Highlights

Generative-aided and Context-aware Forecasting of Mobile Network Traffic

Andrea Pimpinella *, Alessandro E. C. Redondi †
* *Department of Management, Information and Production Engineering (DIGIP), University of Bergamo*
† *Department of Electronics, Information and Bio-Engineering (DEIB), Politecnico di Milano*
contact: andrea.pimpinella@unibg.it

- We use deep learning to forecast mobile traffic, with particular emphasis on traffic peaks.
- We integrated exogenous features with traffic input for closed-loop prediction.
- We generated synthetic data to address scarcity and enhance model generalization.
- We validated performance on real-world data, outperforming baseline methods.

Generative-aided and Context-aware Forecasting of Mobile Network Traffic

Andrea Pimpinella *, Alessandro E. C. Redondi †
* *Department of Management, Information and Production Engineering (DIGIP), University of Bergamo*
† *Department of Electronics, Information and Bio-Engineering (DEIB), Politecnico di Milano*
contact: andrea.pimpinella@unibg.it

Abstract

Mobile cellular networks are experiencing rapid growth in data demand, largely driven by data-intensive applications such as video streaming. In particular, the popularity of live events can induce abrupt and localized traffic surges, often resulting in congestion and performance degradation. **For these reasons, accurate traffic forecasting is expected to play an increasingly important role in future sixth-generation (6G) mobile networks, supporting both real-time operational responses and long-term capacity planning. In this work, we move beyond classical traffic forecasting approaches and propose an AI-based framework that explicitly conditions traffic predictions on contextual information available in advance, while also leveraging generative data augmentation to address data scarcity. Through a comprehensive analysis conducted on two major Italian cities and several real-world datasets spanning four years, we show that traffic dynamics exhibit strong correlations with the occurrence of football matches. Building on this observation, we design a forecasting methodology that combines historical traffic measurements with scheduled event information to forecast future traffic over the prediction horizon. To improve robustness under event-driven and high-load conditions, we further introduce a lightweight synthetic data generation strategy that mitigates the scarcity and imbalance of rare traffic patterns. Experimental results demonstrate that the proposed framework improves forecasting accuracy, particu-**

larly during peak and busy-hour regimes, compared to baseline traffic-only approaches

Keywords: Cellular networks, traffic forecasting, network management, network monitoring.

---

1. Introduction

Mobile cellular networks are experiencing rapid and sustained growth in data demand, with traffic volumes projected to reach 346 exabytes per month by 2028, and subscriptions expected to exceed nine billion by 2027 (Ericcson, 2024). This increase is largely driven by the widespread adoption of data-intensive applications, such as video streaming, which place significant capacity and performance requirements on network infrastructures. In particular, the popularity of live events, including sports and concerts, can cause abrupt and localized surges in mobile traffic, often resulting in congestion and degradation of the Quality of Experience (QoE) perceived by end users (Frömmgen et al., 2016; Sandvine, 2024).

To mitigate these effects, Mobile Network Operators (MNOs) are increasingly adopting proactive management strategies for communication resources (Zhang et al., 2019). Anticipating network behavior enables both real-time operational responses and long-term capacity planning. In this context, mobile traffic forecasting has become a cornerstone for proactive network management, supported by the growing maturity of Machine Learning (ML) and Artificial Intelligence (AI) techniques for time-series prediction and anomaly detection.

The importance of accurate traffic forecasting is expected to increase further with the advent of sixth-generation (6G) mobile networks, envisioned as AI-native systems where intelligence is embedded across all network layers. Recent 6G visions emphasize end-to-end, system-level optimization across radio access, transport, and core network domains, supported by data-driven and AI-based control mechanisms (Saad et al., 2019; Tataria et al., 2021). 6G will support a wide range of high-bandwidth and latency-critical applications, such as extended reality (XR), holographic communications, autonomous mobility, and industrial automation. These heterogeneous services are expected to generate highly dynamic and complex traffic patterns, characterized by rapid demand vari-

ations and correlated load increases across large portions of the network. In this context, the ability to reliably forecast traffic evolution at a system level becomes essential to prevent congestion, support proactive resource orchestration, enable network slicing, and improve energy-efficient operation (Foukas et al., 2017). Consequently, the capability to reliably forecast traffic under such conditions will be fundamental for resource orchestration, network slicing, and energy-efficient operation in future intelligent communication infrastructures.

Despite notable progress, most existing traffic forecasting algorithms are trained exclusively on network-internal KPIs and telemetry, and can therefore be interpreted as classical pattern-extrapolation methods, as they primarily learn periodicity and seasonality from historical traffic measurements. While effective under regular operating conditions, this formulation provides limited ability to anticipate traffic deviations caused by external phenomena that are not recurrent in the traffic history. Among these, sporadic and non-recurrent peaks—often triggered by exogenous events such as public events or large public gatherings—pose particular challenges, as they are rare, weakly represented, and difficult for models to learn effectively. As a result, predictive models tend to overfit isolated cases rather than capture generalizable relationships, yielding poor performance on unseen data. Although the occurrence and timing of such events are often known in advance, their impact on network load is typically inferred indirectly from past traffic observations or handled through open-loop or post-processing corrections, rather than being explicitly incorporated into the forecasting process. In this work, we move beyond this classical formulation and propose an AI-based framework for mobile traffic forecasting that explicitly conditions predictions on scheduled events by embedding future event information directly into the learning architecture. This enables a closed-loop forecasting formulation, where traffic patterns and event-driven deviations are learned jointly within the model. To make this formulation effective in practice, we further introduce a generative data augmentation strategy designed to mitigate the scarcity and imbalance of event-driven high-load samples. The main contributions of this paper are summarized as follows:

- **First, we design a deep-learning architecture for mobile traffic forecasting, which jointly learns from historical traffic measurements and contextual information available in advance, such as the scheduling of football matches.** Differently from our previous work (Pimpinella et al., 2022b), where contextual information was applied in an open-loop manner to post-process the output of a conventional forecasting model, the proposed approach integrates event-related features directly into the model's input, enabling closed-loop learning of the relationships between external phenomena and traffic dynamics.

- To mitigate the scarcity and imbalance of rare high-load traffic samples, a lightweight synthetic-data generation method is introduced to augment the training dataset, thereby improving model generalization. The proposed generative process is based on a time-series decomposition approach, where the trend component is synthesized using Bézier-curve modeling with control points, and peaks are added separately in an additive fashion to reflect real traffic variability.

- We validate the proposed framework on real traffic measurements collected **from two operational mobile networks** in Italy over different periods spanning from 2020 to 2024. We compare our approach with baseline methods that either rely exclusively on network-internal data or do not employ data augmentation, demonstrating higher predictive accuracy and robustness, **with respect to both mid and long-term look-ahead forecast horizons and different granularity of data aggregation in the spatial domain.**

The remainder of the paper is organized as it follows: Section 2 briefly describes literature works related to network traffic forecasting and synthetic data generative approaches. Section 3 introduces the datasets and details the features extraction process, while Section 4 describes the synthetic data generation strategy. Section 5 introduces the baseline and the proposed deep-learning based framework, whose experimental evaluation is contained in Section 6. Finally, Section 7 concludes the paper.

2. Related Works

Deep learning-based cellular traffic (Li et al., 2023; Wang et al., 2024) forecasting have gained popularity in recent years. In (Li et al., 2023), the

authors propose a meta-learning based cell-level network traffic prediction framework, that leverage LSTMs as base-learners and KNN as meta-learner, achieving mean absolute errors below 4%. Similarly, authors in (Wang et al., 2024) leverage LSTMs to to predict traffic generated by users associated to a given base station, and use the prediction outcomes to switch users connections from underutilized base stations to reduce network access energy consumption. In the context of traffic forecasting, research has recently focused on the specific task of predicting busy-hour cellular traffic (Pimpinella et al., 2022a; Li et al., 2024a), i.e., the highest traffic demand in a day or a week. Focusing on busy-hour cellular traffic forecasting in the long term (longer than one month ahead), the authors in (Pimpinella et al., 2022a) design a LSTM-based forecasting approach that first groups network cells with similar busy hour traffic profiles and then fits per-cluster forecasting models to predict the traffic loads. Results on a real cellular network dataset show that busy hour traffic can be forecasted with errors below 10% for look-ahead periods up to 2 months in the future. A similar objective is pursued in (Li et al., 2024a), where authors design two deep-learning based traffic peaks predictors able to outperform state-of-the-art baseline performance in terms of peak prediction accuracy by 50%.

Regarding model's features, besides using solely historical traffic data, forecasting solutions that exploit contextual information have also gained increasing momentum. **In (Zhang et al., 2025), the authors build on the contextual reasoning and generative capabilities of large language models (LLMs), and propose a context-aware wireless traffic prediction framework powered by LLMs. Although the model exhibits strong generalization and prediction performance on real-world 5G datasets, it does not specifically target traffic surges and only considers endogenous contextual information—i.e., context-based features derived from the cellular network itself.** In this vein, some of the works focus on the generally higher traffic volumes faced by cellular networks during occasional occurrences related to social or sport events (Pimpinella et al., 2022b), and propose forecasting architectures which include related information to improve the performance. In (Zanzi et al., 2020), the authors propose a deep learning-based, model-agnostic framework for predicting radio access network capacity. The evaluation, conducted using real-world events, demonstrates the framework's ability to assist mobile operators in anticipating the capacity demands of upcoming sport events. Likewise, in (Bejarano-Luque et al., 2021) the authors develop a data-driven approach to analyze

the influence of local mass events on daily cellular traffic patterns at single base stations. Using a dataset of real traffic traces, the study reveals that i) sport events consistently increase per-cell downlink traffic volume and ii) leveraging such evidence within traffic prediction in the short-term (1-hour ahead) improves accuracy up to approximately 30%, especially in scenarios involving large-scale events.

Regardless of the type of features used, data scarcity is often critical when dealing with mobile traffic prediction, especially when anomalies detection (e.g., forecasting of traffic peaks) is also targeted. A solution to this issue is using the real data available to generate synthetic data, and use them to assist the learning process and improve the final prediction performance. In this vein, deep learning (Wen et al., 2021) and more recently GANs (Lin et al., 2020; Zhang et al., 2023; Hui et al., 2023; Li et al., 2024b; Pandey et al., 2024) have been proved to be particularly successful, achieving overall good generalization and robustness. In (Zhang et al., 2023), the authors uses urban knowledge graphs to generate city-scale cellular traffic traces. Leveraging both city-related environmental factors and spatial and environmental contextual relations between base stations. Similarly, in (Hui et al., 2023) the authors first use GANs to learn temporal multi-periodic and long-term a-periodic cellular traffic patterns and then leverage urban knowledge to enhance traffic generation performance. **GANs have been used also in (Li et al., 2024b), where the authors develop GAN-based generator of user-level mobile traffic capable of learning mobile traffic usage behavior of different user classes and reproduce corresponding traffic patterns. Likewise, authors in (Pandey et al., 2024) introduce a novel approach using Generative Adversarial Networks (GANs) to create synthetic traffic data that closely mimics real-world statistics, while preserving user privacy. Results show the benefits on prediction performance due to including the artificially generated dataset within the learning process. However, while generally providing good data generation performance, generative approaches based on the use of Artificial Intelligence (AI), including those using GANs, present some challenging issues that often limit their practical applications on industrial and commercial settings. In particular, GAN-based approaches are acknowledged to require massive computational training resources, and non-negligible time to generate synthetic data. Moreover, the complexity of the training process, that is generally sensitive to model hyper-parameters**

and highly unstable, represents a complex drawback of such kind of approaches (Vu et al., 2025).

Differently from previous literature studies, including our prior work (Pimpinella et al., 2022b), this paper provides a comprehensive analysis of multiple dimensions of cellular network traffic forecasting, with particular attention to high-load and event-driven operating regimes. Our work distinguishes from the literature according to the following key aspects. First, we design and evaluate several deep learning architectures that jointly learn from endogenous (i.e., network traffic) and exogenous (i.e., football event schedules) variables to improve forecasting accuracy. In particular, our approach operate on traffic aggregates at the city and cell cluster level, making it fundamentally different in scope with respect to state of the art works. Second, we perform a longitudinal comparative analysis across several real-world datasets spanning four years, thereby demonstrating the robustness of the proposed frameworks in mitigating temporal performance drift. Instead of considering short prediction look-aheads, we perform long-term forecasting (one week to one month ahead) and demonstrate the model's effectiveness under both normal and anomalous traffic conditions, achieving high accuracy in forecasting traffic peaks. Finally, we address the challenge of data scarcity by introducing a lightweight, explainable, and AI–free synthetic data generation methodology, which is employed to augment the training datasets and ultimately enhance forecasting performance.

3. Datasets

This study utilizes two types of data. The first type refers to network Key Performance Indicators (KPI) gathered from a major mobile network operator. The second one contains sport events, in particular football matches, which will be used as exogenous input variables for the forecasting models.

3.1. Network KPI Dataset

We leverage three different datasets provided by one of the leading mobile network operators (MNOs) in Europe. Such datasets collects Key Performance Indicator (KPI) measurements from **two** LTE cellular networks, currently operating in Milan, Italy, home to 1.3 million residents (rising to

Table 1: Characteristics of social, spatial and network contexts relative to the cities of Milan and Rome (Italy), as observed in 2024.

	Milan	**Rome**
City Residents [M]	1.3	2.8
Metropolitan Area Residents [M]	3.7	4.5
City Area $[Km^2]$	180	1287
N. of Cells	5000	8000
N. of Access Sites	500	1500

3.7 million when including the metropolitan area) **and Rome, Italy, capital of the country, counting up to 4.5 million people in the metropolitan area. We provide in Table 1 a summary of the characteristics of social, geographical and network contexts relative to both environments. Note that leveraging traffic measurements from both cities, that are populated by users with historically different behaviours and work-life balance, allows to provide heterogeneous performance benchmark and robustly assess the generalization capability of the proposed prediction approach.**

Our analysis concentrates specifically on the hourly sampled downlink traffic volume KPI, representing the total amount of data downloaded each hour from every eNodeB in the network. We consider 4 datasets **(the former three from the cellular network deployed in Milan, while the latter from the cellular network of Rome)**, collected across distinct observation periods:

(A) (**Milan**) the first dataset covers the period from September 28th, 2020, to May 23rd, 2021, spanning almost eight months. In this period, Italy was characterized by a hybrid lockdown due to Covid-19 pandemic, with specific restrictions regarding attendance to public events (e.g., public access to football stadiums was prohibited).

(B) (**Milan**) the second dataset spans from January 10th, 2022, to March 13th, 2022, corresponding to nine weeks. In this period, public access to Italian stadiums was granted to citizens who completed the vaccination process, corresponding to roughly 90% of the whole population at the beginning of 2022.

(C) (**Milan**) the third dataset includes data from February 5th, 2024, to

June 3rd, 2024, spanning four months. No restrictions due to Covid-19 pandemic were active in this period.

(D) (**Rome**) **the fourth dataset includes data from September 28, 2020, to February 28, 2021, spanning five months. Similar considerations observed for dataset A apply to this dataset.**

Rather than forecasting the traffic of each individual eNodeB, we aggregate data at the city level, which offers several advantages for mobile network operators (MNOs). The radio access network (RAN) evolves continuously, with new sites and cells being deployed to expand capacity and improve coverage. Furthermore, the configuration parameters and operating conditions of individual eNodeBs often change to meet varying user demands and service requirements. As a result, cell-level analyses may yield inconsistent insights over time and can lead to reduced forecasting accuracy, while also increasing computational complexity due to the fine granularity of the data. In contrast, city-level aggregation provides a more stable and scalable representation of traffic dynamics, facilitating robust forecasting across extended time horizons. This level of aggregation is particularly relevant in the context of future sixth-generation (6G) networks, which are envisioned as AI-native systems requiring system-level traffic awareness to support end-to-end capacity provisioning, network slicing, and energy-efficient operation across radio access, transport, and core network domains (Saad et al., 2019; Tataria et al., 2021). Moreover, aggregation simplifies the integration of contextual exogenous information—such as major public events or population-scale mobility patterns—which typically affect traffic over large portions of the network rather than individual cells.

We pre-process the data in two steps. First, we aggregate the measurements at the eNodeB level by summing the downlink traffic volumes across all frequency layers and sectors, thus producing a single KPI time series for each eNodeB location, denoted as $v^e(t)$. Second, we aggregate $v^e(t)$ at the municipality level as it follows:

$$\mathbf{v}(t) = \sum_e \; v^e(t) \tag{1}$$

where $\mathbf{v}(t)$ represents the total hourly downlink traffic handled by the core network in the considered area, which is selected as the forecasting target.

3.2. Football Matches Dataset

The second dataset used in this study is a catalog of football matches. It provides details for each game, including the competing teams, the type of competition (e.g., domestic league or international tournament), and the start and end timestamps. As done in (Pimpinella et al., 2022b), this dataset is gathered automatically through a Python-based web crawler that periodically retrieves data from the websites of the Italian Serie A[1] and the major European Leagues (Champions League[2] and Europa League[3]). While no filtering is applied during the crawling process, we only retain in the dataset those games played by at least one team in the set [*AC Milan*, *FC Internazionale*, *AS Roma*, *SS Lazio*, *FC Juventus*, *FC Torino*, *SSC Napoli*, *Atalanta Bergamasca Calcio*, *ACF Fiorentina*], corresponding to the teams with largest fan-bases in Italy.

To ensure proper alignment of traffic and football calendar data sources, both datasets have been synchronized to a common temporal granularity of 1-hour intervals. As for traffic data, in this work the operator logs are natively recorded at 1-hour granularity. Regarding football match starting and ending timestamps, they are mapped to hourly bins by rounding quarter-hour offsets to the nearest lower (e.g., 14:15 is rounded to 14:00) or higher (e.g., 14:45 is rounded to 15:00) integer hour, ensuring any 1-hour bin overlapping with match start/end times is flagged as active. This procedure guarantees precise synchronization while preserving the temporal structure of both data sources.

3.3. Impact of Football Events on Traffic Dynamics

Figure 1 shows two weeks of aggregated traffic **from the cellular network of Milan** for each analyzed period, and the corresponding football events. Furthermore, Table 2 summarizes the characteristics of both the network and football calendar datasets, including the number of available cellu-

[1]https://www.legaseriea.it/it/serie-a/calendario-e-risultati
[2]https://it.uefa.com/uefachampionsleague/fixtures-results/
[3]https://it.uefa.com/uefaeuropaleague/fixtures-results/

Table 2: Details on Mobile Network and Football Calendar datasets. **Columns A, B and C refer to data collected from the cellular network deployed in Milan, while column D refers to data collected from the cellular network of Rome.**

Dataset	**A**	**B**	**C**	**D**
Traffic Samples(#)	5712	1512	3864	3528
Unique Matches (#)	202	45	128	129
Correlated Timestamps (%)	8.67	8.47	7.58	8.67
Avg. Traffic Overhead (%)	16.82	19.58	15.19	21.86

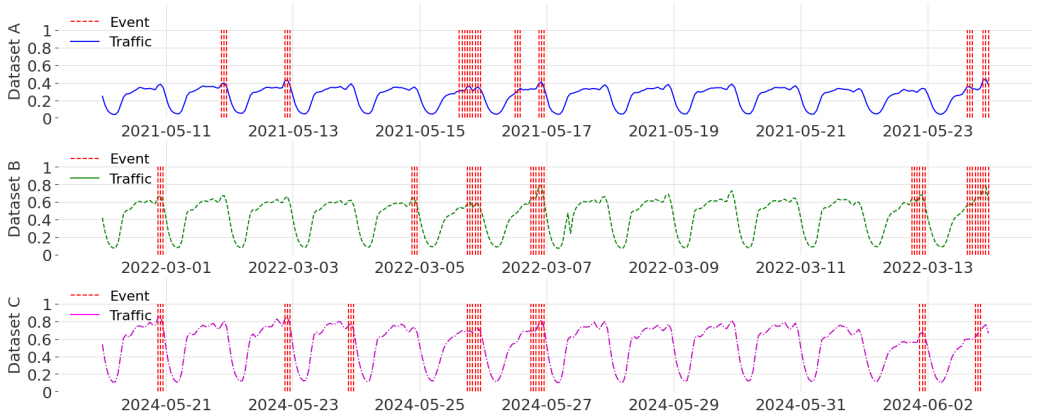
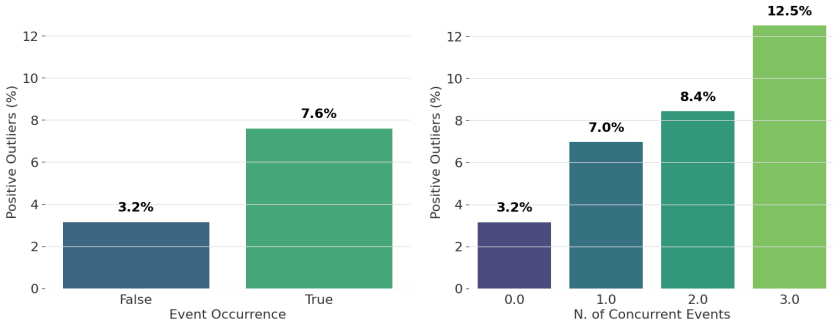


Figure 1: Last two weeks of traffic samples as collected **from network of Milan** in datasets A (top), B (middle), and C (bottom). Vertical (red, dashed) lines highlight football match occurrences. Data are normalized to minimum and maximum traffic value across the three datasets.

lar traffic samples, the number of scheduled matches, and the percentage of samples corresponding to hours when a football match was being played.

To illustrate the effect of football events on mobile traffic, **we keep the focus on Milan cellular network and show** in Figure 3 the distribution of hourly downlink traffic volumes conditioned on the occurrence of a football match, using dataset C as an example. Traffic data are sampled hourly, whereas football matches start at arbitrary timestamps and last approximately two hours. Consequently, each match is associated with three traffic samples: the hour preceding kickoff and the two subsequent ones. As shown in the figure, traffic volumes are on average 15% higher during hours overlapping football matches compared to periods without any ongoing match. **To further clarify the relationship between football events**



(a) (b)

Figure 2: Event flag is raised for 3 hourly timestamps, i.e., 1 hour before + 2 hours to account for match length. The y-axis shows the percentage of positive outliers with respect to total traffic samples observed in dataset C. Left: the fraction of outliers is computed conditionally on the occurrence of a football event. Right: the fraction of outliers is computed conditionally on the occurrence of 0, 1, 2, or 3 concurrent football events.

and cellular traffic anomalies, we show in Figure 2 the potentially multiplicative effect that overlapping occurrences might have on cellular traffic. In particular, the left side of the Figure plots the fraction of outliers computed conditionally on the occurrence of a football event, while the right side plots such fraction computed conditionally on the occurrence of 0, 1, 2, or 3 concurrent football events. As one can see, football events scale up outlier rate by a factor of 2.4, highlighting that traffic anomalies likely occur during football matches. Moreover, the occurrence of concurrent events might further raise the chance to observe a traffic outlier, up to a factor 4 when 3 matches are played simultaneously. For the sake of completeness, Table 2 also reports the average traffic overhead observed during football matches in datasets A, B and D. Notably, these statistics do not reveal substantial longitudinal (i.e., between 2020 and 2024) **or geographical (i.e., between networks of Milan and Rome)** variations in the impact of football events on cellular traffic. This observation suggests a persistent user inclination to consume sport-related video streaming content via cellular networks, even after the relaxation of COVID-19 restrictions on stadium attendance and public gatherings.

Despite such evidences, learning the relationship between the occurrence of football matches and the corresponding traffic overhead—and leveraging it to enhance forecasting accuracy—remains a non-trivial task, primarily

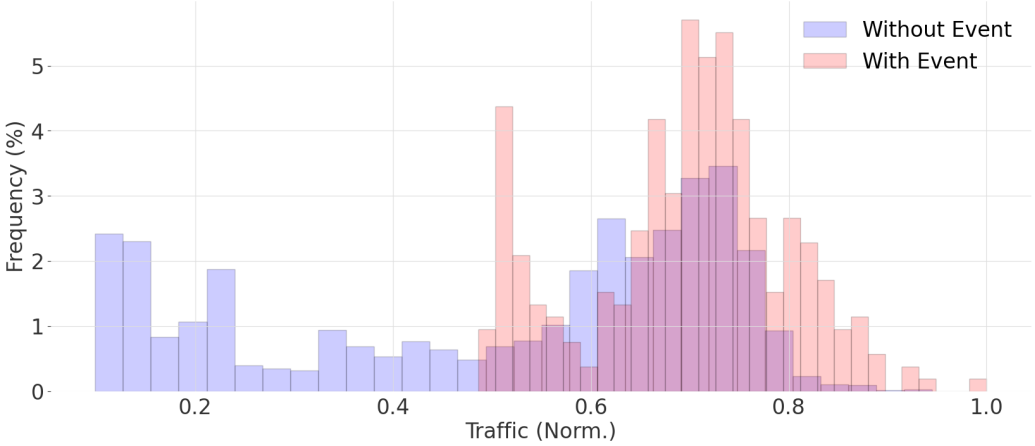


Figure 3: Distribution of (normalized) traffic samples in dataset C (**Milan**, 2024) conditioned to the occurrence of a football match (red), or not (blue).

due to the scarcity of representative data samples. As reported in Table 2, football-related samples account for only about 8% of the total hourly traffic measurements, thus representing a minority class **in all the datasets**. This data imbalance limits the generalization capability of forecasting models. To mitigate this issue, we introduce in the following Section a synthetic data generation framework designed to enrich the training set and improve model robustness in forecasting rare high-load traffic events.

4. Synthetic Traffic Generator

Data scarcity is a well-known challenge in network traffic forecasting, particularly when the goal is to predict occasional and non-recurrent anomalies, such as traffic peaks. These events are sparsely represented in real datasets, limiting the amount of direct training data and thereby increasing the complexity and instability of the forecasting task.

To mitigate this issue, we propose a synthetic data generation strategy that produces realistic realizations of the original traffic signal $\mathbf{v}(t)$. The generation process consists of the following main steps.

1. *Pre-processing:* We first pre-process the available data to align the original mobile-traffic time series with the football calendar dataset. Pre-processing also includes windowing the data into regular weekly segments and filtering out anomalous periods corresponding to major

holidays (e.g., Christmas and Easter) to avoid irregularities that could bias the extraction of seasonal patterns.

2. *STL Decomposition of the Original Series:* We first decompose the traffic signal $\mathbf{v}(t)$ into three additive and statistically independent components — seasonal $S(t)$, trend $T(t)$, and residual $R(t)$ — using Seasonal-Trend decomposition based on Loess (STL).[4] The expected periodicity τ of the time series is empirically set to 168 time steps (i.e., one week), based on visual inspection of the data. This decomposition facilitates the independent modeling and synthetic reconstruction of each component, enabling tailored generation methods for seasonal $S^s(t)$, trend $T^s(t)$, and residual $R^s(t)$ behaviors, which are detailed in the following.

3. *Seasonal Component Generation:* To reproduce the periodic structure of traffic, we compute the Median Weekly Signature $S_m(t)$ from the seasonal component $S(t)$. Each median weekly hourly sample corresponds to the median value of all true traffic measurements collected at the same hour on the same day of the week.

   The synthetic seasonal component $S^s(t)$ is obtained by replicating $S_m(t)$ $K \times N$ times, where N is the number of weeks in the original dataset and K is a user-defined parameter controlling the length of the synthetic series.

4. *Trend Component Generation:* To synthesize smooth, long-term variations consistent with real traffic, we use Bézier curves (Hansford, 2002). Let $B^n(t)$ be a Bézier curve of degree n with control points $\mathcal{P}_n = P_0, \ldots, P_n$, defined as:

$$B^n(t) = \sum_{i=0}^{n} \binom{n}{i} (1-t)^{n-i} \cdot t^i \cdot P_i, \quad t \in [0,1] \tag{2}$$

   By adjusting the number and positions of control points, the curve dynamics can reflect realistic network-traffic trends. For each synthetic trend, K independent Bézier curves of length N weeks are generated and concatenated to form $T^s(t)$. Figure 4 shows five examples of trends

[4]STL = Seasonal and Trend decomposition using Loess.

$B^n(t)$ synthetically generated as Bezier curves (solid lines) along with the real trend component extracted from the data available in dataset C (red, dashed line). We underline that using this approach to generate synthetic traffic trends has several advantages for MNOs. First, the smoothness of the curve ensure gradual changes in the trend, reflecting how network traffic typically evolves over time without abrupt shifts. Second, the curve is confined within the convex hull formed by its control points, which helps maintain realistic bounds on traffic variations and prevents unrealistic deviations. Third, control points provide an intuitive geometric interpretation, simplifying the modeling and tuning process to reflect long-term traffic behavior accurately. Finally, the approach is lightweight and can be flexibly adapted to generate diverse types of traffic trends.

5. *Residual Component Generation:* To capture stochastic variations and the effect of football events, we filter the original residual series $R(t)$ into two subsets:

   - $\mathcal{R}_1$: residuals during football matches,
   - $\mathcal{R}_0$: residuals at all other times.

   Both distributions are approximately Gaussian, as shown in Figure 5, with means and standard deviations $[\mu_1, \sigma_1]$ and $[\mu_0, \sigma_0]$, respectively. We generate 2000 synthetic samples from $\mathcal{R}_1^s \sim \mathcal{N}(\mu_1, \sigma_1)$ and $\mathcal{R}_0^s \sim \mathcal{N}(\mu_0, \sigma_0)$. One week of synthetic residuals is constructed by drawing $\xi\%$ from $\mathcal{R}_1^s$ and $(1-\xi)\%$ from $\mathcal{R}_0^s$, where ξ is a user-defined, positive integer input parameter controlling the density of football match occurrences in the synthetic calendar. We finally build $R^s(t)$ by independently repeating this process $K \times N$ times, stacking along the time axis the generated one-week-long time series.

6. *Synthetic Time Series Generation:* The final synthetic time series traffic $\mathbf{v^s}(t)$ is generated by summing the corresponding three (synthetic) components $S^s(t)$, $T^s(t)$ and $R^s(t)$. Once generated, the synthetic time series can be used together with real-world data to train a network traffic forecasting model. By adjusting K, we can freely vary the length of the synthetic series, which allows us to examine how different volumes of synthetic data influence the final prediction performance.

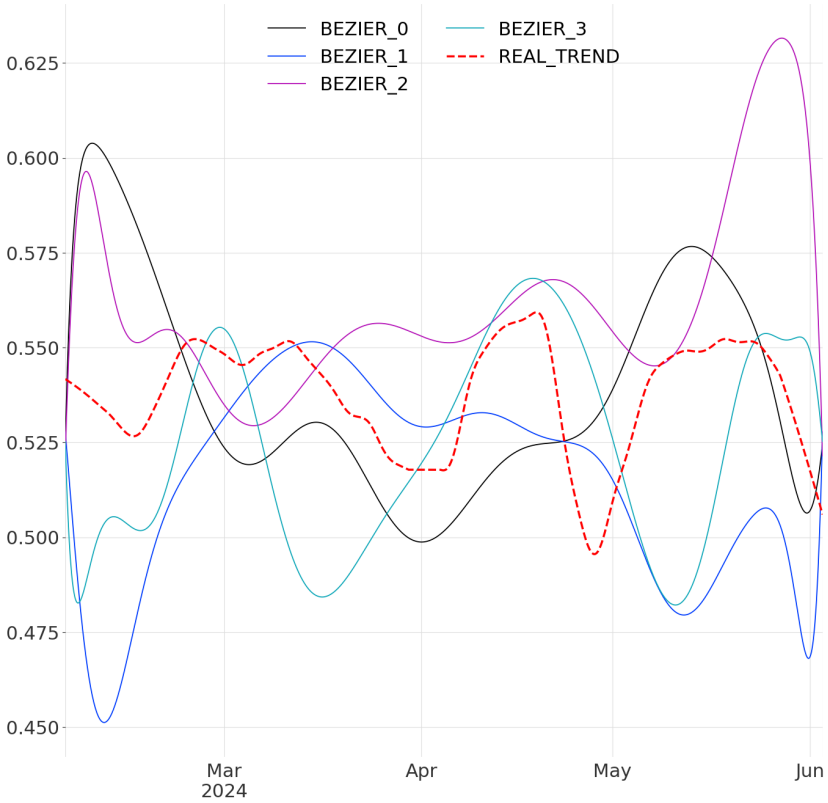


Figure 4: Four randomly trends generated as Bezier curves according to the proposed generation approach. The red dashed line represents the real (normalized) trend observed in data (dataset C, 2024, Milan).

We remark that each block of the proposed generative approach, i.e., seasonal, trend, and residual components generation—is explicitly modeled on the empirical distributions of real (training) traffic traces. This ensures that synthetic realizations faithfully preserve the underlying temporal dynamics and statistical properties observed in authentic network data, including rare anomalies like event-induced traffic peaks. The next Section presents the forecasting methodology adopted in this work.

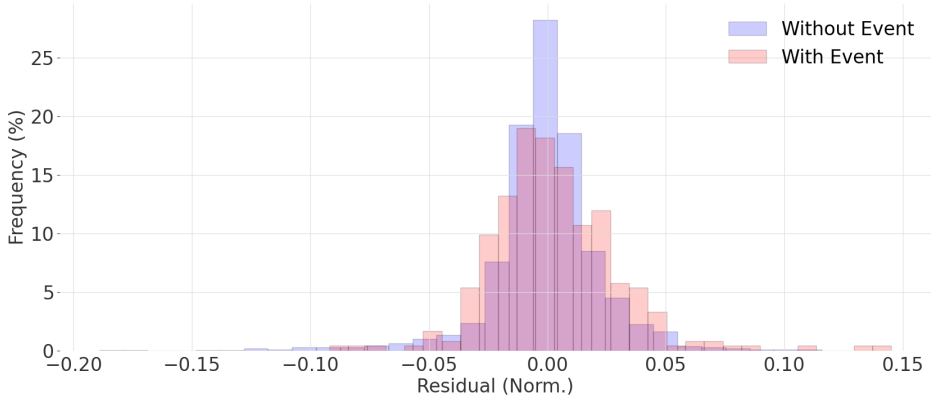


Figure 5: Distribution of (normalized) residual samples conditioned to the occurrence of a football match (red), or not (blue).

5. Forecasting Methodology

5.1. Problem Definition

Given an hourly sampled traffic time series, the objective is to forecast future traffic values over a predefined prediction horizon. While the forecasting target spans the entire traffic signal, particular attention is devoted to high-load regimes—such as busy hours and event-driven traffic peaks—which are the most critical from an operational perspective. In this paper, such regimes are treated as specific operating conditions of the traffic forecasting task and are analyzed through dedicated evaluation metrics. We consider several prediction models characterized by different complexity and features sets. For what regards complexity, we consider baseline models - obtained by Median Weekly Signature (MWS) based traffic profile estimation, alongside advanced deep learning structures, comprising feedforward, recurrent, convolutional layers, or their hybrid combinations. Note that while the baselines perform prediction by copying the traffic profile estimated from input data (and eventually adjusting such profile according to football calendar features, if available), advanced models adopt a classic machine-learning based strategy that perform one-shot mapping from input data to prediction output. Regardless of the model's complexity, we differentiate two cases depending on the features set used. The first option consists of using solely past traffic data samples to predict future values in an auto-regressive manner. Conversely, the second option integrates football calendar information

as additional predictive feature, thus also leveraging the knowledge about the scheduling of events at future time steps. **In particular, our implementation encodes the football calendar as a single, hourly sampled binary feature, where each entry takes the value of 1 if, at the corresponding timestamp, one or more football matches is taking place, and 0 otherwise. Note that the rationale for encoding the football calendar as a single binary feature stems from content providers' revenue maximization strategies, which typically schedule events to avoid overlaps and spread them across time slots. Although preliminary analysis suggests potential benefits from non-binary features (as shown in Figure 2), the low incidence of overlapping events in our dataset renders this extension unnecessary for the current work and suitable for future investigation.** In the following Sections, we provide details about the different prediction strategies.

5.2. Baseline Approach

We use the framework proposed in our previous work (Pimpinella et al., 2022b) as forecasting baseline. It works iteratively alternating the *learning* and *inference* phases as it follows. At time t, the learning phase leverages network data of the past N_l weeks, named $\mathbf{v}_l$, to estimate various parameters such as (i) the expected traffic profile in normal conditions, (ii) the set of events contributing most to traffic peaks in the area under analysis and (iii) the impact, in term of traffic volume, that such events have on the normal traffic profile. Such parameters are useful for the inference phase, which produces estimates for a look-ahead period lasting N_i weeks starting from t, named $\hat{\mathbf{v}}_i$. Generally, N_i is chosen by the MNO according to its own objectives, while N_l should be chosen such to maximize the forecasting accuracy. As mentioned, the two phases alternate, i.e., at time $t + N_i$ a new learning phase starts using the last N_l weeks. Overall, this framework consists of four building blocks:

1. *Traffic Signature Estimation*: the first building block of the framework constructs a MWS from the N_l weeks contained in the current learning phase;

2. *Peak Detection*: Concurrently with a new MWS estimation, the framework also performs traffic peaks detection. The old MWS (i.e., the one computed from the previous learning period) is replicated N_l times to

produce a time series $\mathbf{m}$, corresponding to the estimated median traffic profile during the training period. Then, an error signal $\mathbf{e}$ is computed as:

$$\mathbf{e} = \mathbf{v}_l - \mathbf{m}. \tag{3}$$

Such signal is analysed to find outliers, that is, anomalous peaks in the original traffic trace. This task is performed according to the definition of Tukey's fences, i.e., an observation is flagged as an outlier (peak) if it exceeds the range $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$, where Q_1 and Q_3 are the lower and upper quartiles of $\mathbf{e}$ and k is set equal to 3;

3. *Learning of Exogenous Events*: Once the traffic peaks have been identified in the learning data, the framework leverages the dataset containing football events to find the set $\mathcal{T}$ of teams whose matches are most correlated with the detected outliers. At the end of this process (for details, please refer to (Pimpinella et al., 2022b)), we flag all peaks detected in the training set as either *correlated* or *uncorrelated* with football matches, and we store the corresponding traffic volumes in the sets $\mathcal{C}$ or $\mathcal{N}$, respectively;

4. *Traffic Forecasting*: Once the learning phase has been performed on the N_l weeks of available data, the framework outputs a forecast $\hat{\mathbf{v}}_i$ for the next N_i weeks. Here we rely on a simple additive model, where $\hat{\mathbf{v}}_i$ is obtained by replicating the learnt MWS N_i times and summing a certain traffic value δ at those timestamps where matches of the teams in $\mathcal{T}$ are scheduled to be. In this work, δ is a constant, equal to the 95-th percentile of the distribution of $\mathcal{C}$.

Figure 6 illustrates a representative example of the forecasting process in the week between 01/03/2021 and 07/03/2021. The yellow line shows the forecasted traffic obtained by summing to the MWS the estimated δ volume at the timestamps (bottom axis, red crosses) where the selected teams are playing, while detected peaks are represented as blue circles. While such naive forecasting method is lightweight and less complex than AI-based options, it still relies on a cumbersome learning phase based on a forecast-and-correct mechanism rather than on the joint exploitation of traffic and exogenous input data. In the next Section we address this limitation by designing a deep-learning architecture that simultaneously learns traffic temporal patterns and leverages future context information available from football calendar data.

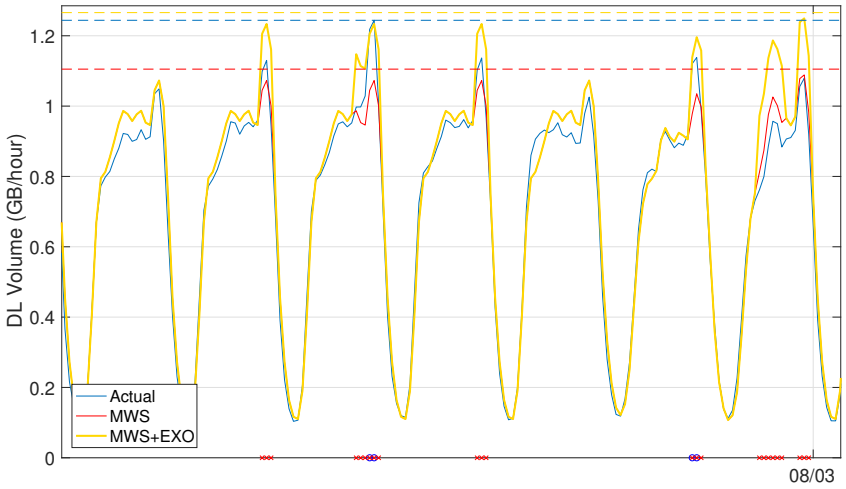


Figure 6: Traffic forecasting process: the estimated peak traffic volume δ is added to the learnt MWS (in red) at those timestamps where matches of the teams in $\mathcal{T}$ will occur. Dashed horizontal lines correspond to the maximum peak observed (blue) or forecasted (red or yellow).

5.3. Deep Learning based Approaches

An alternative approach to time-series forecasting is to model the non linear dependencies among data samples using deep-learning. Figure 7 plots the forecasting framework proposed in this work. In our framework, we employ different forecasting strategies depending on the nature of the input data, i.e., whether they align to *past* or *future* time steps. In fact, while future traffic samples (i.e., those belonging to the test set) represent the prediction target and can obviously not be leveraged as input, contextual features such as football calendar information are often known in advance for the entire season and can therefore be used as priors to enhance the final prediction accuracy.

The upper branch at the right side of the *Features Selector* in Figure 7 depicts the sequence of actions performed by our framework on past data, i.e., those aligning to timestamps before the prediction time ones. Regardless of the selected features, the *Past Learner* (PL) block processes past data samples to capture the corresponding auto-regressive dependencies. As depicted in Figure 8, we consider two implementations of such block:

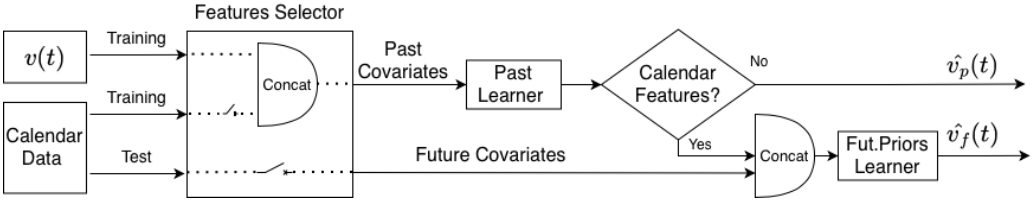


Figure 7: Overview of the proposed forecasting framework. The selected features are processed by Past and Future Priors Learners according to whether data samples align to past or future time steps, respectively. The latter option is enabled only when Calendar Data are included in the model.

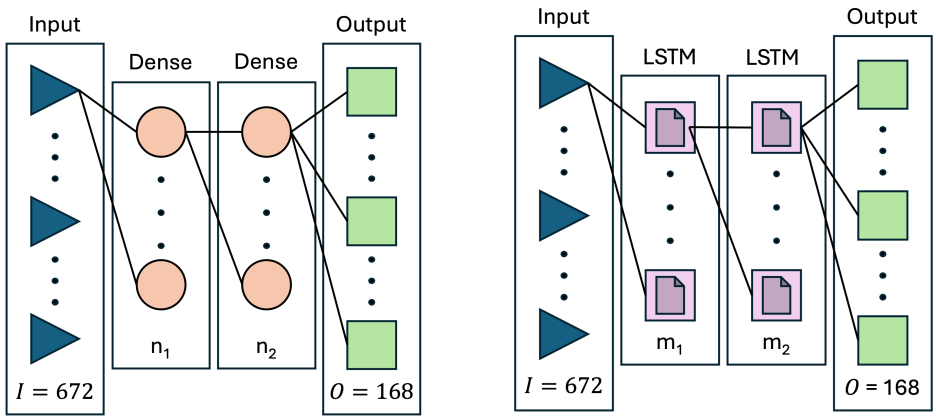
- *Feed-Forward Neural Networks* (FNN): a FNN model (Witten and Frank, 2002) generally consists of h stacked hidden layers, where each hidden layer is formed by n_h neurons fully connected to the units of the previous layer (as shown in the left side of Figure 8);

- *Long Short-Term Memory* (LSTM): these architectures belong to the class of Recurrent Neural Networks (RNNs) (Hochreiter and Schmidhuber, 1997) and typically consist of c stacked hidden layers, each composed of m_c memory cells (right side of Figure 8).

Note that irrespective of whether FNN or LSTM layers are used to implement the *PL*, this block takes as input a batch of I samples and outputs a forecast for the next O time steps (i.e., O represents the forecasting horizon). We refer to the output produced by the *PL* as $\hat{v_p}(t)$.

In addition to the *PL*, when contextual features (such as football calendar information) are available, we leverage a second learning block, namely the *Future Priors Learner* (*FL*), to leverage prior knowledge about future events. Such block is sequentially stacked after the *PL*, such that $\hat{v_p}(t)$ is concatenated with the corresponding co-variate features over the forecast horizon O (i.e., the FL has two input layers with $I_{FL} = O$ units each). The resulting concatenated representation is then processed by the *FL*. Figure 9 plots two different architectural options for the FL, based on either feedforward (FNN, left side) or 1-dimensional convolutional (1D-CONV, right side) hidden layers. Finally, the *FL* outputs a time series $\hat{v_f}(t)$ containing O samples, which represent the final forecast.

To summarize, when only traffic data are included in the model, the output forecast corresponds to $\hat{v_p}(t)$, and the *FL* is de-activated. Conversely, when football calendar data are also leveraged, the framework outputs $\hat{v_f}(t)$,



(a) FNN based architecture.

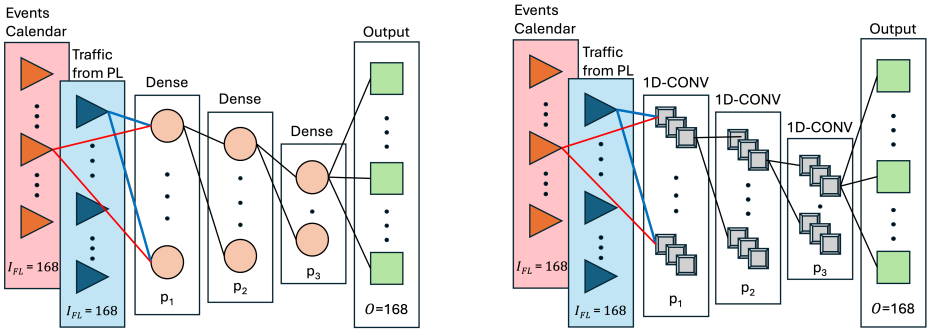
(b) LSTM based architecture.

Figure 8: Past Learner: architecture with h=2 hidden layers, each implemented as either FNN (left) or LSTM (right) layers. In this example, the PL takes as input 1 month of traffic data (i.e., I=672 units) and outputs 1-week ahead (i.e., O=168 units) traffic forecasts (i.e., $\hat{v_p}(t)$).

i.e., the output produced by the *FL*. Note that the parameters of *PL* and *FL* are optimized jointly during training when football calendar information are included in the model, ensuring end-to-end learning of both temporal dependencies and known future information. Details on the tuning of the hyper-parameters are provided in the next Section.

6. Results

In this Section, we evaluate and compare the performance of the proposed forecasting approaches on the available data. The experimental analysis is designed to investigate the interplay between several critical aspects of mobile traffic forecasting under different operating conditions: (i) the trade-off between model complexity—comparing lightweight, machine learning–free approaches with deep-learning-based models—and feature sets, contrasting traffic-only univariate models with multivariate models leveraging football calendar information; (ii) the longitudinal variation of forecasting performance across different datasets spanning four years; (iii) the impact of varying the volume of synthetic data used during the learning phase, as controlled by the hyper-parameter K; **and, iv) the sensitivity of the proposed framework with respect to variations in the forecasting horizon, i.e., mid- versus long-term predictive capabilities, the geographic**



(a) FNN based architecture.

(b) 1D-CONV based architecture.

Figure 9: Future Priors Learner: architecture with 3 hidden layers, implemented as either FNN (left) or 1D-CONV (right) layers. In this example, the FL concatenates at the input 1 week (i.e., I_{FL}=168 units) of traffic data (i.e., $\hat{v_p}(t)$) and future calendar data, and outputs 1-week ahead (i.e., O=168 units) traffic forecasts (i.e., $\hat{v_f}(t)$).

reference context, i.e., Milan versus Rome environments, and the spatial granularity of traffic aggregation, i.e., cluster-level versus city-wide scales. In addition to overall forecasting accuracy, the evaluation explicitly examines performance during high-load and event-driven periods, which are of particular operational relevance. Each approach requires to set the number of weeks of data used as input (i.e., N_l and I for baseline and advanced approaches, respectively) as well as the duration of the forecast horizon (N_i or O). Considering strictly consecutive training and testing series, we performed several experiments setting the look-back window length as equal to 4 weeks **and the forecast horizon as equal to either 1 or 4 weeks**. We provide in Table 3 provides a summary of the different forecasting algorithms as well as the corresponding features sets used.

Hyper-Parameters Tuning

For the baseline, to perform relevant tests and remove bias due to a particular choice of the forecast time window, each experiment is obtained by shifting forward by one week the train and test windows. In addition to the baseline approach presented in Section 5.2 (referred in the following as *B2*), we also assess the performance achieved using a simpler MWS-based forecast, which disregards any exogenous information and relies solely on the learnt MWS. This alternative method, denoted as *B1*, corresponds to setting

Table 3: Features sets used for each forecasting approach.

	Past Covariates		**Future Covariates**
Approach	Traffic	Fb. Calendar	Fb. Calendar
B1	✓	✗	✗
B2	✓	✓	✓
FNN	✓	✗	✗
LSTM	✓	✗	✗
FNN + FNN	✓	✓	✓
LSTM + FNN	✓	✓	✓
LSTM + CONV	✓	✓	✓

the correction traffic volume δ to 0.

For what regards deep learning models, we manually designed the PL using an encoder-decoder based structure consisting of $h = c = 2$ hidden layers with $n_1 = n_2 = 32$ neurons or $m_1 = m_2 = 32$ memory cells, depending on whether FNNs or LSTMs are used, respectively. Regarding the FL (that is only active when future co-variates are included in the model), it is composed of $f = 3$ hidden layers, consisting of $p_1 = 64$, $p_2 = 32$ and $p_3 = 16$ hidden units, respectively. If such layers are implemented as 1-dimensional convolutional layers, a kernel with size equal to 3 is used. Regarding activation functions, the units of fully connected and convolutional layers are activated by $ReLu$, while LSTM cells are activated by $tanh$, irrespective of the type of data they are processing. We choose Mean Squared Error (MSE) as loss metric, while the learning rate (η) is dynamically adjusted using an exponentially decaying function defined as $\eta = \eta_0 \cdot \exp^{-\alpha \cdot t}$, with $\eta_0 = 10^{-3}$ being the initial learning rate, $\alpha = 10^{-3}$ being the exponential decay rate, and t representing the current learning iteration. The learning process stops after $T = 200$ epochs.

Performance Metrics

As summarized in Table 4, we chronologically split the data available in each dataset in training (80%) and test set (20%), such that the training set includes data from the earliest time periods, while the test set covers the last, subsequent, non-overlapping intervals. While this strategy offers a well-balanced compromise between training set size and test set statistical consistency, it also prevents any temporal leakage, and ensures that all models are

Table 4: Train (top) and test (bottom) set split of the available data.

	Training		
Dataset	**From**	**To**	**Length (# weeks)**
A	28 Sep. 2020	18 Apr. 2021	27
B	10 Jan. 2022	27 Feb. 2022	7
C	05 Feb. 2024	05 May 2024	13
D	28 Sep. 2020	31 Jan. 2021	18

	Test		
Dataset	**From**	**To**	**Length (# weeks)**
A	19 Apr. 2021	23 May 2021	5
B	28 Feb. 2022	13 Mar. 2022	2
C	06 May 2024	02 Jun. 2024	4
D	01 Feb. 2021	28 Feb 2021	4

trained only on past information and evaluated on future observations. In particular, we evaluate performance considering the following three error metrics:

- Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{M}} \frac{(\hat{\mathbf{v}}_i - \mathbf{v}_i)}{\mathbf{v}_i}, \tag{4}$$

- MAPE on Events (E-MAPE):

$$\text{E-MAPE} = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} \frac{(\hat{\mathbf{v}}_i - \mathbf{v}_i)}{\mathbf{v}_i}, \tag{5}$$

- Busy Hour Error (BHE):

$$\text{BHE} = P_{70}[\{\frac{\hat{\mathbf{v}}_d^b - \mathbf{v}_d^b}{\mathbf{v}_d^b}, \forall d \in \mathcal{T}\}] \tag{6}$$

where $\mathbf{v}_i$ and $\hat{\mathbf{v}}_i$ are the i-th test set ground-truth and predicted traffic samples, $|\mathcal{M}|$ is the subset of test set timestamps where at least one football match is scheduled to be, while $\mathbf{v}_d^b$ and $\hat{\mathbf{v}}_d^b$ are the test set ground-truth

daily busy-hour traffic sample and corresponding forecast, respectively. Since we have one busy-hour sample per each day d in the test set, and thus a distribution of busy-hour forecast errors, we set BHE as equal to the 70-th percentile error (P_{70}) of such distribution.

Such three performance metrics enable a comprehensive evaluation of the proposed forecasting framework. On the one hand, the MAPE quantifies the overall prediction accuracy regardless of the operating regimes. On the other hand, BHE and E-MAPE measure the performance obtained considering two specific operational areas that are of particular relevance for mobile network operators in capacity planning and resource allocation. Specifically, BHE assesses the framework's capability to predict peak daily traffic values, while E-MAPE represents a modified version of the MAPE which considers exclusively those traffic samples associated with the occurrence of football matches. This latter metric thereby allows assessing the robustness of the model during event-driven traffic fluctuations, as derived from the available football events dataset. Collectively, the considered metrics encompass the key aspects necessary to evaluate the framework's ability to provide reliable traffic forecasts under both typical, high-load and event-induced operating conditions.

In the next sections, we first investigate how baseline and advanced forecasting models perform under real data usage (Section 6.1), and then how training datasets enriched with different volumes of synthetically generated data affect predictive accuracy and model tuning (Section 6.2). **Finally, we conduct a sensitivity analysis to evaluate the framework's robustness against key aspects related to cellular traffic forecasting, including the forecasting horizon, the reference networking geographical context and the traffic data aggregation granularity (Section 6.3)**. Overall, our analysis provide a comprehensive basis for interpreting performance trends longitudinally across different data compositions.

6.1. Forecasting Performance

We focus on Milan cellular network to compare the performance of the different forecasting approaches when only real, (i.e., non synthetic) data are used to train the models**, considering a forecasting horizon O equal to 1 week.** Section 6.1.1 assesses the benefits of using advanced forecasting techniques over baseline approaches, while Section 6.1.2 investigates the

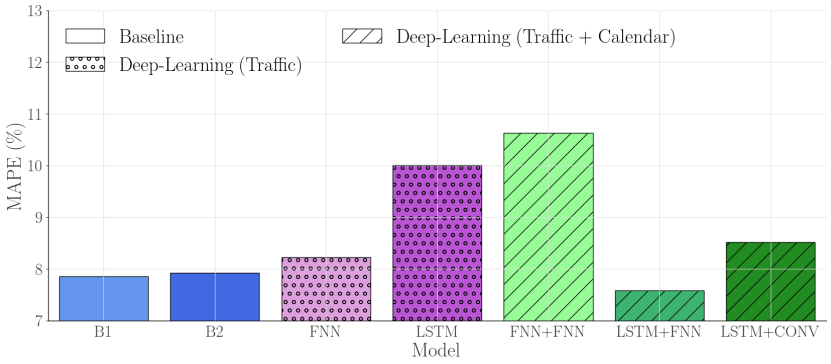


Figure 10: MAPE - Dataset A.

longitudinal variation of performance across datasets A, B and C.

6.1.1. Baseline vs Deep Learning Approach

As first set of experiments, we use dataset A and compare the forecasting performance of deep-learning based approaches with those obtained from baselines B1 and B2. In the following, we assess the performance of the different models with respect to i) forecasting traffic regardless of the occurrence of a football match and ii) the specific task of forecasting traffic *during* the occurrence of a football event.

General Forecasting Performance

Figure 10 shows the barplot of performance as measured by MAPE, where baseline approaches correspond to blue bars, while deep-learning based approaches correspond to dotted-purple (future covariates are not considered for prediction) and dashed-green (viceversa) bars, respectively. As one can see, LSTM+FNN-A outperforms all the other approaches, achieving a MAPE equal to 7.59%. Close but worse forecasting performance are obtained in scenario B1-A, B2-A and FNN-A, which yield 3.3%, 4.2% and 7.7% higher MAPE, respectively. Conversely, LSTM-A, FNN+FNN-A and LSTM-CONV-A obtain a MAPE higher than 8.5%, with FNN+FNN-A performing at worst. Being able to predict the maximum traffic volume the network will have to deliver to end users is equally critical for MNOs, as it enables more efficient resource management and mitigates the risk of network congestion. To evaluate how the considered models capture such network anomalies, we plot in Figure 11 the BHE obtained from each of the proposed approach. Negative values of the BHE mean traffic underestimation, a situation which

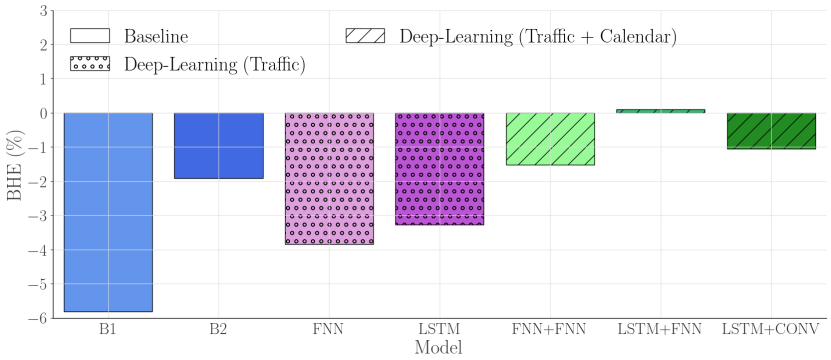


Figure 11: BHE - Dataset A.

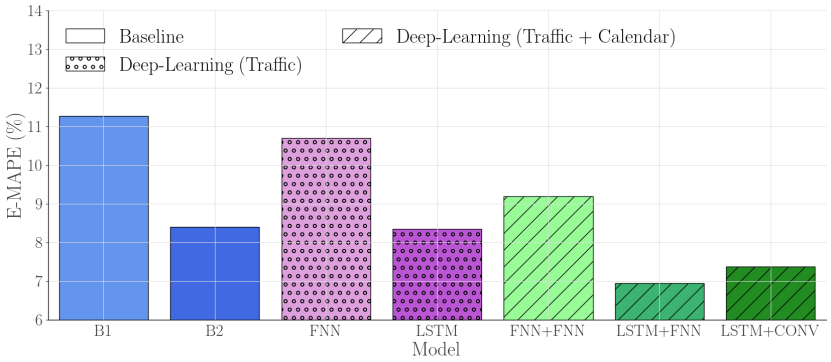


Figure 12: E-MAPE - Dataset A.

is potentially harmful for MNOs and that should be avoided at all costs. As one can see, both the simple MWS-based approaches and past covariates-only based deep learning methods suffer from such a problem, always producing forecasts which underestimate the maximum traffic peak between 2% and 6%. Conversely, scenario LSTM+FNN-A produces much better forecasts, such that the 70th percentile estimation error of maximum traffic peak is only 0.1% above of the actual observed traffic peak.

Forecasting Performance and Football Events

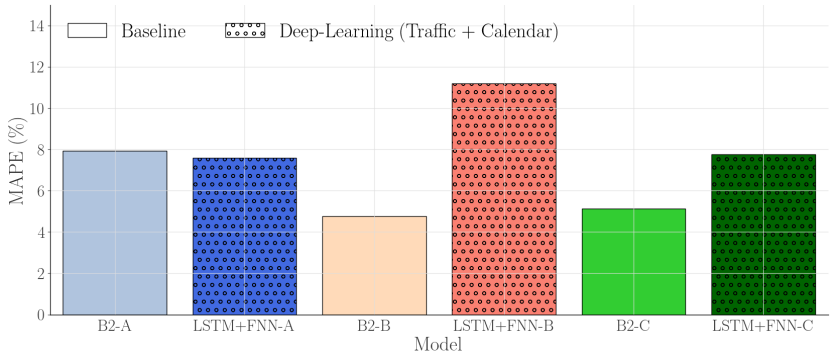
We now assess the capability of the considered models to capture the impact of the occurrence of football matches on network traffic, which often determines an increase of the overall generate volume, as shown in Section 3. With this aim, we plot in Figure 12 the E-MAPE obtained from the different approaches. Overall, we consistently observe the same trend across

both baseline and deep learning approaches: the models leveraging football calendar information (i.e., when both past and future covariates are considered for prediction) obtain better E-MAPE than their corresponding network traffic-only based versions. Among the models leveraging football calendar, all deep-learning approaches outperform B2, with best E-MAPE equal to 7% obtained in scenario LSTM+FNN-A. In the next Section, we investigate how the use of distinct training and test datasets across various observation periods impacts the evaluation of the different forecasting approaches.

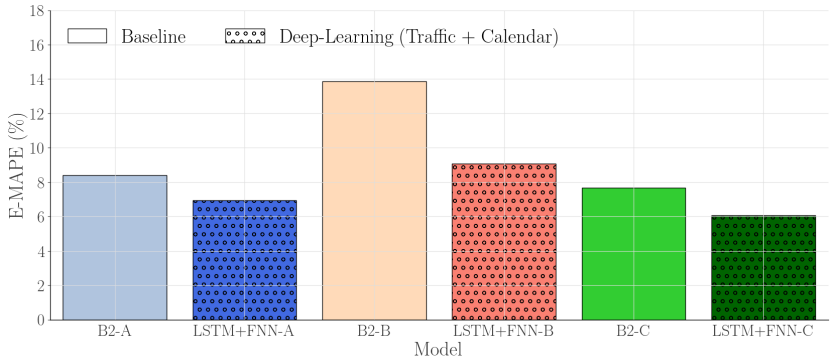
6.1.2. Longitudinal Performance Analysis

Studying the performance drift across time is of paramount importance to comprehensively evaluate the stability of deep-learning based forecasting approaches. To analyze this aspect, we focus on LSTM+FNN model (which performed best on dataset A) and update its performance by retraining and evaluating it on datasets B (LSTM+FNN-B) and C (LSTM+FNN-C). Figure 13 shows the MAPE (top), E-MAPE (middle) and BHE (bottom) obtained after such updates, and compare them against baseline B2, which also leverages football calendar data to perform prediction. Considering the MAPE (Figure 13a), we observe that B2 outperforms LSTM+FNN on both B and C datasets, achieving 37.2% and 41.3% lower error, respectively. Also, we observe that LSTM+FNN obtains lowest performance when trained and evaluated on dataset B: this is a direct consequence of the reduced amount of training data available in datasets B with respect to dataset A (75% less samples, as reported in Table 4). Conversely, LSTM+FNN-C does not suffer from the same issue, and performs comparably to LSTM+FNN-A, but worse then the corresponding baseline approach. Regarding BHE (Figure 13c), LSTM+FNN-C performs better than the baseline (the opposite happens for LSTM+FNN-B and B2-B), while both LSTM+FNN-B and LSTM+FNN-C perform worse than LSTM+FNN-A.

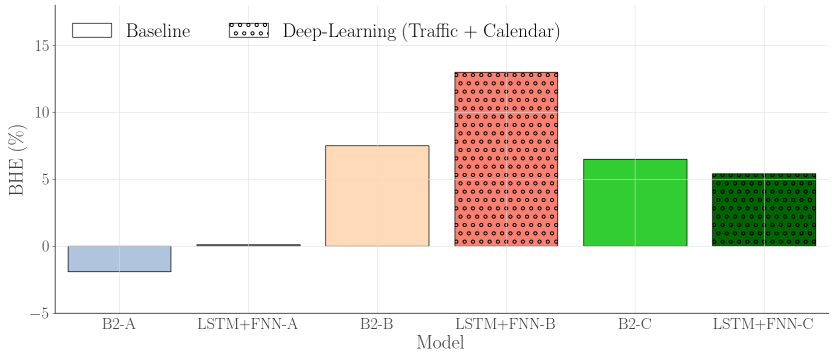
As final comment, results from scenarios B and C highlight the relatively low robustness of the additive approach underlying baseline B2: if smoother traffic conditions are observed during the occurrence of a football match, approach B2 might lead to severe forecasting errors, up to roughly 15% and 7% above actual traffic values for E-MAPE and BHE, respectively. Also, the analysis demonstrates that the insufficient availability of data samples for training deep-learning forecasting algorithms may either lead to a degradation in performance (as observed in scenario B) or cause such approaches to perform comparably to, or even worse than, simple baseline methods (as in



(a) MAPE



(b) E-MAPE



(c) BHE

Figure 13: Performance (top: MAPE, middle: E-MAPE, bottom: BHE) over datasets A (2020-2021, blue bars), B (red bars, 2022) and C (green bars, 2024) datasets.

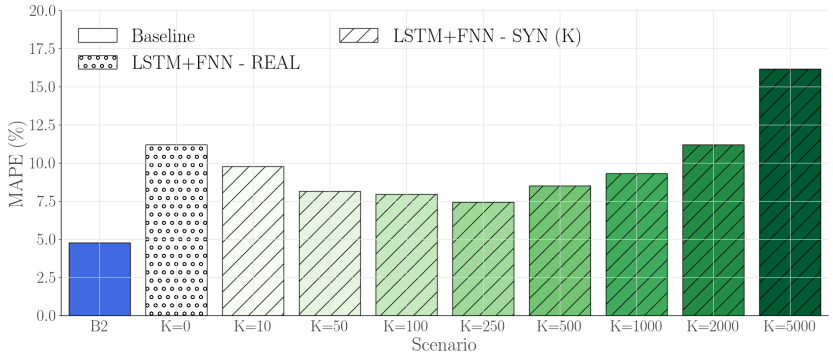
scenario C). In the next section, we address this issue by augmenting datasets B and C with synthetically generated data samples.

6.2. Benefits of Synthetic Data Generation

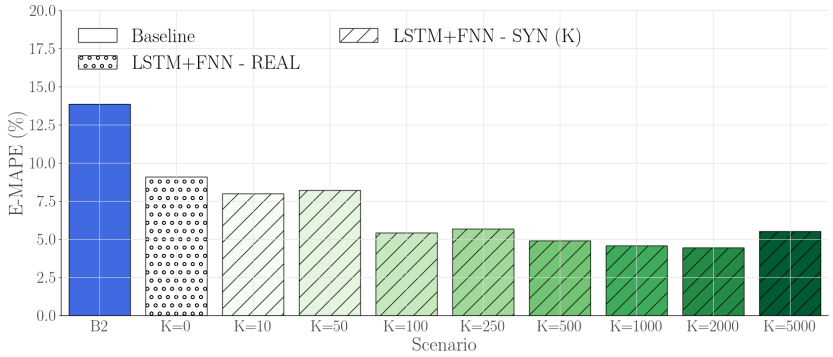
The availability of sufficient data for the accurate training of deep-learning models is crucial to achieving high performance standards. This is especially critical in anomaly detection scenarios, as often the anomalies (e.g., the traffic peaks caused by the live streaming of football matches) are strongly under-represented in the training dataset. To analyze this aspect, **we keep the attention on Milan cellular network** and resort to the data generation approach described in Section 4 to evaluate the forecasting performance of the different approaches when synthetically generated data are used to augment the training dataset. **For this study, the forecasting horizon is kept fixed and equal to 1 week. Note that only the training data are considered throughout the data generation process, while test data are left aside and used only for performance evaluation.** In the following Sections, we provide insights regarding key issues related to the use of synthetic data, that is the the trade-off related to the proportion of synthetic with respect to real data (Section 6.2.1) and the impact that a proper fine-tuning of the models on real data samples has on final performance (Section 6.2.2).

6.2.1. Synthetic Augmentation of Training Set

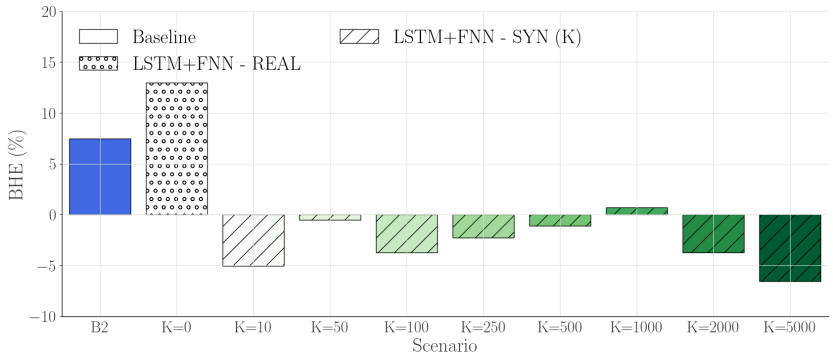
We analyze the trade-off existing between forecasting performance and K, i.e., the parameter controlling the amount of synthetic data used for training the forecasting models. In this experiment, we focus on the smallest of the available datasets, that is, dataset B, and on algorithm LSTM+FNN. Moreover, to isolate the impact of K on final performance, we completely ignore real data during the training process, and only use them for performance evaluation. We let K take values in the set $[10, 50, 100, 250, 500, 1000, 2000, 5000]$, and compare in Figure 14 the MAPE (14a), E-MAPE (14b) and BHE (14b) obtained for each scenario against the baseline approach (i.e., B2) and the case where model LSTM+FNN is trained only on real data (corresponding to $K = 0$). As one can see, the MAPE of LSTM-FNN improves for increasing values of K up to $K = 250$, beyond which performance start to decrease. A similar trend is also observed in BHE metric. This sugggests that increasingly aggressive data augmentation benefits performance until it does not lead to over-fitting, which instead represents an impairment. Conversely, the E-MAPE decreases steadily with K, suggesting that synthetic data helps the model to better capture the relationship between the occurrence of a football match and its impact on network traffic. Overall, training based on synthetic



(a) MAPE



(b) E-MAPE



(c) BHE

Figure 14: Impact of K on MAPE (top), E-MAPE (middle) and BHE (bottom) over Dataset B (2022). Models are trained on synthetic data samples, while real data are not used in this experiment.

data is preferable over using only real data according to all metrics, due to the scarce availability of the latter compared to the former - which also can

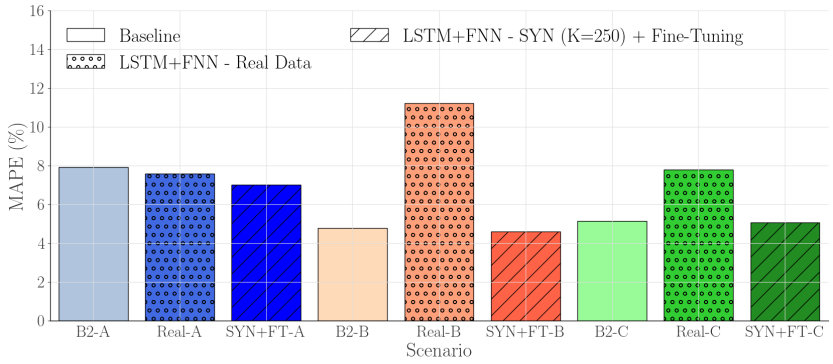
be augmented at will by means of the proposed generative approach. Nevertheless, while the use of synthetic data let LSTM-FNN further improve its performance with respect to the baseline in terms of E-MAPE (up to 67.9%) and BHE (up to 90.8%), B2 still achieves a MAPE that is 35.8% lower than the best one obtained by LSTM-FNN model on dataset B. In the next section, we show how to leverage the combination of synthetic and real data to further improve the performance of LSTM-FNN approach.

6.2.2. Impact of Fine-Tuning

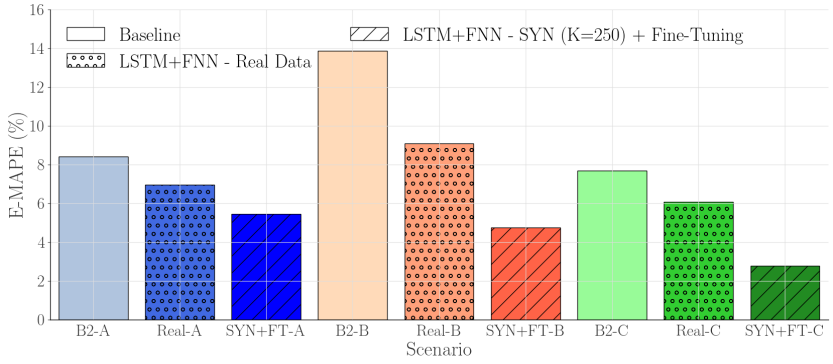
As last experiment, we fine-tune model LSTM+FNN on the real training data available in scenarios A, B, and C. Figure 15 compares the performance obtained as measured by the MAPE (top), E-MAPE (middle) and BHE (bottom) with those achieved by the baseline B2 and model LSTM+FNN when trained on real data samples only, i.e., when synthetic data are not used during the training process. In Figure 15, we refer to this latter model as *Real*, while its fine-tuned version is referred to as *SYN+FT*. As shown, the performance of LSTM+FNN improve across all metrics and scenarios after fine-tuning, making LSTM+FNN outperform the baseline approach on average by 5.5%, 55% and 51.6% with respect to MAPE, E-MAPE and BHE, respectively. Relatively to the specific contribution of fine-tuning to final performance, the greatest improvements of *SYN+FT* approach with respect to *Real* is observed in scenario B, where MAPE, E-MAPE and BHE are improved by 59.1%, 48.0% and 85.6%, respectively. This demonstrates how the combination of synthetic data generation and fine-tuning can benefit the quality of forecasts when scarce real data samples are available for training deep-learning models. Overall, results confirm that fine-tuning on real data represents a last fundamental training task to increase the accuracy of forecasting.

6.3. Sensitivity to Forecasting Horizon and Spatial Context

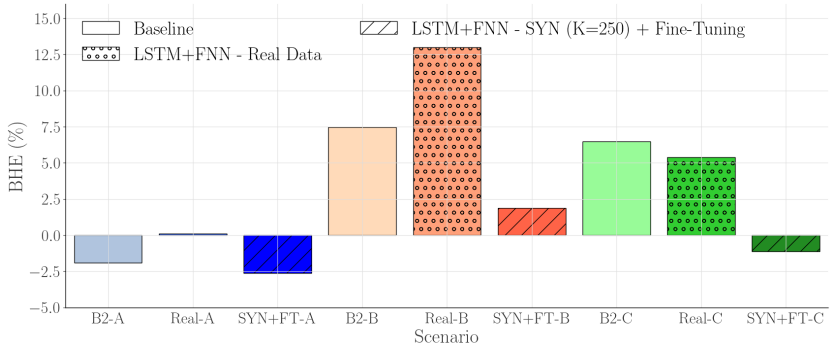
As final set of experiments, we conduct a sensitivity analysis to assess the robustness of the proposed prediction framework with respect to key modeling parameters typical of cellular traffic forecasting. First, Section 6.3.1 examines how extending the prediction horizon impacts performance, with respect to the same geographical context (i.e., network in Milan) and considering the most recent data collection campaign (dataset C). Secondly, Section 6.3.2



(a) MAPE



(b) E-MAPE



(c) BHE

Figure 15: Impact of fine-tuning on MAPE (top), E-MAPE (middle) and BHE (bottom) over datasets A (2020-2021, blue bars), B (2022, red bars), and C (2024, green bars). Models are trained on synthetic time series with $K = 250$ and then fine-tuned on real data.

evaluates how the framework adapts to a different geographical scenario, that is the cellular network deployed in the city of Rome, and compares the performance obtained when aggregating traffic data city-wide versus those corresponding to a smaller, cluster-based spatial aggregation granularity. For both analysis, we consider the traffic data observed during collection campaign D.

6.3.1. Forecasting Horizon

Longer traffic forecasting horizons enhance proactivity in both network resource management and detection of anomalies, but often suffer from reduced accuracy due to increased model training complexity. To investigate this aspect, we show in Table 5 the performance obtained when models LSTM+FNN and LSTM+CONV are used to produce 4-weeks ahead forecasts. In both cases, models are trained on synthetic time series of length $K \in [250, 500, 1000, 2000]$ and finally fine-tuned on real traffic samples. As one can see, the lowest MAPE is obtained by baseline B2, which yields 1.5% less percentage points than the best deep learning model, i.e, LSTM + FNN with K=250, achieving a MAPE equal to 5.69%. However, despite the slightly superior MAPE shown by B2, it shows substantially reduced forecasting performance during both high-load and event-induced operating regimes. Conversely, the proposed framework does not generally degrade its performance under such operational conditions, improving both E-MAPE and BHE up to more than 70% compared to B2 in the best case. Among the different tested options, model LSTM+CONV trained with either K=500 or K=2000 shows the best trade-off across the three metrics. Overall, our approach demonstrates a substantial robustness with respect to increased long-term forecasting horizons, achieving performance that are comparable to those obtained using 1-week look-ahead models.

6.3.2. Spatial Context

It is worth to evaluate the robustness of the proposed approach when applied to different networking environments, in terms of both users' behaviour and network size. In the following, we investigate on this aspect and assess the performance of the proposed approach with respect to traffic data aggregated over i) the

Table 5: **Prediction performance obtained with dataset C collected from Milan cellular network, when the forecasting horizon O equals 4 weeks. All the considered deep learning models leverage synthetic data (with variable K) and are eventually fine-tuned on real observations. Bold values highlight the top-3 performing models per each metric.**

	MAPE	**E-MAPE**	**BH**
B2	**4.25**	9.57	5.58
LSTM+FNN - Real	9.89	8.04	6.14
LSTM+FNN - K=250	**5.69**	5.51	-3.05
LSTM+FNN - K=500	6.76	5.34	-2.60
LSTM+FNN - K=1000	6.60	5.01	-3.21
LSTM+FNN - K=2000	7.78	5.88	-3.39
LSTM+CONV - Real	12.79	12.80	**1.38**
LSTM+CONV - K=250	7.65	6.50	-1.76
LSTM+CONV - K=500	6.51	**4.22**	**-0.38**
LSTM+CONV - K=1000	7.12	**4.46**	-1.89
LSTM+CONV - K=2000	**6.31**	**3.80**	**-1.49**
Best 1-week ahead:	5.05	2.76	-1.13

whole cellular network deployed in the city of Rome and ii) selected groups of cells (i.e., clusters) that show similar traffic pattern characteristics. In both experiments, we set the forecasting horizon as equal to 1 week.

City-wide Traffic Aggregation

The first column of Table 6 reports the MAPE and E-MAPE obtained when the proposed framework is used to forecast network traffic cumulated over the whole cellular network deployed in the city of Rome, when synthetic data are either excluded (*Real*) or leveraged ($K \in [100, 250, 500, 1000]$) during the training process. When synthetic data are used for training, models are then fine-tuned on real data samples: as one can see, while such models perform slightly worse (but comparably) than the baseline in terms of MAPE, they strongly outperform B2 on event-inducedd traffic anomalies, reducing the E-MAPE by more than 4 times. Consistently to what observed in Milan networking scenario, the combination of training the models on synthetic data and then fine-

Table 6: **Prediction performance obtained using dataset D collected from the cellular network deployed in Rome, with respect to city-wide (1st column) and cluster-based (2nd column: Residential Cluster, and 3rd column: Business Cluster) spatial aggregations. Bold values highlight the top-3 performing models per each metric. The forecasting horizon O is set as equal to 1 week.**

	City		Residential		Business	
	MAPE	**E-MAPE**	**MAPE**	**E-MAPE**	**MAPE**	**E-MAPE**
B2	**9.66**	27.02	10.65	39.46	**9.84**	**10.02**
Real	12.39	7.66	**8.61**	7.76	15.81	20.98
K=100	10.54	6.35	**7.25**	**6.61**	**12.24**	17.25
K=250	**10.24**	**5.64**	14.79	7.00	**9.38**	14.85
K=500	10.54	**5.89**	**9.23**	**6.21**	15.24	**11.78**
K=1000	**10.42**	**5.54**	9.65	**5.87**	13.87	**12.93**

tuning them on real data leads on average to more than 25% performance improvement across both metrics. Note that the severe traffic over-estimation yielded by B2 might accidentally lead to low busy hour errors, despite the substantially unreliable behaviour of such model when used to predict event-induced traffic anomalies. For such reason, and to avoid misleading conclusions, results on BHE are omitted in this case.

Cluster-based Traffic Aggregation

Besides forecasting traffic at the city-level, the ability to accurately predict traffic volumes cumulatively delivered by specific groups of cells, e.g., selected according to traffic pattern similarities, would enable fine-grained resource management strategies such as differentiated network slicing and adaptive capacity provisioning for distinct cell clusters, thereby improving utilization and QoS guarantees across the network. To evaluate the performance of our approach at lower traffic data aggregation granularity, we focus on the city of Rome and select two distinct clusters of cells grouped by their weekly traffic patterns. Figure 16 shows the cumulative traffic delivered by these cell groups over one week. The red profile, which aggregates traffic of roughly 1500 cells (i.e., 37 % of the network, responsible of 48% of the overall traffic), exhibits a strong evening peak between 8:00 p.m. and 10:00 p.m., regardless of the day, characteristic of *residential* areas with uniform but

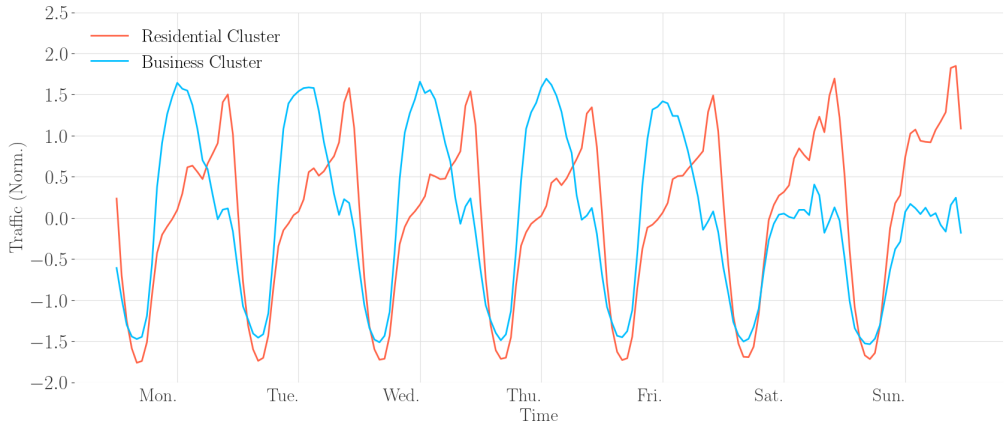


Figure 16: One week of (normalized) traffic data (dataset A) aggregated over the clusters of Residential (red) and Business (blue) cells deployed in Rome.

lower demand during working hours. Conversely, the blue profile (∼500 cells, corresponding to 12% of the network and generating 6% of overall traffic) shows heavier traffic during weekdays from 9:00 a.m. to 6:00 p.m., reflecting typical European *business* hours, with consistently lower volumes otherwise.

Table 6 reports the MAPE and E-MAPE (columns 2 and 3) obtained by our approach when forecasting traffic for both clusters. For the residential cluster, we generally observe that our approach outperforms the baseline across the majority of configurations and for both metrics. The best performance trade-off is achieved by training on synthetic data ($K = 100$) followed by fine-tuning on real data, which yields a MAPE of 7.25% and E-MAPE of 6.61%, corresponding to 1.5x and 6x improvement over the baseline, respectively. Conversely, method B2 outperforms our framework for the business cell cluster. Moreover, in this latter scenario, E-MAPE consistently exceeds MAPE across all forecasting approaches in most of the cases, indicating a reduced impact of football events occurrences on network load for business cells group. This effect is reasonable, as home environments enjoy substantially more football-related contents than workplace areas.

7. Conclusions

This paper addressed the challenge of integrating contextual information and synthetic data generation into deep-learning-based mobile traffic forecasting. By jointly learning from historical traffic measurements and football event–related information, the proposed framework produces more reliable traffic forecasts compared to baseline approaches. Results obtained on real network data validate the effectiveness of the proposed method, highlighting the potential of AI-driven, context-aware forecasting solutions to support proactive network management and capacity planning in future 6G networks, particularly under high-load and event-driven operating conditions.

The joint use of synthetic data samples and exogenous features boosted the performance of the proposed two-staged, **1-week ahead**, deep-learning architecture up to 59.1%, 48.0% and 85.6% with respect to MAPE, E-MAPE and BHE, ultimately making it outperform the baseline on average by 5.5%, 55% and 51.6%, respectively. Moreover, results highlight the specific benefits that data augmentation strategies have on anomaly detection, corresponding to an increase of the quality of forecasts up to approximately 35%. **Extensive experiments designed to test the performance on either longer forecasting horizons (up to 1 month ahead) or different spatial contexts confirm the robustness of our approach with respect to different prediction tasks and contextual domains.**

Several insightful trade-offs emerged from the analysis of the results. First, deep-learning strategies proved superior capabilities in learning the non-linear impact that exogenous variables have on network traffic than simple, MWS-based, additive approaches, this reinforcing the importance that AI-based traffic prediction is expected to have in 6G mobile networks. Nevertheless, approaches based on MWS copying remain effective entry-level forecasting options for MNOs, **especially for long-term forecasting horizons**, and could be considered in the design of lower-complexity forecasting systems. Second, data scarcity is a critical issue when using deep-learning approaches, especially with respect to anomaly detection tasks. Designing ad-hoc data augmentation strategies and selecting the correct volume of synthetic data to use is of paramount importance to achieving high performance. Finally, besides using synthetic data, fine-tuning the models on real observations revealed decisive to enhance performance: the careful design of such

process represents the last crucial step of the learning pipeline.

While this work focuses on traffic forecasting, the proposed approach is well suited to support proactive network optimization. By providing reliable traffic predictions under event-driven and high-load conditions, the forecasting model can be integrated into higher-level management and orchestration frameworks to enable proactive capacity scaling, network slice dimensioning, resource reallocation, or energy-efficient operation. Such closed-loop optimization architectures are widely envisioned for AI-native 5G and 6G networks, and the integration of the proposed forecasting component within these frameworks represents a natural direction for future work.

References

Bejarano-Luque, J.L., Toril, M., Fernández-Navarro, M., Gijón, C., Luna-Ramírez, S., 2021. A deep-learning model for estimating the impact of social events on traffic demand on a cell basis. IEEE Access 9, 71673–71686.

Ericcson, 2024. Ericcson mobility report. https://www.ericsson.com/en/reports-and-papers/mobility-report/. Last Accessed: 2024-12-09.

Foukas, X., Patounas, G., Elmokashfi, A., Marina, M.K., 2017. Network slicing in 5g: Survey and challenges. IEEE communications magazine 55, 94–100.

Frömmgen, A., Heuschkel, J., Jahnke, P., Cuozzo, F., Schweizer, I., Eugster, P., Mühlhäuser, M., Buchmann, A., 2016. Crowdsourcing measurements of mobile network performance and mobility during a large scale event, in: International Conference on Passive and Active Network Measurement, Springer. pp. 70–82.

Hansford, D., 2002. Bézier techniques.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 9, 1735–1780.

Hui, S., Wang, H., Li, T., Yang, X., Wang, X., Feng, J., Zhu, L., Deng, C., Hui, P., Jin, D., Li, Y., 2023. Large-scale urban cellular traffic generation via knowledge-enhanced gans with multi-periodic patterns, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 4195–4206. URL: https://doi.org/10.1145/3580305.3599853, doi:10.1145/3580305.3599853.

Li, F., Zhang, Z., Chu, X., Zhang, J., Qiu, S., Zhang, J., 2023. A meta-learning based framework for cell-level mobile network traffic prediction. IEEE Transactions on Wireless Communications 22, 4264–4280.

Li, S., Magli, E., Francini, G., Ghinamo, G., 2024a. Deep learning based prediction of traffic peaks in mobile networks. Computer Networks 240, 110167. URL: https://www.sciencedirect.com/science/article/pii/S1389128623006126, doi:https://doi.org/10.1016/j.comnet.2023.110167.

Li, T., Hui, S., Zhang, S., Wang, H., Zhang, Y., Hui, P., Jin, D., Li, Y., 2024b. Mobile user traffic generation via multi-scale hierarchical gan. ACM Transactions on Knowledge Discovery from Data 18, 1–19.

Lin, Z., Jain, A., Wang, C., Fanti, G., Sekar, V., 2020. Using gans for sharing networked time series data: Challenges, initial promise, and open questions, in: Proceedings of the ACM Internet Measurement Conference, Association for Computing Machinery, New York, NY, USA. p. 464–483. URL: https://doi.org/10.1145/3419394.3423643, doi:10.1145/3419394.3423643.

Pandey, C., Tiwari, V., Rodrigues, J.J., Roy, D.S., 2024. 5gt-gan-net: Internet traffic data forecasting with supervised loss based synthetic data over 5g. IEEE Transactions on Mobile Computing 23, 10694–10705.

Pimpinella, A., Giusto, F.D., Redondi, A.E.C., Venturini, L., Pavon, A., 2022a. Forecasting busy-hour downlink traffic in cellular networks, in: ICC 2022 - IEEE International Conference on Communications, pp. 4336–4341. doi:10.1109/ICC45855.2022.9838982.

Pimpinella, A., Redondi, A.E., Pavon, A., Venturini, L., 2022b. Using the (crystal) ball: Forecasting network traffic peaks with football events, in:

GLOBECOM 2022 - 2022 IEEE Global Communications Conference, pp. 4334–4339. doi:10.1109/GLOBECOM48099.2022.10001059.

Saad, W., Bennis, M., Chen, M., 2019. A vision of 6g wireless systems: Applications, trends, technologies, and open research problems. IEEE network 34, 134–142.

Sandvine, 2024. The global internet phenomena report. URL: https://www.sandvine.com/global-internet-phenomena-report-2024.

Tataria, H., Shafi, M., Molisch, A.F., Dohler, M., Sjöland, H., Tufvesson, F., 2021. 6g wireless systems: Vision, requirements, challenges, insights, and opportunities. Proceedings of the IEEE 109, 1166–1199.

Vu, T.H., Kumar Jagatheesaperumal, S., Nguyen, M.D., Van Huynh, N., Kim, S., Pham, Q.V., 2025. Applications of generative ai (gai) for mobile and wireless networking: A survey. IEEE Internet of Things Journal 12, 1266–1290. doi:10.1109/JIOT.2024.3487627.

Wang, X., Lyu, B., Guo, C., Xu, J., Zukerman, M., 2024. A base station sleeping strategy in heterogeneous cellular networks based on user traffic prediction. IEEE Transactions on Green Communications and Networking 8, 134–149. doi:10.1109/TGCN.2023.3324486.

Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., Xu, H., 2021. Time series data augmentation for deep learning: A survey, in: Zhou, Z.H. (Ed.), Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, International Joint Conferences on Artificial Intelligence Organization. pp. 4653–4660. URL: https://doi.org/10.24963/ijcai.2021/631, doi:10.24963/ijcai.2021/631. survey Track.

Witten, I.H., Frank, E., 2002. Data mining: practical machine learning tools and techniques with java implementations. Acm Sigmod Record 31, 76–77.

Zanzi, L., Sciancalepore, V., Garcia-Saavedra, A., Costa-Pérez, X., Agapiou, G., Schotten, H.D., 2020. Arena: A data-driven radio access networks analysis of football events. IEEE Transactions on Network and Service Management 17, 2634–2647. doi:10.1109/TNSM.2020.3032829.

Zhang, C., Patras, P., Haddadi, H., 2019. Deep learning in mobile and wireless networking: A survey. IEEE Communications Surveys Tutorials 21, 2224–2287. doi:10.1109/COMST.2019.2904897.

Zhang, H., Sediq, A.B., Afana, A., Erol-Kantarci, M., 2025. Mobile traffic prediction using llms with efficient in-context demonstration selection. IEEE Transactions on Communications 73, 11170–11185. doi:10.1109/TCOMM.2025.3579703.

Zhang, S., Li, T., Hui, S., Li, G., Liang, Y., Yu, L., Jin, D., Li, Y., 2023. Deep transfer learning for city-scale cellular traffic generation through urban knowledge graph, in: Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining, pp. 4842–4851.