure 4 and Figure 6 captions mention "far-edge nodes" before the term is introduced.
8. Figure 7, core layer: all cells sit at 17.9-18.4 with +-0.0; explain the missing variance.
9. The S0-S3 labels appear once and vanish, and the Methods announce four model families while Table 1 lists six.
10. "R^2 decreases by 3-5%": percent or percentage points?

11. References: Zhang et al. 2018 has an arXiv identifier where pages should be; Musumeci et al. 2019 and Mehri et al. 2024 are incomplete.
12. Data availability: "cluserting" and "ageggated" are typos, and NetMob'23 is licence-gated rather than "publicly available".
13. Keywords: I agree with Reviewer 1 that more specific terms would help discoverability.

Recommendation
The direction of the work is worthwhile, and most of what I raise above is fixable. A revised version would need the full pipeline deposited, event features included, together with the code and protocol behind the two tuned models and an explanation of their headline result; one consistent definition of capacity, static baseline and orchestration, with Table 1 regenerated to match; a robustness design that states clearly what is being perturbed; the layer results reconciled with Table 1; significance tests for the model comparisons; and claims in the abstract and conclusions that stay within what a single-city, 4G-era study can show. If the event-feature and asymmetric-loss code cannot be recovered, I would ask the authors to rewrite the claims around the experiments that were actually run, which would still make a useful paper. More generally, the manuscript would benefit from a clearer narrative structure. Although the individual sections are generally well written, the overall experimental workflow is difficult to follow because several key concepts and assumptions are introduced only after they become relevant. A more structured presentation would substantially improve readability.

References
1. Bega D, Gramaglia M, Fiore M, Banchs A, et al.: DeepCog: Optimizing Resource Provisioning in Network Slicing With AI-Based Capacity Forecasting. *IEEE Journal on Selected Areas in Communications*. 2020; **38** (2): 361-376 Publisher Full Text

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?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
No

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

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

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Applied artificial intelligence and machine learning, data analytics, and

decision support systems, with long experience in applied ICT research and published work on AI in next-generation networks. We can assess the forecasting methodology, the experimental design, the reproducibility of the code and data, and the consistency of the reported results; We are not a specialists in radio access network engineering, so we do not assess the telecom-infrastructure aspects in depth.

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

Author Response 27 Sep 2026

Marco MAMEI

Dear Reviewer, Thank you for the detailed and constructive review. We agree that the previous version did not sufficiently align the manuscript, the deposited artifact and several operational definitions. The revision therefore focuses on reproducibility and on narrowing claims to what the reported experiment directly supports.
In particular, we have expanded the public software pipeline, rewritten the feature/split/loss/metric methodology, corrected the node count and operational metric, replaced the unsupported robustness/hierarchy experiment with an input-level event-feed sensitivity analysis, and revised the abstract, discussion, and conclusions accordingly.
Summary of the main revisions **Area Revision** Software artifact Added event-aware/no-event benchmarks, standard and extended Optuna LSTM scripts, direction-aware custom-loss LSTM, standalone PatchTST-style code, scheduled-event feature generation, and optional Google Trends ingestion/alignment. Event features Specified the complete traffic-only feature set and the added event-aware variables; documented hourly alignment, missing-value handling, lead/lag windows, optional geographic relevance, and the retrospective status of Google Trends. Training protocol Specified one-hour forecast horizon, chronological train/validation/test split, approximate dates, no shuffling, Optuna search spaces, and the role of validation. Asymmetric loss Added the full direction-aware loss and directional diagnostics; reran the corrected full 1000-trial experiment. The previously reported 7.49 x 10^6 RMSE is replaced by 2.24 x 10^7.
The corrected result shows lower RMSE/R2 performance than opt_std but higher operational OR at all three thresholds. Operational metric Replaced inconsistent capacity definitions with the percentile-based definition implemented in code and renamed the metric as relative overload reduction (OR, %), not percentage points. Robustness analysis Removed the old hierarchy/demand-edit perturbation experiment and replaced it with a 27-condition input-level event-feed sensitivity test that keeps traffic targets and trained model weights fixed. Scope/statistics Corrected 195 “nodes” to 39 spatial nodes x 5 service clusters; results are now described as descriptive rather than inferentially significant; single-city/2019 limitations are explicit. Claims Replaced the erroneous opt_cst accuracy headline with corrected full-search results, updated OR values, and a full no-event opt_cst comparison. Unsupported 20-25 pp, 100 pp, threefold-RMSE, and direct energy/cost-saving claims remain removed. Point-by-point response Major issue 1(a) - Event-aware code and data were missing **Reviewer comment.**
The deposited benchmark was explicitly event-agnostic and did not include the event calendar,

Google Trends data, or web/event feature construction. **Response.** The revised artifact now contains a common event-aware/no-event interface and dedicated event-feature generation utilities. The main historical case study is explicitly based on scheduled-event proxies. Optional archived/online Google Trends adapters are also released, but the manuscript no longer treats retrospectively retrieved Trends values as evidence of real-time availability.
Changes in the revision.
The Methods now list the base traffic features (lag1-lag24, hour-of-day and day-of-week one-hot encodings) and the added contextual features (event activity/count, intensity, relevance, before/after/proximity indicators, event type, and holidays). The Software Availability section now describes the expanded artifact. Major issue 1(b) - Optuna and asymmetric-loss implementation were absent **Reviewer comment.**
The previous deposit contained neither Optuna nor the opt_std/opt_cst implementations, while the manuscript stated that models were tuned. **Response.** We agree that the artifact was incomplete. The revised release includes the standard Optuna LSTM, an extended Optuna implementation, and the direction-aware custom-loss implementation. We also corrected the manuscript so that Ridge and Random Forest are no longer described as Bayesian-tuned: the released benchmark uses fixed Ridge and Random Forest settings. Optuna is used for the optimized LSTM variants. **Changes in the revision.**
The Methods now specify the LSTM search space: 50-500 units, 50-500 maximum epochs, batch size 2-128, learning rate 10^-5 to 10^-1, and 1000 trials for the full search. The complete optimization procedure is released. The specific best-performing configurations from the original full search are not distributed; they can be regenerated by rerunning the complete search. Reduced-trial runs are intended only for functional verification. Major issue 1(c) - Legacy prediction files did not identify the model or test block **Reviewer comment.** *The deposited prediction files contained a single unidentified predicted column over the full period and could not be used to reproduce Table 1.* **Response.**
The revised workflow no longer relies on those legacy prediction files as the authoritative reproduction path. Each model script generates its own per-node metrics, predictions, run configuration, and summary from the chronological split. **Changes in the revision.** The manuscript now specifies the one-hour horizon and the train/validation/test protocol. For a complete series spanning 17 March-31 May 2019, after the 24-hour lag warm-up the approximate blocks are 18 March-4 May (train), 5-16 May (validation), and 17-31 May (test). Major issue 1(d) / Major issue 4 - Capacity and overload metric were inconsistent **Reviewer comment.** *The manuscript used average demand, peak demand, and percentile capacity definitions, while the code computed a percentile-based relative percentage.* **Response.** We have made the code definition authoritative. For q in {0.70,0.80,0.90}, static capacity is the q-th percentile of observed demand in the evaluation block. Forecast-informed capacity is max(C_q, prediction).
Overload reduction is 100 x (O_static - O_dynamic) / O_static. We now call this OR@70/80/90 and report it as a relative percentage, not percentage points. **Changes in the revision.** The Experimental Design, Operational Metrics, Table 1, Figure 5, captions, Results, Discussion, and Conclusions were all revised to use the same definition. Figure 5 was regenerated with one consistent scale. Claims of direct capex or energy savings were removed. Major issue 2 - Event features and forecast-time availability were underspecified **Reviewer comment.** *The feature list, calendar construction, Google Trends queries, node/hour mapping, scaling, and forecast-time availability were not specified; contemporaneous Trends values could cause leakage.* **Response.**

The revision now distinguishes the traffic-only baseline from the event-aware extension. Scheduled events are aligned to the hourly index, missing values are zero, intensity/relevance variables are bounded, and event windows are expanded by configurable lead/lag intervals. The released configuration exposes 6 hours before and 4 hours after as defaults and supports optional spatial relevance when coordinates are available. Google Trends is explicitly optional. **Changes in the revision.**
Most importantly, the manuscript now states that the historical 2019 Google Trends series were retrieved retrospectively. Therefore same-hour Trends values are not claimed to have been available in real time. The numerical event-ablation result is framed primarily as a scheduled-event-context experiment; the Trends code is provided for exploratory/prospective extensions with an explicit lag/publication-delay model. Major issue 3 - Surprising opt_cst result and insufficient loss/training detail **Reviewer comment.** *The asymmetric loss should favor overestimation, yet opt_cst also strongly improves symmetric metrics; the loss and validation protocol were not sufficiently documented.*

Response. We sincerely thank the reviewer for identifying this issue. The unusually low RMSE reported for the asymmetric-loss LSTM in the original submission was indeed incorrect. During our re-examination, we identified an inconsistency in the post-processing and aggregation of that result. We therefore reran the complete experiment with the consolidated implementation and the common chronological train/validation/test and metric-computation pipeline.
The corrected 1000-trial run gives LSTM opt_cst MAE = 1.60 x 10^7, RMSE = 2.24 x 10^7, MAPE = 24.01%, and R2 = 0.901. By comparison, opt_std reaches RMSE = 1.94 x 10^7, MAPE = 26.35%, and R2 = 0.936. The corrected result therefore no longer shows an across-the-board symmetric-error advantage: opt_cst has higher RMSE and lower R2 than opt_std, although its MAPE is lower. We apologize for the original reporting error and have replaced the affected value throughout the manuscript. Importantly, the corrected results provide the expected metric trade-off. opt_std is the stronger configuration when RMSE/R2 are the primary objectives, whereas opt_cst improves the operational overload benchmark: OR@70/80/90 = 71.3%, 69.1%, and 61.2%, compared with 65.5%, 61.8%, and 59.2% for opt_std. We therefore interpret the asymmetric loss as an operationally motivated objective rather than as a mechanism for minimizing every conventional forecasting error. This interpretation is also consistent with our earlier work, where the customized loss was motivated by the operational asymmetry between missing traffic peaks and moderate over-provisioning. The revised manuscript now makes that distinction explicit and reports the symmetric and operational metrics side by side instead of inferring an operational advantage from an anomalously low RMSE.

Changes in the revision. We revised the Abstract, Methods, Table 1, Figures 4-6, optimized-LSTM results, Discussion, and Conclusions. The corrected full-search opt_cst values are now reported directly, and the former 7.49 x 10^6 value is explicitly identified as erroneous. Figure 4 includes the corrected opt_cst symmetric metrics. Figure 6 now reports the full no-event comparison; for opt_cst, the no-event result is RMSE = 2.59 x 10^7, MAPE = 28.9%, and R2 = 0.812. Relative to that no-event configuration, event-aware opt_cst reduces RMSE by 13.4% and MAPE by 16.9%, with an R2 increase of 0.089.
Major issue 5 - Robustness experiment did not perturb the claimed event inputs **Reviewer**

comment. *The reported 27-scenario experiment edited demand-like quantities rather than clearly corrupting event-feature inputs, and the text/figure did not support the claimed monotonic degradation.* **Response.** The previous implementation did not isolate event-information uncertainty correctly, because some perturbations were applied to demand-like quantities rather than exclusively to the exogenous event inputs. We therefore redesigned and reran the robustness experiment. In the revised analysis, the observed mobile-traffic series and test targets are never modified. Perturbations are applied exclusively to the event-information layer before the corresponding event features are generated. Because only three perturbable scheduled events fall inside the held-out evaluation block, percentage severities would collapse to the same integer perturbations. We therefore use absolute counts. The 27 factorial conditions combine 0, 1, or 2 cancelled event records; 0, 1, or 2 false-positive event records inserted into originally event-free periods; and temporal displacements of 0, 1, or 2 hours applied to selected surviving events. Public holidays are kept deterministic. After each perturbation, all event-derived features are regenerated from the modified event calendar. The forecasting model is trained on unperturbed data and then kept fixed. The same held-out traffic targets are used in every condition; only the exogenous event inputs change. The C0/FP0/SHIFT0 condition reproduces the clean evaluation exactly, providing a direct within-run reference for the sensitivity deltas. The corrected results do not show the monotonic degradation claimed in the previous version. Across all 27 conditions, delta RMSE ranges from -1.90% to +0.21%, and delta OR@80 ranges from -1.37 to +1.19 percentage points relative to the clean robustness-run baseline. Two spurious events with no cancellations produce the clearest adverse OR@80 change (approximately -1.34 to -1.37 percentage points across shift levels), whereas temporal shifts alone have only small effects. Event cancellations can occasionally improve the metrics, showing that scheduled-event presence is not uniformly beneficial when an event is weakly aligned with realized traffic. Changes in the revision. We rewrote the robustness methodology to define the perturbation process before the Results, regenerated all 27 conditions by modifying event records rather than traffic observations, and added a new Figure 7 reporting delta OR@80 and delta RMSE relative to the clean condition. The manuscript now describes the result as modest, non-monotonic sensitivity to event-feed errors and no longer claims graceful or monotonic degradation. Major issue 6 - Undefined network hierarchy and incorrect node count **Reviewer comment.** *The manuscript introduced far-edge/near-edge/core layers without defining them, and "195 nodes" actually represented 39 spatial nodes x 5 service clusters.* **Response.** We agree. The revised case study does not infer a near-edge/core hierarchy from data that do not expose such topology. We explicitly define 39 spatial aggregation nodes and five service clusters, yielding 195 node-slice time series. Figure 3 is now described as a conceptual abstraction only.

Changes in the revision. The unsupported far-edge/near-edge/core hierarchy results and the old hierarchy-based Figure 7 were removed. The new Figure 7 is a different experiment: it reports only event-input sensitivity deltas at the node-slice forecasting level and does not introduce a near-edge/core hierarchy. Figure 4/6 captions and the Results/Limitations sections distinguish 39 spatial nodes from 195 node-slice reporting series.

Major issue 7 - Unsupported headline quantities **Reviewer comment.** *Several headline numbers were not supported or used inconsistent denominators/scales.*

Response. In addition to correcting the denominator/scaling issues identified by the reviewer, we reran the affected asymmetric-loss and no-event experiments. The erroneous 7.49 x 10^6 opt_cst RMSE is replaced by the corrected 2.24 x 10^7 value. The matched no-event opt_cst RMSE is 2.59 x 10^7, so event-aware inputs reduce RMSE by 13.4% when the no-event result is used as the denominator; MAPE decreases from 28.9% to 24.01%, a 16.9% reduction. We also removed the unsupported 20-25 percentage-point event claim, all "100 pp" language, the unspecified threefold-RMSE statement, and direct energy/cost-saving claims. Changes in the revision. The revised paper separates three questions that were previously conflated: symmetric forecasting accuracy, the operational effect of asymmetric training, and the value of event-aware contextual inputs. Table 1 and Figures 4-6 now use the corrected full-search values and explicit denominators. Figure 7 separately quantifies sensitivity to event-feed errors. Claims are made only on the metric actually reported.
Major issue 8 - Baselines, Transformer/GNN scope, and forecast horizon **Reviewer comment.** *The one-step horizon was not stated; persistence is a weak seasonal baseline; SARIMA/seasonal-naive references were absent; Figure 1 suggested model families not all evaluated.* **Response.** We now state the one-hour forecast horizon explicitly and clarify that PatchTST-style is the evaluated Transformer baseline. Graph models are not evaluated because the dataset does not expose an operational adjacency graph; this is stated explicitly rather than leaving the implication ambiguous. We retained persistence for continuity with the original benchmark but agree that t-24, t-168, and SARIMA-class references are useful additions. **Changes in the revision.** Because we do not have preserved, independently verified results for those additional baselines, we did not insert post-hoc numbers into Table 1. Instead, the revised Discussion and Limitations explicitly identify seasonal-naive and SARIMA baselines as a priority for future replication and avoid presenting the current benchmark as exhaustive. Major issue 9 - Related work and DeepCog precedent **Reviewer comment.** *The asymmetric capacity-forecasting objective has a direct precedent in DeepCog; several social-signal citations concerned road traffic rather than cellular demand; the 70% variance claim lacked conditions.* **Response.** We agree. We added both DeepCog references (INFOCOM 2019 and JSAC 2020) and contrast their cost-aware capacity-forecasting objective with our event-context focus. We also rewrote the social-sensing discussion so that Yang et al. 2016 and Shafiq et al. 2013 are identified as direct cellular-network evidence, while road-traffic/mobility papers are described as adjacent-domain evidence. **Changes in the revision.** The Yang et al. result is now qualified: the Greater London study used over 0.6 million geotagged tweets and over one million mobile-demand records and reported approximately 71-79% explained variance for long-term uplink/downlink demand. “Essien and Petrovic” was corrected to “Essien and Petrounias.” The inaccurate Musumeci citation/sentence was removed. Major issue 10 - Scope and statistics **Reviewer comment.** *The study covers one city and roughly 77 days of 2019 traffic, yet the conclusions generalized to 5G/6G; the effective spatial sample is 39 nodes, and significance tests were requested.* **Response.** We agree with the scope concern and substantially narrowed the claims. The revised manuscript explicitly describes a single-city, historical, 4G-era case study and no longer claims direct operational 5G/6G performance. We also correct the effective spatial sample to 39 nodes and explain that the five service-cluster series per node are correlated. **Changes in the revision.** We did not add post-hoc significance tests because the original experiment preserves a single chronological split and not an independent set of repeated stochastic runs suitable for the suggested inferential analysis.

Instead, all language implying statistical significance has been removed; figures are described as descriptive mean +/- standard deviation. Repeated-seed and spatially blocked inference is listed as future work. We believe this is preferable to reporting p-values on 195 correlated series as if they were independent. Minor issues **Figure/caption placement:** The Figure 1 caption is kept as a separate caption paragraph and the surrounding text has been checked for sentence continuity. **Author name:** “Essien and Petrovic” was corrected to “Essien and Petrounias.” **Figure 3 notation:** The text now uses T_{s,t} for traffic of slice s at time t; the caption states that Figure 3 is conceptual and does not reconstruct physical topology. **Table 1 notation and units:** Table 1 was rebuilt with decimal points, explicit normalized-traffic units for MAE/RMSE, and OR@70/80/90 (%) instead of Delta-pp. **Figure 5 scale:** Figure 5 was regenerated on one consistent overload-reduction percentage scale; the former percentage-point terminology was removed.

Figure 6 terminology: “Histogram” was corrected to “bar chart”; the figure now reports the full no-event comparison, with persistence omitted because it does not consume event features and is unchanged by the ablation. Figure 7/core variance: The old far-edge/near-edge/core robustness figure was removed. The new Figure 7 reports only input-level robustness deltas (OR@80 and RMSE) for the corrected 27-condition event-feed perturbation experiment, so the unexplained core-layer variance issue no longer applies. Model counts/labels: The Methods now state five base forecasting families plus two optimized LSTM variants; transient S0-S3 shorthand was removed from the main Results text. **R2 change:** Changes in R2 are described as absolute changes in R2, not percentages. **References:** DeepCog references were added; the inaccurate Musumeci citation was removed; the remaining arXiv references are identified as preprints where applicable. **Data availability:** NetMob'23 is no longer called unconditionally “publicly available”; the revision states that it must be obtained under the original dataset access/license conditions. **Keywords:** Keywords were made more specific to mobile traffic forecasting, event-aware forecasting, LSTM, network slicing, and resource provisioning.
Closing note. We particularly appreciate the reviewer's observation concerning the asymmetric-loss result. It exposed a reporting inconsistency and prompted a complete 1000-trial rerun under the consolidated pipeline. The corrected result (RMSE = 2.24 x 10^7 rather than 7.49 x 10^6) now shows the intended trade-off clearly: opt_std is stronger on RMSE/R2, whereas opt_cst yields higher overload reduction. The same audit also led us to replace the old robustness experiment with a direct event-input perturbation analysis and to report the event/no-event comparison with explicit denominators.
Sincerely,
Marcello Pietri and Marco Mamei

Competing Interests: No competing interests were disclosed.

Reviewer Report 20 May 2026

https://doi.org/10.21956/openreseurope.25267.r74497

© 2026 P 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.

?

Ashok P
Symbiosis Institute of Digital and Telecom Management (SIDTM), Pune, Maharashtra, India

Summary of the Paper
This paper presents a combined literature review and empirical case study on AI-driven, event-aware mobile network demand forecasting. The authors argue that web and social signals (e.g., Google Trends, event listings) serve as leading indicators of localized traffic surges during mass events. Using the NetMob'23 dataset restricted to the city of Nantes, they benchmark four forecasting families—Naïve, Ridge Regression, Random Forest, and LSTM—under a simulated network slicing orchestration loop. Their principal claims are that LSTM with a custom asymmetric underestimation-penalizing loss outperforms all baselines, that event-aware features reduce RMSE by up to 30%, and that the framework degrades gracefully under corrupted event inputs.

Overall Assessment and Contribution

The problem is genuinely important and practically motivated. Proactive, event-aware orchestration for 5G/6G is a timely research frontier, and the end-to-end framing—from social signal acquisition through forecasting to orchestration impact—is conceptually sound. However, in its current form the paper suffers from a fundamental identity crisis: it is simultaneously a mini-survey, a reproducibility extension of prior work (Pietri et al. 2025), and an original empirical study. None of these three roles is executed with the depth a top-tier venue would require. The empirical novelty over the authors' own prior publication is marginal, the feature engineering underpinning the key "event-awareness" claim is critically underspecified, and the experimental scope (single city, single time window, no modern sequence-to-sequence baselines) does not support the broad conclusions drawn about 5G/6G infrastructure.
Strengths

- **Practical relevance.** The problem of anticipating event-driven demand surges is operationally significant for network operators and is well-motivated in the introduction.
- **Asymmetric loss function.** The opt_cst variant with asymmetric underestimation penalty is a sensible and domain-motivated design choice, and its formulation and impact are at least partially explained.
- **Robustness analysis.** The 27-scenario perturbation experiment—covering event cancellations, false positives, and temporal shifts—is the most original and thorough component of the paper. The graceful degradation finding is both credible and useful.
- **Open data and code.** Code is available on GitHub and data is archived on Zenodo with a CC-BY 4.0 license. This is commendable and facilitates reproducibility.
- **Clear pipeline figure.** Figure 1 and Figure 2 succinctly communicate the conceptual architecture.

Major Issues / Concerns
1. Critical underspecification of the event features — the paper's central claim. The title promises an analysis of "web and social event signals," yet the Methods section never specifies what these features actually are. The reader is told that "event-related features" derived from

"web or scheduled events" are used, and Figure 2 labels them as "Web, Google Trends." No detail is given on: which Google Trends queries were selected, at what temporal or spatial resolution, how they were aligned with the 15-minute or hourly traffic windows, how they were normalized, or how the scheduled event calendar was constructed and sourced. This is the single most damaging omission in the paper. Without this information, the 30% RMSE reduction attributed to event-awareness is unverifiable and unreproducible, regardless of code availability. The authors must provide a dedicated feature engineering subsection with all details, including the full feature set fed to each model.

2. Single-city, single-period experiment severely limits generalizability. All empirical results derive exclusively from Nantes during Spring 2019 — a period predating widespread 5G deployment. The paper's conclusions are framed in terms of "5G/6G infrastructures," yet the dataset is a 4G-era LTE trace. The authors explicitly acknowledge the "limited visibility of the mobile network infrastructure within the dataset," but draw no methodological boundary around their conclusions. At minimum, the authors should run experiments on at least one additional city from the NetMob'23 dataset (which covers 20 metropolitan areas) and explicitly limit their 5G/6G claims to conceptual extensions.

3. Absence of modern time-series baselines. The model zoo stops at vanilla LSTM (Hochreiter & Schmidhuber, 1997). Transformer-based architectures (e.g., Informer, PatchTST, TimesNet, iTransformer) have become standard baselines in time-series forecasting benchmarks since 2021 and are absent. Similarly, state-of-the-art mobile traffic forecasting works such as STGCN or STAEformer, which explicitly model spatial correlations across cells, are not benchmarked. The absence of these comparators means the LSTM's claimed superiority is established against a field from the previous decade.

4. The Δpp metric normalization is circular and misleading. Reporting all operational gains normalized to LSTM_opt_cst = 100% obscures absolute performance and inflates the apparent superiority of the proposed method. The unnormalized absolute values in Table 1 already convey the message; the normalized columns add no information except to make the best model trivially attain a perfect score by construction. Worse, the Ridge model achieves higher raw Δpp at 70% capacity (61.7 pp vs. 59.6 pp for standard LSTM), which contradicts the narrative of LSTM superiority in orchestration terms. This inconsistency is noted but not adequately explained.

5. The orchestration simulation rests on unvalidated assumptions. The network model assumes a linear mapping of traffic to resource consumption, a fixed capacity parameter C expressed as a fraction of observed peak, and a one-to-one mapping of geographic cells to network nodes. These are strong and unrealistic abstractions for any real 5G deployment, where non-linear queue dynamics, handover overhead, and slice isolation constraints dominate. The authors describe these as simplifications but do not quantify their impact on the Δpp results. A sensitivity analysis varying the key assumptions (e.g., non-linear load curves, capacity uncertainty) is needed before operational claims can be taken seriously.

6. Absence of inferential statistics. Results are reported as means with standard deviations but no hypothesis tests are performed. Given the high variance visible in Figures 4 and 6 (e.g., RMSE error bars that overlap substantially between Ridge, Random Forest, and LSTM), it is unclear whether the differences are statistically significant. A Diebold-Mariano test or Wilcoxon signed-rank test across nodes would be standard practice.

Minor Issues / Questions for Clarification :

- **Section "From social signals to network demand"** cites Yang et al. (2016) as showing Twitter explains "up to 70% of variance." The phrase "up to" is doing heavy lifting; the authors should clarify the conditions under which this figure holds.

- **MAPE for LSTM_opt_cst is reported as 17.7%** in the body text but the abstract claims "up to 30%" error reduction without a clear denominator. These figures should be reconciled.
- **Figure 5** is poorly readable: the x-axis orders models inconsistently (LSTM, RandomForest, Ridge, Naïve, LSTM opt_std, LSTM opt_cst) and the line colour legend overlaps. The figure should be redesigned for clarity.
- **The 80/20 temporal split** is mentioned without specifying the cut date. For seasonal traffic data, the split point matters. Was the test set a contiguous future block or randomly sampled? Random sampling would constitute data leakage in a temporal forecasting setting.
- **Optuna hyperparameter search:** 10^3 trials are reported but the search space (epochs 50–500, neurons 50–500, batch 2–128, learning rate 10^−5–10^−1) is continuous. Were the same best settings applied across all 195 nodes and all 5 service clusters, or per-node? This affects whether the comparison is fair.
- **The term "slides"** appears in Figure 3's caption ("Multiple network slides are instantiated") — this should be "slices."
- **Keywords** are limited to "5G networks, demand forecasting, social media." More specific terms (LSTM, event detection, network slicing, proactive orchestration) would improve discoverability.

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?
Partly

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

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: Wireless Communication, Optical Communication, Free Space Optics, Telecom Networks, AI/ML in Telecom

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.

Author Response 27 Sep 2026

Marco MAMEI

Dear Reviewer, We sincerely thank the reviewer for the detailed and constructive assessment of our manuscript. The comments helped us identify several points where the previous version was unclear, over-generalized, or insufficiently specified. We have revised the manuscript substantially, with particular attention to the event-aware feature engineering, the scope of the empirical claims, the set of forecasting baselines, the operational metric definition, and the reproducibility of the experimental setup. Following the additional reproducibility audit prompted during the second review round, we also corrected the asymmetric-loss experiment, the overload metric terminology, the event-ablation numbers, and the robustness analysis. The responses below reflect the final revised manuscript. Below we provide a point-by-point response.

1. Critical underspecification of the event features We agree with the reviewer that the previous version did not describe the event-aware features with sufficient precision, especially given that this is central to the paper's contribution. To address this issue, we added a dedicated subsection entitled “Event-aware feature engineering”. In this subsection, we now explicitly distinguish between endogenous traffic features and exogenous event-aware contextual features. We specify that endogenous features include lagged traffic values, hour-of-day indicators, day-of-week indicators, and service-slice identifiers. We also clarify that, in the reproducible case study, the event-aware variables are primarily based on scheduled-event proxies and derived temporal indicators, rather than on a live Google Trends or real-time social-media ingestion pipeline. The revised text now explains how events are encoded as binary and intensity-like variables, how they are aligned with hourly traffic windows, how lead and lag indicators are used to account for early arrivals and delayed departures, and how missing event values are encoded as zero. We also clarify that the same event-aware feature set is used across models and that the no-event ablation removes only the event-related variables while preserving traffic lags and temporal encodings. We further revised the wording around web and social signals. Google Trends and social-media streams are now discussed as compatible sources of event-awareness in the broader pipeline, while the empirical results are explicitly interpreted as evidence of scheduled-event contextual features rather than as a direct evaluation of a live Google Trends pipeline.

2. Single-city, single-period experiment and generalizability We agree that the empirical evaluation is limited by its focus on Nantes during Spring 2019. We have revised the manuscript to make this boundary explicit and to avoid over-generalizing the results to operational 5G/6G deployments. The abstract, introduction, discussion, and conclusions were revised to use more cautious wording. We now refer to the empirical evaluation as a single-city case study based on real-world mobile traffic data, rather than as a direct validation on current operational 5G infrastructure. In the conclusions, we explicitly state that the results should be interpreted within the limits of a single-city case study based on Spring 2019 mobile traffic traces, and that broader validation across additional cities, longer time windows, and more recent deployments is required before generalizing the findings to operational 5G/6G systems. We did not add a second city in this revision because the main goal of the empirical component is to provide a reproducible, controlled case study

supporting the proposed event-aware forecasting and orchestration pipeline. However, we agree that multi-city validation is necessary, and we now identify replication across additional NetMob'23 cities and current operational deployments as a primary direction for future work.

3. Absence of modern time-series baselines We agree with the reviewer that the previous benchmark was too limited and that the claim about LSTM performance needed to be evaluated against a more modern forecasting baseline. To address this, we revised the State of the Art section to include recent Transformer-based and graph-based spatio-temporal forecasting models. In particular, we added discussion and citations for PatchTST, STGCN, and STAEformer. We then added a PatchTST-style Transformer baseline to the empirical comparison. The PatchTST-style baseline was evaluated using the same chronological protocol as the other models: the final 20% is held out for testing, while the preceding 80% is split chronologically into training and validation, yielding an effective 64%/16%/20% train/validation/test partition after lag construction. The same node-level setup, event-aware input variables, accuracy metrics, and operational OR metrics are used for comparison. We also added the relevant implementation parameters, including the 24-hour input context, patch length, stride, model dimension, number of attention heads, encoder layers, dropout, optimizer, learning rate, weight decay, batch size, and validation-based early stopping. The revised results show that the PatchTST-style baseline improves over the naïve predictor in terms of forecasting error, but does not outperform the standard LSTM, Random Forest, or Ridge models in this specific node-level hourly forecasting setting. We report this result explicitly in the Results and Discussion sections, and we revised the wording to avoid making a general claim of LSTM superiority beyond the evaluated baselines. Regarding graph-based models such as STGCN and STAEformer, we added a dedicated discussion explaining why their direct application is not straightforward in the present dataset. NetMob'23 provides spatially aggregated mobile traffic measurements, but it does not expose the operational radio topology, handover relations, or network-level adjacency graph. Applying graph-based models would therefore require constructing an inferred topology, introducing additional modelling assumptions that would affect comparability. We now identify deeper evaluation of Transformer-based and graph-based models under richer multivariate and topology-aware settings as future work.

4. OR metric definition and normalization We agree that the previous operational presentation was misleading because the manuscript mixed percentage-point terminology with a code implementation that actually returned a relative percentage reduction. We revised the operational metric definition and presentation accordingly. For q in {0.70, 0.80, 0.90}, the static capacity C_q is the q-th percentile of observed evaluation demand. Forecast-informed capacity at time t is max(C_q, P_t), and overload reduction is OR_q = 100 x (O_static - O_dynamic) / O_static. Table 1 and Figure 5 therefore report OR@70/80/90 as relative percentages, not percentage points. The corrected full-search opt_cst values are 71.3%, 69.1%, and 61.2%, compared with 65.5%, 61.8%, and 59.2% for opt_std. This makes the comparison among models more transparent while avoiding the former normalization ambiguity. Ridge and other simple baselines remain competitive under selected thresholds, so the revised text does not interpret operational performance as a monotonic function of model complexity.

5. Orchestration simulation assumptions We agree that the orchestration model relies on simplifying assumptions and that the previous version did not sufficiently delimit the operational interpretation of the overload-reduction results. We revised the Experimental Design section to clarify that the orchestration component is an abstract percentile-capacity evaluation loop rather than a full 5G network simulator. The model abstracts detailed RAN/core mechanisms and uses the same retrospective static-capacity benchmark for every forecasting model. Forecasts only raise capacity above the static percentile when P_t exceeds C_q; the resulting overload is compared with the static reference through OR. We also revised the framing of the results. OR is presented as a retrospective orchestration-oriented indicator, not as a direct estimate of operator-level performance, energy savings, or capex. A deeper sensitivity analysis involving non-linear load curves, capacity uncertainty, handover effects, and slice-isolation constraints remains future work.

6. Absence of inferential statistics We agree that the previous version relied primarily on descriptive statistics and that this limits the strength of claims about differences among models, especially when error bars overlap. In the revision, we have therefore avoided interpreting the results as formal statistical significance claims. The wording in the Results, Discussion, and Conclusions was revised to describe the findings as empirical differences observed in the considered case study and under the adopted experimental protocol. We also report mean and standard deviation consistently in the figures and tables, and we explicitly discuss cases where simpler models remain competitive. We acknowledge that formal paired statistical testing, such as Wilcoxon signed-rank tests across node-level results or Diebold-Mariano tests on paired forecast errors, would provide a stronger inferential basis. We now treat this as a limitation of the present empirical study and as an appropriate extension for future work, particularly in conjunction with multi-city validation.

Minor comments
1. Clarification of the “up to 70% variance” statement We revised the related-work discussion to avoid over-interpreting this prior result. The statement is now presented as evidence from a specific prior empirical setting rather than as a general property of social signals.

2. Clarification of the “up to 30%” error reduction We replaced the previous “up to 30%” wording with the corrected full-search opt_cst comparison and an explicit denominator. With event-aware inputs, opt_cst obtains RMSE = 2.24 x 10^7, MAPE = 24.01%, and R2 = 0.901; without event-related features, the corresponding values are RMSE = 2.59 x 10^7, MAPE = 28.9%, and R2 = 0.812. Using the no-event configuration as the denominator, event-aware inputs reduce RMSE by 13.4% and MAPE by 16.9%, while R2 increases by 0.089.

3. Figure 5 readability Figure 5 was regenerated as a grouped bar chart showing OR@70, OR@80, and OR@90 on one consistent relative-percentage scale. The model ordering is consistent with Table 1, and the former percentage-point terminology and normalized-to-best representation have been removed.

4. Temporal split We clarified the complete chronological partition. The final 20% of the post-lag supervised samples is held out as a contiguous test block. Within the preceding 80%, the final 20% is reserved for validation, yielding an effective 64%/16%/20% train/validation/test split. No random shuffling is used, and the test block is not used by

Optuna.

5. Optuna hyperparameter search We clarified the optimization setup for the LSTM variants and reran the corrected asymmetric-loss experiment with the full 1000-trial pipeline. The final opt_cst result is MAE = 1.60 x 10^7, RMSE = 2.24 x 10^7, MAPE = 24.01%, and R2 = 0.901; opt_std retains the lower RMSE of 1.94 x 10^7. We also added the PatchTST-style training parameters and the validation/model-selection procedure to improve reproducibility.

6. Typographical issue in Figure 3 We corrected the typo “slides” to “slices”.

7. Keywords We updated the keywords to improve discoverability, including terms such as LSTM, event-aware forecasting, network slicing, and proactive orchestration.

Overall, we thank the reviewer again for the detailed comments. The final revised manuscript is more precise about the event-aware feature set, the historical scope of the case study, the train/validation/test protocol, and the operational benchmark. It now reports the corrected full-search asymmetric-loss results, a full no-event opt_cst ablation with explicit denominators, and an input-level robustness experiment that perturbs event information without modifying observed traffic.

Competing Interests: No competing interests were disclosed.