# RAN Energy Consumption Prediction in Network Expansion Scenarios

Andrea Pimpinella*, Alessandro E. C. Redondi†, Luisa Venturini‡, Andrea Pavon‡, Mircea Nitescu††
* *Department of Management, Information and Production Engineering (DIGIP), University of Bergamo*
† *Department of Electronics, Information and Bio-Engineering (DEIB), Politecnico di Milano*
‡ *Vodafone Servizi e Tecnologie, Ivrea, Italy,* †† *Vodafone Romania*
*contact: andrea.pimpinella@unibg.it*

***Abstract*—As mobile data traffic continues to increase and network operators expand their infrastructures, energy consumption in Radio Access Networks (RAN) becomes a critical concern, particularly during network expansion. This paper addresses the problem of predicting RAN energy consumption in network expansion scenarios. We develop and evaluate different forecasting strategies, including per-site and network-wide models, as well as Machine Learning (ML)-based approaches, using real-world data from a LTE network. Our results show that while simple network-wide models perform well when the base station (BS) configurations remain constant, ML models are more effective in scenarios where BS configurations change during network expansion. The insights from this study can help mobile network operators improve energy efficiency by adapting their networks to traffic patterns and expansion processes, supporting both cost management and sustainability goals.**

***Index Terms*—Cellular Networks, Energy Consumption Forecasting, Network Monitoring, Network Management**

## I. Introduction

Worldwide electricity consumption of communication networks increased from 200 TWh in 2017 to 334 TWh by 2022, accounting for over 1.5% of global electricity usage [1], [2]. With mobile data traffic expected to reach 131 Exabytes per month by 2024, network operators face growing operational challenges, particularly in managing energy consumption. Radio access networks (RAN), which account for about 80% of a mobile network's total energy use, are a major contributor to these costs [3], [4].

As mobile traffic continues to surge, operators face the dual challenge of managing rising energy costs while meeting sustainability targets. Predicting the energy consumption of the RAN is critical for Mobile Network Operators (MNOs) for several reasons:

- **Cost Management:** Energy costs form a major part of operational expenses, with the RAN consuming up to 80% of network energy. Accurate energy forecasting helps operators manage costs, optimize budgeting, and identify opportunities for energy savings.
- **Network Optimization:** Predictive models allow operators to adjust network configurations dynamically based on traffic. Base stations can be powered down during low-traffic periods, reducing energy consumption without impacting service quality.
- **Environmental Impact:** Reducing RAN energy use helps lower carbon emissions, allowing operators to meet sustainability goals and implement green strategies more effectively.
- **Capacity Planning:** As 5G expands, predicting energy demands enables efficient planning for network upgrades, helping operators adopt newer, more energy-efficient technologies.
- **Energy-Efficient Design:** Insights into energy consumption patterns support better network design, enabling the use of features like base station sleep modes to enhance energy efficiency.

At the same time, predicting RAN energy consumption is challenging due to several factors: first, significant fluctuations in traffic based on time and user behavior complicate the creation of accurate predictive models. Secondly, the diverse components of the network, such as different BS hardware/software configurations, each have unique energy consumption profiles. Additionally, the integration of various RAN technologies, such as 4G and 5G, adds complexity. Dynamic management strategies, such as load balancing and sleep modes, further complicate accurate forecasting. Finally, frequent upgrades and reconfigurations of the network necessitate continuous updates to predictive models.

### A. Related work

Regarding energy models, Auer et al. [5] present a model dividing BS energy usage into fixed power, which is independent of traffic, and variable power, which scales with traffic load, while Capone et al. [1] provide an operator-focused model that also considers sleep modes, where BSs can reduce energy consumption during periods of low traffic. In the context of advanced network architectures, Khan et al. explore the energy-saving potential of Cloud Radio Access Networks (C-RAN) through dynamic resource allocation [6]. Their work highlights how centralizing Baseband Units (BBUs) enables efficient resource sharing among Remote Radio Heads (RRHs). Further expanding the scope of energy efficiency in advanced RAN, Liang et al. [7] review the integration of Machine Learning (ML) techniques into Open Radio Access Networks (O-RAN). They discuss how ML can optimize resource allocation and improve network performance, leading to substantial energy savings.

### B. Contribution

Differently from the related work, this paper addresses the problem of predicting total RAN energy consumption in the specific scenario of network expansion, i.e., when new BSs are added to an existing deployment. Based on a real dataset of a LTE RAN, we develop several strategies for predicting future total energy consumption, exploring granular per-site approaches, global network-wide approaches, and ML based approaches. The different strategies are tested and evaluated on different network expansion datasets, where the distribution of BS configurations either remains unchanged or changes. Extensive experiments show that in the former case, simple network-wide approaches provide accurate predictions, while in the latter case, ML based approaches are necessary.

The remainder of the paper is structured as it follows: Section I-A briefly reviews related works in the area of RAN energy consumption prediction, while Section I-B clarifies the contributions of this work. Then, Section II details the used dataset and the different forecasting approaches. Section III focuses on the experimental results obtained, while Section IV concludes the work.

## II. Methodology

### A. Dataset

This study leverages a dataset provided by Vodafone, a leading European MNO, which contains hourly-sampled Key Performance Indicators (KPIs) collected from each active eNodeB of the LTE network deployed in Italy in a period of four weeks, from October 7th, 2023, to November 4th, 2023. Each eNodeB is composed of several radio cells, working at different carrier frequencies and providing coverage to different areas (sectors). We focus on two well known and widely used KPIs, specifically i) the downlink traffic volume $v_i(t)$, reporting the amount of bytes downloaded each hour $t$ from each eNodeB $i$, and ii) the energy consumption $e_i(t)$, measuring for each eNodeB the sum over the set of active cells of the energy consumed by base-band and radio front-end units.

We restrict the attention on the LTE network of a medium-sized Italian city consisting of 400 tri-sector eNodeBs, where each sector provides connectivity to cellular users through one or more logically overlapping radio cells that operate at either 700 MHz, 800 MHz, 1400 MHz, 1800 MHz, 2100 MHz, or 2600 MHz. Although eNodeBs may support any combination of such carriers, organized in any number of sectors, in practice a small subset of configurations is used. In particular, in the used dataset, eNodeBs work according to only 10 different configurations, each one characterized by a specific number and type of active frequency layers: Figure 1 illustrates the distribution of the fraction of eNodeBs per each configuration, with respect to the overall number of eNodeBs in the network. In the next Section we comment on how the number and type of radio cells activated in a network site impact on its energy consumption.

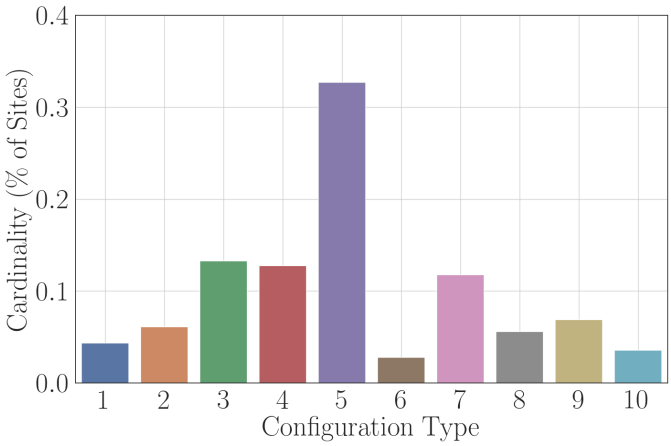


Fig. 1: Distribution of LTE radio access points configuration types in the considered network. As an example, one configuration may correspond to eNodeBs with 3 frequency layers working at 800, 1800 and 2100 MHz.

### B. Energy Consumption and Downlink Traffic

It is widely accepted that the energy consumed by an eNodeB can be modeled as a linear function of the served traffic [1]. Figure 2 (top) illustrates this relationship for 2 eNodeBs taken from configurations 3 and 5, representing 15% and 32% of the network, respectively. eNodeBs within the same configuration exhibit diverse linear behaviors due to factors like antenna transmit power, hardware types, or software versions, complicating per-site energy forecasting. Therefore, it is more practical for MNOs to focus on aggregate energy consumption, combining all eNodeBs in a configuration or across configurations. The bottom of Figure 2 shows energy consumption versus downlink traffic load, as result of the summation of data samples over three set of eNodeBs per each hour of operation. Each considered set always contains 100 eNodeBs taken from selected groups of configurations, precisely: i) blue samples represent a network consisting of 10 eNodeBs sampled by each configuration (i.e., configurations are equally distributed); ii) green samples correspond to 20 eNodeBs per configuration type in the set $\{2, 4, 7, 8, 9\}$; iii) orange samples correspond to 50 eNodeBs taken from both configurations 3 and 5. As one can see, regardless of the distribution of configurations, the relation between downlink traffic and energy consumption fits consistently a linear curve. However, while the slope of the lines (i.e., the energy cost of each transmitted byte) is weakly affected by the distribution of configurations, the values of the offset do depend on such distribution. Moreover, fixing the distribution, we observe that different eNodeBs have different offset consumption: these differences highlight the preference for global forecasting models over per-site ones.

### C. Energy Consumption Forecasting Approaches

We focus on scenarios of network expansions, i.e., where the MNO either deploys new network sites (e.g., to enhance

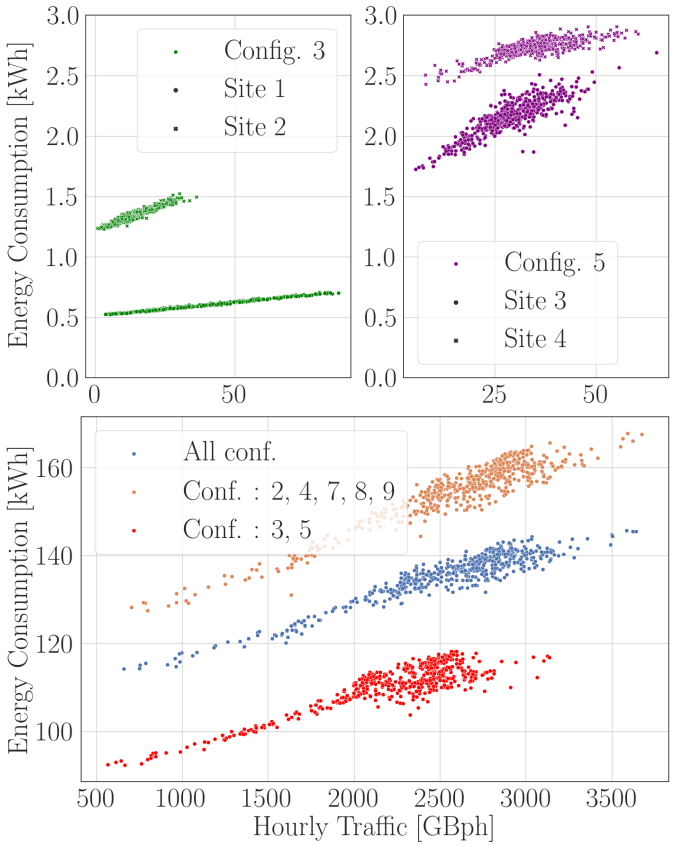


Fig. 2: Top: Scatterplot of downlink traffic [GBph] served vs energy [kWh] consumed by 2 eNodeBs configured as either type 3 (green) and 5 (purple) taken as example, respectively. Bottom: Scatterplot of sum downlink traffic [GBph] vs sum energy consumption [kWh], for 3 different sets of 100 eNodeBs.

coverage), installs new frequency layers on top of already deployed eNodeBs (e.g., to boost capacity), or both. To mimic such scenarios, we let the size of the network increase from a *starting* set of $\mathcal{N}_\text{s}$ eNodeBs to a *final* set of $\mathcal{N}_\text{f} > \mathcal{N}_\text{s}$ eNodeBs, that corresponds to a network expansion rate $\eta$ equal to $(\mathcal{N}_\text{f}-\mathcal{N}_\text{s})/\mathcal{N}_\text{s}$. The starting set will be used to train the forecasting approaches, while performance will be evaluated on the final set. Note that the deployment of new eNodeBs at forecasting time might change the distribution of sites' configurations observed during the training phase. To control this possibility, we define with $\mathbf{c}_\text{s}$ and $\mathbf{c}_\text{f}$ the distributions observed at training and forecasting time respectively, where $\mathbf{c}_\text{s}$ and $\mathbf{c}_\text{f}$ are vectors of 10 entries and each entry $c_j$ equals the fraction of eNodeBs configured as type $j$. Regarding the forecasting approaches, we use solely methods based on linear regression, considering the linear relationship between traffic load and energy consumption. However, we explore several approaches characterized by different complexity and feature sets. For what concerns complexity, we focus on baseline models, obtained taking either a per-site or a network-wise perspective, and on ML based models, where a set of features is fed into a regressor to output the target. Regarding the feature sets, we focus on two cases:

1) A baseline case where only traffic load is considered as input feature,
2) An advanced scenario where $\mathcal{N}_\text{s}$, $\mathcal{N}_\text{f}$, $\mathbf{c}_\text{s}$ and $\mathbf{c}_\text{f}$ are also considered for prediction in addition to traffic load.

Note that while traffic load is known at training phase, it is not known exactly at forecasting time. A possible option is to use one of the many cellular network traffic predictors available in literature [8]–[11], predict the future network traffic load and finally consider it as input to forecast energy consumption. Here we opt for an oracle, i.e., we forecast energy consumption taking as input the ground truth traffic load observed at forecasting time. We leave the analysis on the interplay between possible traffic load forecasting errors and energy consumption prediction to future work.

In the following, we introduce the forecasting approaches and provide details on how they are trained and then used to forecast energy consumption.

*1) Site:* this model estimates the overall energy consumed by a cellular network $E(t)$ at hour $t$ as the sum of the hourly consumption of the single eNodeBs. First, at training phase, we learn the per-site relationship between the downlink traffic load $v_{\text{s},i}(t)$ and energy $E_{\text{s},i}(t)$, summarized for each eNodeB $i$ as:

$$\hat{E}_{\text{s},i}(t) = \theta \cdot v_{\text{s},i}(t) + \phi, \ \forall i = 0 \dots \mathcal{N}_\text{s} \tag{1}$$

Secondly, at forecasting phase, we use the parameters $\theta$ and $\phi$ learnt during training to output $E_\text{f}(t)$ as:

$$\hat{E}_\text{f}(t) = \sum_i^{\mathcal{N}_\text{f}} \hat{E}_{\text{f},i}(t) \tag{2}$$

Note that $\theta$ and $\phi$ do not depend on $i$, and represent the average single-site relationship between traffic load and energy consumption, regardless of the site's configuration. Also, this approach does not take as input neither the size of the network, nor the distribution of configurations: for these reasons, no retraining is needed after the expansion of the network.

*2) Network:* in this case, $E(t)$ is directly estimated as linear function of the network-wise, hourly sum of downlink traffic $V(t) = \sum_i v_i(t)$. At training phase, the model is trained on the starting set of $\mathcal{N}_\text{s}$ eNodeBs according to the following relation:

$$\hat{E}_\text{s}(t) = \Theta \cdot \sum_i^{\mathcal{N}_\text{s}} v_{\text{s},i}(t) + \Phi = \Theta \cdot V_\text{s}(t) + \Phi \tag{3}$$

Then, parameters $\Theta$ and $\Phi$ are used at forecasting time to output the energy consumed by the (expanded) network as it follows:

$$\hat{E}_\text{f}(t) = \Theta \cdot V_\text{f}(t) + (1+\eta) \cdot \Phi \tag{4}$$

where $\eta$ is the expansion rate of the network, and $V_\text{f}(t) = \sum_i^{\mathcal{N}_\text{f}} v_i(t)$, i.e., the hourly sum of the downlink traffic with

respect to the final set of eNodeBs. Also in this case, we do not need to retrain the model after the expansion: the correcting factor $(1+\eta)$ is used to take into account the additional energy overhead due to the deployment of new network sites.

*3) Machine Learning (ML):* this model takes a ML-like approach and estimates $E(t)$ as it follows. First, $V_\mathrm{s}(t)$, $\mathcal{N}_\mathrm{s}$ and $\mathbf{c}_\mathrm{s}$ are used to train a regressor $\mathcal{L}(\cdot)$, as:

$$\hat{E}_\mathrm{s}(t) = \mathcal{L}\left(V_\mathrm{s}(t), \mathcal{N}_\mathrm{s}, \mathbf{c}_\mathrm{s}\right) \tag{5}$$

Then, $\mathcal{L}(\cdot)$ is used to predict the energy $E_\mathrm{f}(t)$ consumed hourly by the expanded network taking as inputs $V_\mathrm{f}(t)$, $\mathcal{N}_\mathrm{f}$ and $\mathbf{c}_\mathrm{f}$:

$$\hat{E}_\mathrm{f}(t) = \mathcal{L}\left(V_\mathrm{f}(t), \mathcal{N}_\mathrm{f}, \mathbf{c}_\mathrm{f}\right) \tag{6}$$

Note that to achieve high forecasting performance and increase model's robustness to network expansions, it is crucial to let the model train on as many different network instances as possible, characterized by different combinations of $\mathcal{N}_\mathrm{f}$ and $\mathbf{c}_\mathrm{f}$. Details on this aspect are provided in Section III-B.

As a final remark, note that the *Site* approach requires traffic load predictions per each eNodeB. In case of network expansion, therefore, such an approach may be subject to additional noise caused by the change in traffic load at each eNodeB caused by the expansion itself. Conversely, *Network* and *ML* approaches are resilient to this issue by design, as they require as input the total traffic load.

## III. Experiments and Results

To train and evaluate the performance of our models, we consider strictly consecutive training and testing periods, using data collected from October 7th to October 27th, 2023, for training (i.e., 3 weeks) and leaving the period from October 28th to November 4th, 2023, to testing (i.e., 1 week). As performance metric, we use the Mean Absolute Performance Error (MAPE), defined as the average relative error between ground truth $E(t)$ and predicted $\hat{E}(t)$ energy consumption:

$$\mathrm{MAPE} = \frac{1}{|T|}\sum_{t\in T}\frac{E(t)-\hat{E}(t)}{E(t)}, \tag{7}$$

where $T$ is the set of test hours and $|T| = 168$, i.e., the number of hourly samples in the test week.

We perform several experiments considering the interplay of two challenges arising from network expansion: i) the variation of the number of eNodeBs in the network (i.e., from $\mathcal{N}_\mathrm{s}$ to $\mathcal{N}_\mathrm{f}$), and ii) the variation of the distribution of their configurations (i.e., from $\mathbf{c}_\mathrm{s}$ to $\mathbf{c}_\mathrm{f}$), from training to testing periods. According to this idea, we recognize two different experimental scenarios:

- **S1**: in this scenario, the distribution of site type configurations remains unchanged (i.e., $\mathbf{c}_\mathrm{f} \sim \mathbf{c}_\mathrm{s}$), while we test independently different cases where the number of nodes $\mathcal{N}_\mathrm{f}$ is set equal to $\mathcal{N}_\mathrm{s} + \eta \cdot \mathcal{N}_\mathrm{s}$, being $\eta$ an expansion factor that takes a value from 0 (i.e., $\mathcal{N}_\mathrm{f} = \mathcal{N}_\mathrm{s}$) to 100% (i.e., $\mathcal{N}_\mathrm{f} = 2\mathcal{N}_\mathrm{s}$);

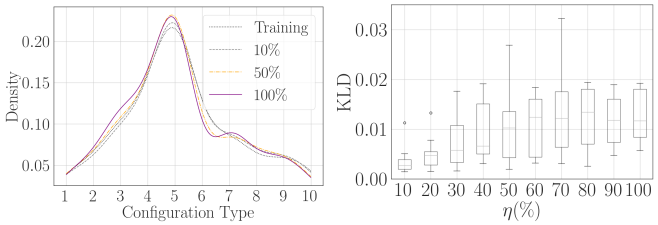


Fig. 3: Left: Training (black, dotted) and test distribution of configurations, for $\eta$ equal to 25% (grey, dotted), 50% (yellow, dash-dotted), and 100% (purple). Right: Boxplot of KLD between training and test distribution of configurations, for $\eta$ ranging from 10% to 100%.

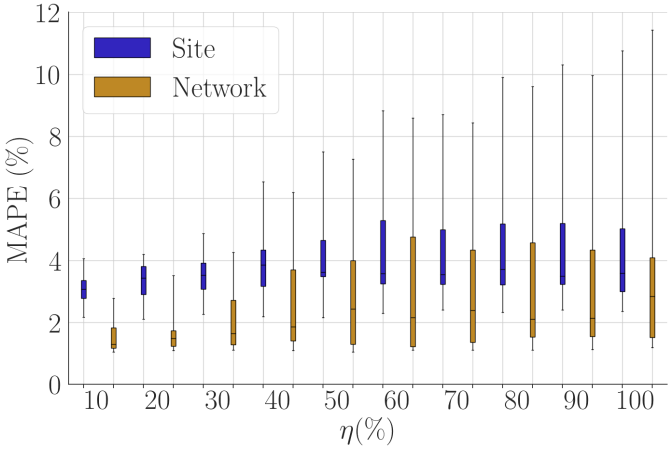


Fig. 4: Boxplots of MAPE of Site (blue) and Network (orange) approaches, with respect to 100 different combinations of training and side sets drawings.

- **S2**: in this case, we let $\mathcal{N}_\mathrm{f}$ vary as in S1, but also we vary the distribution $\mathbf{c}_\mathrm{s}$, such that $\mathbf{c}_\mathrm{f} \neq \mathbf{c}_\mathrm{s}$. Note that in this scenario we impose that $c_{j,s} > 0, \forall j \in 1, \ldots, 10$, i.e., all the configurations are available for training and also no new configuration is added during the testing phase.

In the following Sections, we show the results obtained from each experimental scenario.

### A. Scenario S1

This scenario involves a MNO deploying new eNodeBs without significantly changing the configuration distribution. For the experiment, 200 eNodeBs are randomly selected from 400 sites for training, leaving the remaining 200 as *side* set $\mathcal{S}$. The configuration distribution of the training set mirrors that of the real network (Figure 1). During testing, 10 iterations are run, each adding 10% (20 eNodeBs) from the side set to the training set. For example, iteration 1 evaluates forecasting on 220 eNodeBs—200 from training and 20 added during testing.

Figure 3 compares the eNodeB configuration distributions during training and testing at 10%, 50%, and 100% new eNodeBs. The distributions remain similar, as confirmed by Kullback–Leibler Divergence (KLD) results, with distances below 0.03.

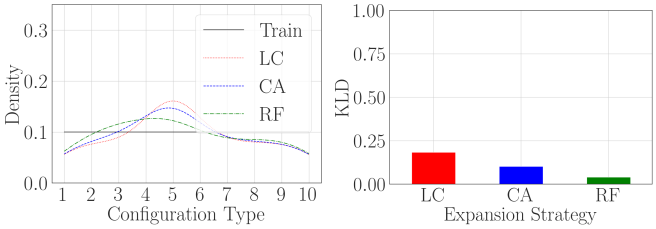


Fig. 5: Left: Training (black) and test distribution of configurations, according to expansion strategies LC (red, dotted), CA (blue, dashed), and RF (green, dash-dotted), for $\eta = 25\%$. Right: Average KLD between training and test distributions of configurations.

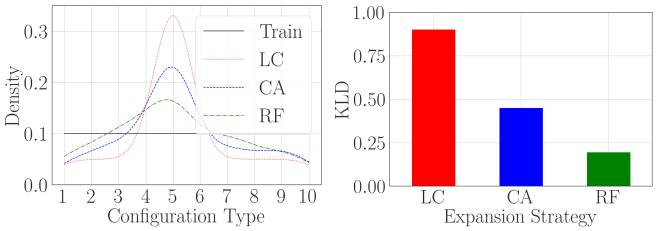


Fig. 6: Left: Training (black) and test distribution of configurations, according to expansion strategies LC (red, dotted), CA (blue, dashed), and RF (green, dash-dotted), for $\eta = 100\%$. Right: Average KLD between training and test distributions of configurations.

Figure 4 shows the boxplots of the MAPEs obtained with the *Site* and *Network* approaches. Note that we do not consider the *ML* model for this scenario, as the distributions of configurations do not change. As shown, 75-th percentile of MAPEs are below 5.5%: when $\mathbf{c}_\mathrm{f}$ is close to $\mathbf{c}_\mathrm{s}$, it is relatively easy for MNOs to predict the energy consumption of the network given the traffic load, even if the network expands in size (i.e., $\eta > 0$), with no need of retraining. Also, we observe that *Network* (blue) outperforms *Site* (orange), reducing on average 50-th and 75-th percentile errors by 50% and 30% respectively: forecasting approaches based on data aggregated at the network level are preferable in static scenarios.

### B. Scenario S2

In this scenario, the distribution of eNodeBs' configurations at training time is different than the one observed during testing. In addition to the *Site* and *Network* approaches presented before, here we also rely on a ML-based approach, training a linear regressor which explicitly takes in consideration the distribution of the configurations. For the training phase, we randomly select a set $\mathcal{N}_\mathrm{p}$ of 100 sites from the original set of 400 sites and create a new dataset with the goal of showing to the ML model the greatest number of distributions. We design the training dataset according to a process lasting 1000 iterations. At the $k$-th iteration, the following steps are performed:

1) draw a vector $\mathbf{c}_\mathrm{k}$ composed of 10 random weights, strictly taking values between 0 and 1, representing a specific distribution of configurations;
2) from $\mathcal{N}_\mathrm{p}$, draw a random number $\mathcal{N}_\mathrm{k} \leq 100$ eNodeBs whose configurations are distributed according to $\mathbf{c}_\mathrm{k}$;
3) generate a new network instance composed by the eNodeBs selected at step 2, computing the corresponding ground truth total energy consumption $E_\mathrm{k}(t) = \sum_i^{\mathcal{N}_\mathrm{k}} E_i(t)$, for all time samples in training set (i.e., 504 samples);
4) Add the observations $(E_\mathrm{k}(t), \sum_i^{\mathcal{N}_k} v_i(t), \mathcal{N}_\mathrm{k}, \mathbf{c}_\mathrm{k})$ to the training set.

Finally, we train the linear regressor $\mathcal{L}(\cdot)$ so that $E_\mathrm{k}(t) = \mathcal{L}(\sum_i^{\mathcal{N}_k} v_i(t), \mathcal{N}_\mathrm{k}, \mathbf{c}_\mathrm{k})$. For the testing phase we consider the set of remaining 300 sites and we randomly pick $\mathcal{N}_\mathrm{s} = 100$ eNodeBs from that, such to make each configuration appear exactly 10 times. We use this instance to train the *Site* and *Network* approaches, as done in III-A. The remaining 200 sites are left to the side set $\mathcal{S}$ and used for simulating network expansion. This is performed with a controlled drawing of eNodeBs from $\mathcal{S}$, in order to change the distribution of configurations with respect to $\mathcal{N}_\mathrm{s}$. We perform such operation according to 3 different strategies:

- Largest Configuration (**LC**): draw side eNodeBs only from conf. 5, i.e., the largest configuration in the real network;
- Cardinality (**CA**): draw side eNodeBs with weights proportional to the density of configurations in the real network;
- Random Full (**RF**): randomly draw eNodeBs from the side set;

The three different strategies mimic different expansions of the network, and all correspond to a final distribution of configuration different from the starting one. The left sides of Figures 5 and 6 plot the distributions obtained at testing time after expanding the network according to each of the three mentioned strategies, with respect to the cases of $\eta = 25\%$ and $\eta = 100\%$. Also, we plot on the right side the corresponding values of KLD. As one can see, test distributions substantially differ from the training time distribution (corresponding to the black line in both left plots), and distances increase with $\eta$. In fact, when $\eta = 25\%$ (i.e., $\mathcal{N}_\mathrm{f} = 125$ sites), average distances between $\mathbf{c}_\mathrm{s}$ and $\mathbf{c}_\mathrm{f}$ are between 0.06 (RF) and 0.23 (LC), i.e., from 2 to 7 times larger than the largest distance observed in scenario S1, while they further increase by a factor of 4 when $\eta$ raises to $100\%$ (i.e., $\mathcal{N}_\mathrm{f} = 200$ sites).

We observe that both large values of $\eta$ (i.e., increase of network size) and large distances between $\mathbf{c}_\mathrm{s}$ and $\mathbf{c}_\mathrm{f}$ map to a decrease of forecasting performance. To illustrate, we plot in Figures 7 and 8 the boxplots of MAPEs obtained after using Site (orange), Network (blue) and ML (green) approaches to forecast energy consumption. Referring to Figure 7, that corresponds to $\eta = 25\%$, Network performs better than Site, providing on average 2.2% lower median MAPE. Moreover, ML model outperforms both Site and Network for all strategies, yielding average 50-th and 75-th percentiles of MAPE below 3% and 4%, respectively. This can be explained considering that the distribution of eNodeBs' configurations in the network is an input of ML model, which makes it robust to changes due to network expansion processes. A network size

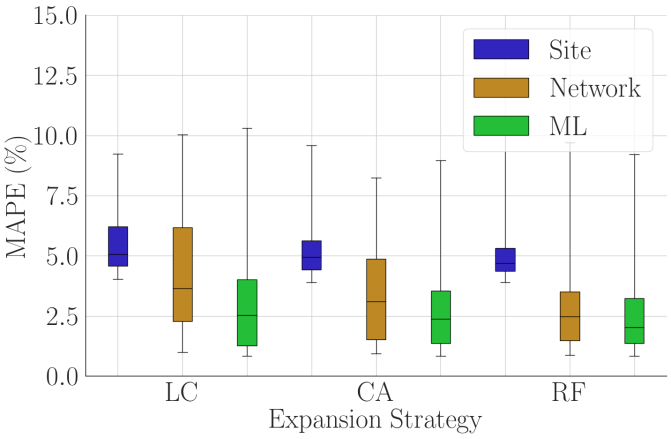


Fig. 7: Boxplots of MAPE for the different approaches and expansion strategies, with $\eta = 25\%$ and 100 different combinations of training and side sets drawings.

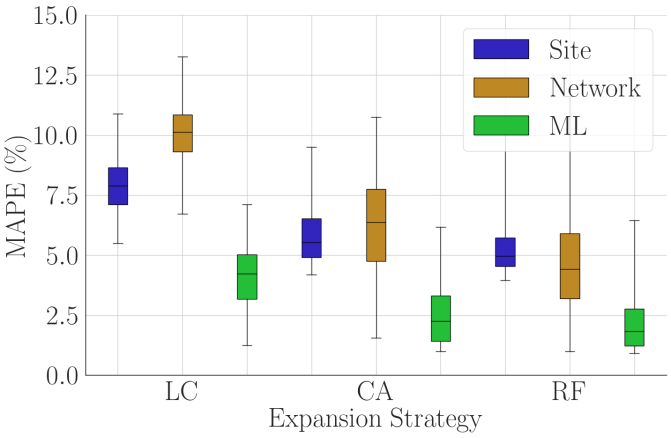


Fig. 8: Boxplots of MAPE for the different approaches and expansion strategies, with $\eta = 100\%$ and 100 different combinations of training and side sets drawings.

increase is also detrimental for performance if the model is not robust to generalize over new scenarios: as one can see in Figure 8, a four-fold increase of $\eta$ leads to a general increase of forecasting errors for Site and Network models, corresponding to average losses of 1.33 and 3.5 percentage points on MAPE, respectively, while the impact on ML approach is rather negligible (-0.5%). This confirms the overall robustness of model ML, that besides the distribution of configurations includes as feature also the size $\mathcal{N}_\mathrm{f}$ of the network. As final remark for this scenario, we observe that with the increase of $\eta$, and regardless of the forecasting approach, the larger the value of KLD between training and test distributions, the worse the effect on performance: strategy LC, corresponding to a KLD of 0.87 when $\eta = 100\%$, generates the highest loss of performance, equal to 3.2%, 6.3% and 1.5% higher median MAPE for Site, Network and ML approach, respectively.

## IV. Conclusions

This paper addresses the challenge of forecasting RAN energy consumption in network expansion scenarios. We examined various forecasting approaches, focusing on scenarios where the expansion either preserves or alters the eNodeB configuration distribution. In the former case, we demonstrated that a simple model trained on aggregate energy consumption, adjusted with a correction factor for the number of added sites, delivers good performance (below 5.5% MAPE at 75-th percentile), and is preferable over atomic approaches that take per-site perspectives. However, when the distribution of configurations changes, we found that a machine learning model explicitly considering such distribution during training achieves the best results. Future works will explore scenarios where the deployment of new sites includes previously unseen configurations as well as analyze the impact of imperfect traffic load predictions.

## V. Acknowledgments

This work was supported in part by project *SERICS* (PE00000014) under the NRRP MUR program funded by the EU - NGEU, and PRIN Project *NEWTON* (2022ZA8T22).

## References

[1] A. Capone, S. D'Elia, I. Filippini, A. E. Redondi, and M. Zangani, "Modeling energy consumption of mobile radio networks: An operator perspective," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 120–126, 2017.

[2] D. Lundén, J. Malmodin, P. Bergmark, and N. Lövehagen, "Electricity consumption and operational carbon emissions of european telecom network operators," *Sustainability*, vol. 14, no. 5, p. 2637, 2022.

[3] T. Saraiva, D. Duarte, I. Pinto, and P. Vieira, "An improved bbu/rru energy consumption predictor for 4g and legacy mobile networks using mixed statistical models," in *2020 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2020, pp. 320–325.

[4] J. Malmodin, N. Lövehagen, P. Bergmark, and D. Lundén, "Ict sector electricity consumption and greenhouse gas emissions–2020 outcome," *Telecommunications Policy*, vol. 48, no. 3, p. 102701, 2024.

[5] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume *et al.*, "How much energy is needed to run a wireless network?" *IEEE wireless communications*, vol. 18, no. 5, pp. 40–49, 2011.

[6] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, "Reducing energy consumption by dynamic resource allocation in c-ran," in *2015 European Conference on Networks and Communications (EuCNC)*. IEEE, 2015, pp. 169–174.

[7] X. Liang, Q. Wang, A. Al-Tahmeesschi, S. Chetty, D. Grace, and H. Ahmadi, "Energy consumption of machine learning enhanced open ran: A comprehensive review," *IEEE Access*, 2024.

[8] S. Dawoud, A. Uzun, S. Göndör, and A. Küpper, "Optimizing the power consumption of mobile networks based on traffic prediction," in *2014 IEEE 38th Annual Computer Software and Applications Conference*. IEEE, 2014, pp. 279–288.

[9] S. Zhang, S. Zhao, M. Yuan, J. Zeng, J. Yao, M. R. Lyu, and I. King, "Traffic prediction based power saving in cellular networks: A machine learning method," in *Proceedings of the 25th ACM SIGSPATIAL international conference on advances in geographic information systems*, 2017, pp. 1–10.

[10] A. Azari, F. Salehi, P. Papapetrou, and C. Cavdar, "Energy and resource efficiency by user traffic prediction and classification in cellular networks," *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 2, pp. 1082–1095, 2021.

[11] A. Pimpinella, F. D. Giusto, A. E. C. Redondi, L. Venturini, and A. Pavon, "Forecasting busy-hour downlink traffic in cellular networks," in *ICC 2022 - IEEE International Conference on Communications*, 2022, pp. 4336–4341.