Premsankar, Gopika; Piao, Guangyuan; Nicholson, Patrick K.; Di Francesco, Mario; Lugones, Diego

Data-driven Energy Conservation in Cellular Networks: A Systems Approach

Published in:
IEEE Transactions on Network and Service Management

DOI:
10.1109/TNSM.2021.3083073

Published: 01/09/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Premsankar, G., Piao, G., Nicholson, P. K., Di Francesco, M., & Lugones, D. (2021). Data-driven Energy Conservation in Cellular Networks: A Systems Approach. *IEEE Transactions on Network and Service Management*, *18*(3), 3567-3582. Article 9439539. https://doi.org/10.1109/TNSM.2021.3083073

Data-driven Energy Conservation in Cellular Networks: A Systems Approach

Gopika Premsankar, Guangyuan Piao, Patrick K. Nicholson, Mario Di Francesco, Diego Lugones

***Abstract*—The energy consumption of mobile networks is already substantial nowadays, and only expected to further increase with the roll-out of 5G. Base stations are the key elements in this context: reducing their energy consumption is of paramount importance for network operators, not only to lower operating costs, but also to meet sustainable development goals. Today's base stations are typically over-provisioned, i.e., they comprise multiple cells to meet the peak load in a region. Therefore, substantial energy savings are possible by switching off cells that are under-utilized. This article proposes a data-driven approach to determine the time periods when a cell can be switched off. Forecasting is used to accurately predict network utilization and automatically find the time intervals to reliably switch off a cell. We carefully analyze the requirements of the system as a whole, from data collection to forecasting methods, to enable effective energy savings in practice. Considering several real-world traces from LTE networks, we show that an average of 10.24% energy savings is possible. We explore the trade-offs between energy savings and overhead in switching off cells, and provide insights into the choice of methods accordingly. In particular, we show that the accuracy of forecasting is not the most important factor in achieving energy savings; instead, the prediction (uncertainty) interval plays a key role in being able to achieve energy savings with less impact on end-users. Finally, we propose a model to generate utilization traces that match the distribution of real-world traces obtained from cellular networks.**

***Index Terms*—Energy savings, LTE, forecasting, time series**

I. Introduction

With eight billion mobile subscriptions worldwide [1], the market growth of the telecommunication industry has been accompanied by a substantial increase in the energy consumption of mobile networks. Cellular base stations are the key elements in this context, as they consume 80% of the energy expenditure in operating mobile networks [2, 3]. Besides, more and more base stations are being deployed in 5G to support a 1,000-fold increase in traffic, while simultaneously aiming at a 50% reduction in energy consumption [4]. Consequently, mobile networks are required to increase their energy efficiency by several orders of magnitude over the coming years. Indeed, a lower energy consumption is of paramount importance for cellular network operators already now [5], not only to reduce operating costs, but also to meet sustainable development goals [6, 7].

However, today's cellular networks are typically over-dimensioned: cells are deployed to meet the peak demands — also known as the *busy hour*. As a consequence, significant energy savings can be achieved by switching off cells during periods of low utilization. Base stations support switching off under-utilized cells during pre-configured time periods [8, 9]. Unfortunately, in practice, it is very challenging to determine the time periods when a cell can be switched off. Traffic patterns significantly differ from cell to cell and also vary over time. Figure 1 illustrates these two issues based on a Long-Term Evolution (LTE) traffic trace of a mobile network operator. We first plot the CDF of the durations (per day) when the utilization is low enough for a cell to be switched off for two representative cells. The durations when the cell can be switched off show variations between weeks in one case (Figure 1a), whereas it is more consistent in the other (Figure 1b). However, a closer look at the actual time intervals when the latter cell can be switched off (marked as light bands in Figure 1c) reveals that the utilization patterns significantly change between weeks. For instance, the opportunities for energy savings on Wednesday and Saturday in the first week are very different from those in the last week. Thus, in practice, it is not at all trivial to determine (in advance) the time periods when a cell can be switched off, especially given current best practices: network operators typically determine such periods based on the time of the day or simple thresholds [9]. Clearly, such an approach is error-prone and not scalable as the number of cells increases.

Recent advances in both machine learning (ML) and communication technologies offer the opportunity to solve such a challenge [10]. On the one hand, deep neural networks as well as novel techniques for time series forecasting have gained increasing accuracy [11], in addition to the ability to characterize the uncertainty of predictions [12]. On the other hand, modern cellular networks (i.e., 4G and 5G) allow for significant flexibility in managing radio resources through software components running in virtualized environments, for instance, at the edge of the access network [10, 13]. However, the opportunities brought forward by these developments come with a cost: the resulting system is complex and software-based solutions need to be carefully designed by taking a holistic approach into account to avoid issues.

In this context, we propose a **data-driven approach** to determine the time periods during which an under-utilized cell

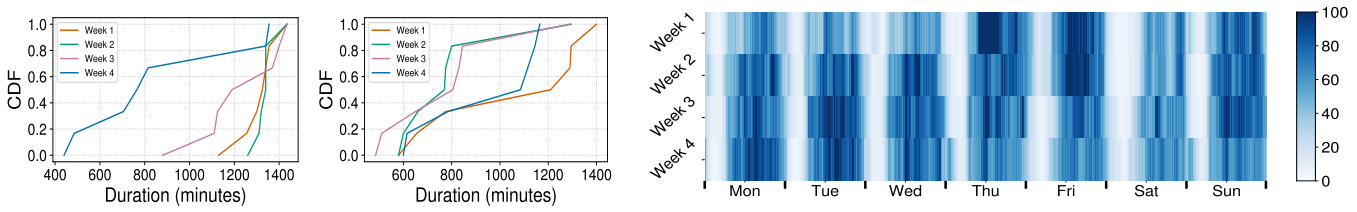
Gopika Premsankar and Guangyuan Piao contributed equally to this work. Part of this work was carried out when they were at Nokia Bell Labs in Dublin, Ireland. This work was partially supported by the Academy of Finland under grants number 319710, 326346, and 332307.

G. Premsankar is with Ivey Business School, Western University in London, ON, Canada. E-mail: gpremsankar@ivey.ca

G. Piao is with Maynooth University in Kildare, Ireland. E-mail: Guangyuan.Piao@mu.ie

P. K. Nicholson and D. Lugones are with Nokia Bell Labs in Dublin, Ireland. E-mail: {pat.nicholson, diego.lugones}@nokia-bell-labs.com

M. Di Francesco is with the Department of Computer Science, Aalto University in Espoo, Finland. E-mail: mario.di.francesco@aalto.fi



(a) (b) (c)

Fig. 1: Cumulative distribution function of the duration of possible off periods for a cell with (a) high and (b) low variance in resource utilization. (c) Heatmap showing the resource utilization of a cell over days in a week for different weeks.

can be switched off. Forecasting is used to accurately predict network utilization and automatically find the time intervals to reliably switch off a cell (Section II). Such an approach does not require global information on the network, but only of the utilization metrics for co-located cells served by a single base station. As a consequence, it scales to tens of thousands of cells and can easily be adopted by network operators. We carefully analyze the requirements of *the system as a whole* – from data collection to forecasting methods – to enable effective energy savings in practice and evaluate the proposed solution based on data from a real LTE network. To the best of our knowledge, this is the first comprehensive study on the energy savings that can be achieved by switching off under-utilized cells (Section VII). In fact, the state-of-the-art has so far addressed switching off *entire* base stations [14–16] as opposed to individual cells *within* a base station. Moreover, existing solutions in the literature have been evaluated through simulations in which traffic is assumed to follow a statistical distribution [17–19], perfectly known in advance [20] or predicted without taking the related uncertainty into account [11, 21].

The main contributions of this article are as follows. First, we **establish the requirements for accurate data collection of utilization data** in the context of power management at cellular base stations (Section III). Specifically, we evaluate the suitability of utilization data with different granularities for forecasting purposes. Moreover, we propose a model to expand utilization traces with coarse granularity so as to match the distribution of real-world data from cellular networks at a millisecond timescale. Second, we **analyze the accuracy as well as fitting time of several forecasting methods** (Section IV). We consider different options to decide when to switch off cells for energy conservation accordingly. We also define control strategies to overcome inaccurate predictions or unexpected events during power management. Finally, we **carry out an extensive evaluation of energy savings at LTE base stations** over a large and diverse set of utilization patterns (Section V). Specifically, we characterize utilization traces from a real operator across different dimensions to understand which methods work best in specific scenarios. We explore the trade-offs between energy savings and overhead in switching off cells, and provide insights into the choice of methods accordingly.

Our results show that switching off decisions have to be tailored to the characteristics of the traffic in considered cells, depending on the key performance metrics targeted by mobile operators – for instance, the number of user migrations (Section VI). However, this is not achieved by simply using the most accurate forecaster; instead, the uncertainty interval is critical in busy cells to able to achieve energy savings with less impact on end-users. Our results indicate that an average of 10.24% percent of energy savings is possible within a base station when the capacity cell is switched off according to the predicted traffic. This represents annual savings of approximately USD 87.6 million savings in China alone[1] with 3 million base stations [22].

II. Saving Energy in Cellular Networks

This section introduces a reference architecture for reducing energy consumption of cellular networks. It also details the components for planning and implementing power management policies at a base station. Last, it introduces a novel approach for power management at cellular base stations.

A. System architecture

The reference architecture follows the principles defined by the O-RAN Alliance [13] to enable ML-powered functionalities in modern cellular networks and illustrated in Figure 2.

The **base station** is the element of the cellular network handling physical radio resources, and consequently, the main source of energy consumption in the network. The base station provides utilization-related metrics in real-time and accepts power management commands, such as switching on or off the hardware associated with certain radio resources (e.g., radio channels). Metrics are collected at a **control** module which processes and stores them into a *radio network database* for later use. The controller uses such metrics to execute a power-management policy and to address the related consequences (e.g., load balancing and allocation of radio resources). A **planning** module predicts utilization based on historical data with a certain granularity as available in the radio network database. In particular, it obtains a *forecast* according to certain methods – such as those discussed in Section IV – which could also employ ML, particularly, neural networks. Predicted utilization is leveraged to define *power management policies* applied to the base station. Such policies include two components: an *energy savings plan*, describing the time periods (in a future horizon) during which radio resources can

[1]Estimated with a cost of 0.08 USD/kWh according to https://www.china-briefing.com/news/china-electricity-prices-industrial-consumers/

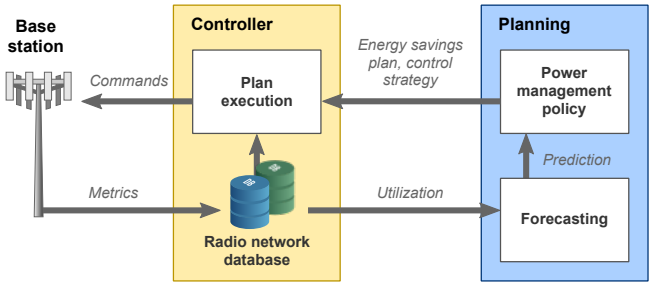


Fig. 2: Overview of an energy saving system for cellular networks based on forecasting. Several metrics are collected from the base station and also stored in a radio network database. Utilization data are sent to a forecasting module and the predicted values are leveraged to define power management policies. Such policies are then applied by issuing commands to switch on (off) radio resources, while still controlling the actual utilization in real-time to ensure a target utilization level.

be switched off; and a *control strategy* applied while executing power management to ensure a target level of utilization.

The control module operates in real-time and is located at the base station or at the edge of the access network. The planning module, instead, operates over longer timescales and is located at the core network or at a geographically-close cloud data center [13]. The technical aspects of the operations associated with these modules are detailed in the next two sections.

B. Power management approach

Cellular base stations provide network connectivity with a layered architecture of overlapping cells that differ in capacity and coverage [21]. In practice, a single base station comprises different types of co-located cells (i.e., layers): a *coverage cell* operating at lower frequencies over a large coverage area, and one or more *capacity cells* operating at higher frequencies to provide higher data rates to users in closer proximity (Figure 3a). Cells with such a *coverage-capacity relationship* are indeed very common in current LTE networks and are also an important element of 5G deployments [23].

Accordingly, our approach to power management explicitly targets coverage-capacity cells at a single base station. In particular, it switches off a capacity cell when its utilization is low, as long as the users eventually served by that cell can be accommodated by the associated coverage cell. Figure 3b shows a sample allocation of users over time according to the proposed approach. Initially, the capacity cell serves four users and the coverage cell serves only one user. At time t_1, two users leave the capacity cell and the related utilization become low. These two users can actually be served by the coverage cell, in addition to the existing user (already connected to the coverage cell). Thus, the capacity cell is switched off and the two users moved to the coverage cell. At time t_2, one more user enters the coverage cell, raising the utilization over an acceptable threshold. As a consequence, the capacity cell is switched on again, and two users are moved from the coverage to the capacity cell. Note that the proposed approach relies on the coverage cell being always[2] on, as the related area cannot otherwise be served.

Our approach is *scalable* for different reasons. First, it only operates at individual base stations, thus, it does not require complex coordination across the cellular network. In contrast, solutions that switch on or off entire base stations require exchanging messages to coordinate decisions between different base stations [17], which may incur a substantial overhead in large networks. Second, our solution only requires identifying base stations with a coverage-capacity relationship beforehand. The related information is readily available to network operators and does not rely on special tools or complex pre-processing of possibly large radio network maps.

III. Data-driven power management

Data-driven power management requires collection and processing of utilization metrics. Accordingly, the rest of this section discusses metric selection and the choice of a suitable granularity for forecasting purposes. The discussion concludes with a method to generate accurate, fine-grained utilization from data available at a coarser granularity.

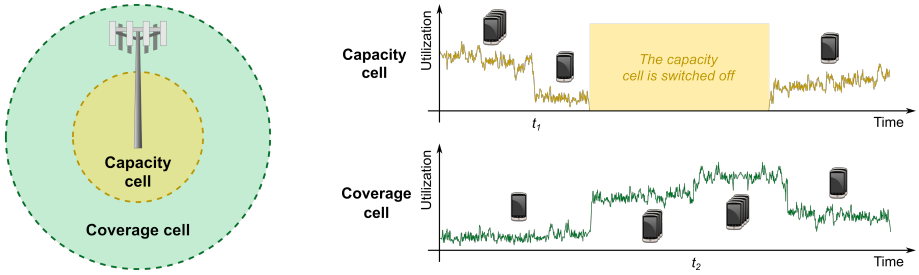
A. Metrics and granularity

The previous section has simply referred to utilization, even though there are different ways to characterize such in practice. Finding a suitable metric is particularly important, so that network operators can meaningfully specify when a cell should be switched off [24]. For this reason, the widely-used *physical resource block* (PRB) utilization [25, 26] is considered next – the PRB is the smallest unit of frequency and time allocated to a user in a cell.

The joint utilization history of a capacity-coverage pair can be seen as a time series, thus, it is possible to use forecasting methods to predict future utilization. Before choosing a forecasting method (Section IV), we must first establish the input data and granularity to be used. We note that the utilization of individual cells is not suitable for forecasting, as the related utilization goes to zero for the period they are switched off. Forecasting should not consider such periods as valid data to avoid biased estimates, as they result from the switchoff as a side effect. To solve this issue, the forecaster considers the *joint* utilization of the capacity and coverage cells (i.e., the sum of their individual utilizations) as input. Choosing an appropriate forecasting method, however, relies on selecting the appropriate timescale to predict PRB utilization. In fact, a too detailed characterization of PRB utilization incurs in significant overhead, while a too coarse granularity may not be representative of the actual utilization dynamics.

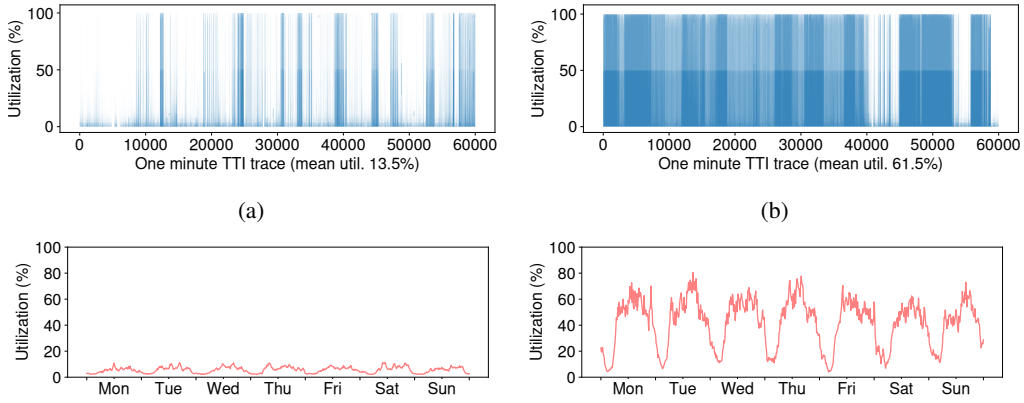
In this respect, we characterize the properties of PRB utilization traces at different timescales. In LTE networks, the radio resources of a cell (i.e., PRBs) are scheduled and allocated to users at a *transmission time interval* (TTI) of 1 ms. If we could accurately forecast utilization at such a

[2]This design choice does not actually prevent to obtain considerable energy savings, as it will be shown in Section V.



(a) (b)

Fig. 3: (a) Base station with co-located coverage-capacity cells. (b) Sample allocation of users to cells over time: the coverage cell is always on while the capacity cell is switched off during periods of low utilization. Users of a switched off capacity cell are moved to the coverage cell as long as resources therein are sufficient: the capacity cell can be switched on again (and users assigned to it) when necessary.



(a) (b)

(c) (d)

Fig. 4: Sample utilization traces from an actual mobile network with different granularities: data for one minute of traffic for a cell with (a) low and (b) high utilization at a millisecond-level granularity; data for one week of traffic for a cell with (c) low and (d) high utilization collected over 15-minute intervals.

small timescale (every millisecond), we could make real-time decisions to switch off cells and achieve larger energy savings. In fact, prediction of packet arrivals in each TTI has been found to be feasible in [27, 28]; however, an analysis of the PRB utilization at this granularity has not been studied so far. Our analysis has found that the utilization trace at the TTI level for each pair of coverage and capacity cells is actually a *stationary* time series, without any trend or seasonality – thus, it does not have a predictable pattern in the long-term [29]. We use the *Augmented Dickey-Fuller test* [30], a statistical method for testing whether a time series is stationary or not: the null hypothesis is that the time series is non-stationary. The p-value of the Augmented Dickey-Fuller test for the utilization at TTI level is near zero, meaning that the time series is stationary. Figures 4a and 4b show two representative examples of real traces from base stations collected at 1 ms intervals over a period of 1 minute with a mean utilization of 13.5% and 61.5%, respectively. An analysis of the distribution of the PRB utilization shows that the utilization tends to have small values (below 20%) or large values (nearly 100%) but not the ones in the middle, irrespective of the mean utilization.

Although TTI-level traces can be obtained from the base stations, monitoring the utilization at this granularity results in a large overhead and is typically carried out only when strictly necessary [31]. Thus, metrics are generally aggregated – typically in 15-minute intervals – and sporadically sent (e.g., once a day) to the network operator for processing and analysis using appropriate tools and dashboards [31]. In contrast to the utilization time series at the TTI level, the traces at 15 minute intervals are indeed non-stationary (e.g., with p-values of 0.74 ± 0.24 for 80 pairs of cells in our dataset) and clearly exhibit daily patterns regardless of overall average utilization (see Figures 4c and 4d). Therefore, we consider aggregated utilization traces in the rest of the article.

B. Trace expansion

As previously discussed, a granularity in the order of minutes is appropriate for forecasting, resulting in low overhead and storage requirements. However, data at much smaller timescales (e.g., at the TTI level) is still required to evaluate a control strategy during power management. To address such an issue, this section proposes a method to generate utilization data at the TTI level from traces with a higher granularity, namely, seconds or even minutes. The corresponding process

Timestamp	U	T
02:20:02:477	0.16	1
02:20:02:478	0.04	7
02:20:02:485	0.04	8
02:20:02:493	0.25	$\cdots$

(a)

Util. range	Sampled discrete distribution					
$[0-20]\%$	$(1, 0.38)$	$(2, 0.33)$	$(3, 0.12)$	$\cdots$	$(106, 0.003)$	$\cdots$
$[20-40]\%$	$(1, 0.50)$	$(2, 0.36)$	$(3, 0.08)$	$\cdots$	$(296, 0.0025)$	$\cdots$
$[40-60]\%$	$(1, 0.65)$	$(2, 0.29)$	$(3, 0.03)$	$\cdots$	$(13, 0.0017)$	$\cdots$
$[60-100]\%$	$(1, 0.71)$	$(2, 0.27)$	$(3, 0.015)$	$\cdots$	$(10, 0.0014)$	$\cdots$

(b)

TABLE I: (a) Sample PRB utilization at the TTI level. (b) Sampled discrete distributions for time duration T aggregated by utilization ranges.

operates on the data in the radio information database – equivalently, utilization traces from a dataset – and it is referred to as *expanding* the aggregated values. In the following, the properties of a real TTI trace are described first and then leveraged to derive a model to generate such a trace from data with coarser granularity. The correspondence between the two traces is finally established through an evaluation of their statistical properties.

1) Characterization of TTI-level traces: Table Ia shows an example of the PRB utilization at the TTI level, where the first column denotes the timestamps (with millisecond accuracy), the second column indicates the corresponding PRB utilization at each timestamp and the third column indicates the time duration between two consecutive records. The TTI-level trace typically has only non-zero utilization records to minimize the size of the file, i.e., the utilization is 0 at all timestamps not present in the TTI trace.

Thus, there are two random variables that are important for expanding the aggregated trace: one is T – the time duration between timestamps (i.e., [1, 7, 8, ...]), and the other is U – the utilization at each timestamp (i.e., [0.16, 0.04, 0.04, 0.25, ...]). T is a discrete random variable which also depends on the aggregated PRB utilization. That is, if the aggregated utilization is high, the probability of $T = 1$ ms is high as well (i.e., the PRBs are utilized more often) compared to when the aggregated utilization is low. Accordingly, we sampled 1,000 discrete distributions for T from the real TTI traces for different aggregated utilization values (in ranges of [0%-20%], [20%-40%], [40%-60%], [60%-100%]) as shown in Table Ib. The table shows that the percentage of occurrence of 1 ms increases as the aggregate PRB utilization increases.

The U values are assumed to be independent and identically distributed (i.i.d.). The time series plots (Figures 4a and 4b) show that there is no obvious trend in the utilization, indicating that the data appears to be identically distributed. Next, we check whether the U values are independent by plotting lag plots, i.e., scatter plots of each utilization value (U) in the real TTI trace against a point at a different time lag, k (for instance, 1 ms ahead). We consider different lag values $k = 1, 50, 500, 1000$. Figure 5 shows the points are randomly distributed in all lag plots demonstrating that the data is independent and not correlated. Based on our analysis of the TTI-level trace, we conclude the U sequence is i.i.d. Next, we describe our approach to generating such traces from the aggregated data.

2) Generation of TTI-level traces from aggregated data: For each row in the aggregated trace, we first obtain a list of timestamps at 1 ms intervals and then get the corresponding utilization values for the expanded timestamps. The list of timestamps are obtained by sampling the time durations from the discrete distribution (Table Ib) for the corresponding utilization range.

Algorithm 1: GETBETAPARAMS$(m, t = .5)$

Input: m, an aggregated mean to be expanded
Output: α, β for the corresponding beta distribution
1 **if** $m < 0.5$ **then** $v = t \cdot m$
2 **else** $v = t \cdot |m - 1|$
3 $\alpha = ((1 - m)/v - 1/m) \cdot m^2$
4 $\beta = ((1 - m)/v - 1/m) \cdot m \cdot (1 - m)$
5 **return** α, β

TABLE II: The p-values of the paired t-test results for the percentages of different sleep modes (SM) in real and sampled traces for different signaling synchronization (SS) periodicity.

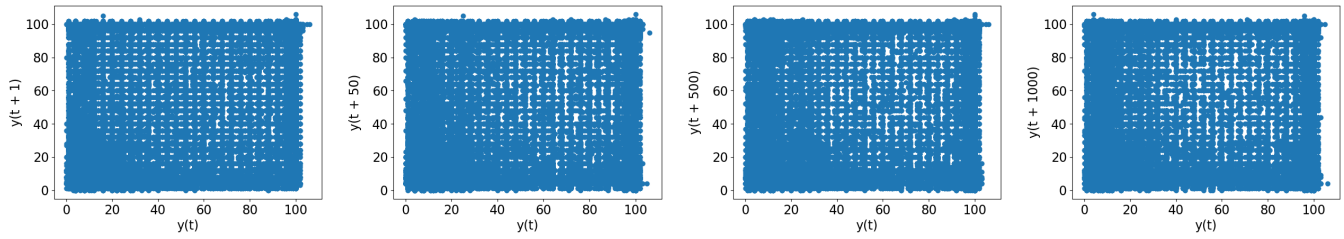
SS	SM_1	SM_2	SM_3	SM_4
5	0.22	0.18	N/A	N/A
10	0.22	0.56	N/A	N/A
40	0.23	0.37	0.11	N/A
100	0.23	0.50	0.04	N/A

We observe that the distribution of utilization values follows a U-shape (Figure 6a), with several occurrences of both high and low utilization values. Accordingly, we choose the beta distribution to model the utilization as it follows a U-shape when its parameters, α and β, are less than one [32]. This distribution is commonly used to model such patterns that are constrained between 0 and 1 [33, 34]. We propose Algorithm 1 to determine α and β, given the aggregated PRB utilization, m. We use a linear relationship between the variance v and m through an empirically determined coefficient t (lines 1–2). The variance decreases when the aggregated utilization m is close to 0 or 1. For instance, if m is 1, the expanded utilization traces are all set to 1 (i.e., the variance is 0). Next, α and β are set (lines 3–4) according to properties of the beta distribution [32]:

$$m = \frac{\alpha}{\alpha + \beta}, \quad v = \frac{\alpha \cdot \beta}{(\alpha + \beta)^2 \cdot (\alpha + \beta + 1)} \tag{1}$$

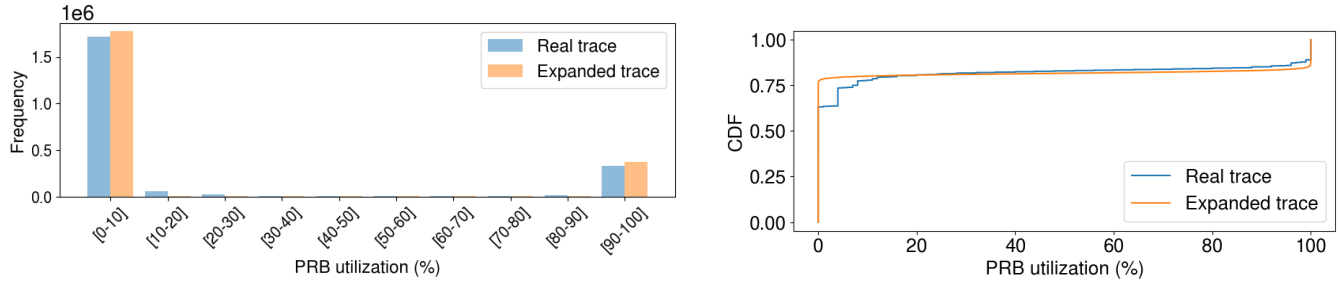
The beta distribution with α and β (set according to Algorithm 1) is used to sample the utilization values with the number of samples set to the size of the initially obtained list of timestamps.

3) Evaluation: We evaluated the properties of the expanded trace as compared to a real TTI trace, both in terms of the time durations between samples as well as the PRB utilization. To



(a) (b) (c) (d)

Fig. 5: Lag plots of utilization (U) values from original TTI trace at lag values (k) set to (a) 1, (b) 50, (c) 500, and (d) 1000.



(a) (b)

Fig. 6: (a) Histogram and (b) empiricial CDF of the PRB utilization for a real trace at TTI-level and the corresponding one for the sampled trace using our generation model.

this end, we used a TTI trace collected over a duration of 37 minutes in an LTE network. The aggregated one-minute utilization levels was input to our expanding approach; we then compared the properties of the real and expanded trace.

First, we test the properties of the time durations in the expanded trace. We particularly focus on obtaining an accurate measurement of energy consumption from the expanded trace. In this context, the time durations in the TTI trace when the utilization is 0 are particularly important. This is because base stations can go into *sleep modes* (SM), wherein a subset of its components is deactivated to conserve energy [3], whenever it is idle for a sufficiently long duration. Specifically, a cell enters SM1 if the utilization is zero for 0.0714 ms, SM2 after 1 ms, SM3 after 10 ms and SM4 after 1,000 ms (Table IV). A deeper sleep mode allows the base station to consume less energy. However, a base station has to periodically wake up to send signaling messages, i.e., synchronization signaling (SS), to provide network connectivity to users [20]. In our evaluation, we consider SS periodicity values supported by both LTE (5 ms) and 5G networks (10, 40 and 100 ms) [35, 36].

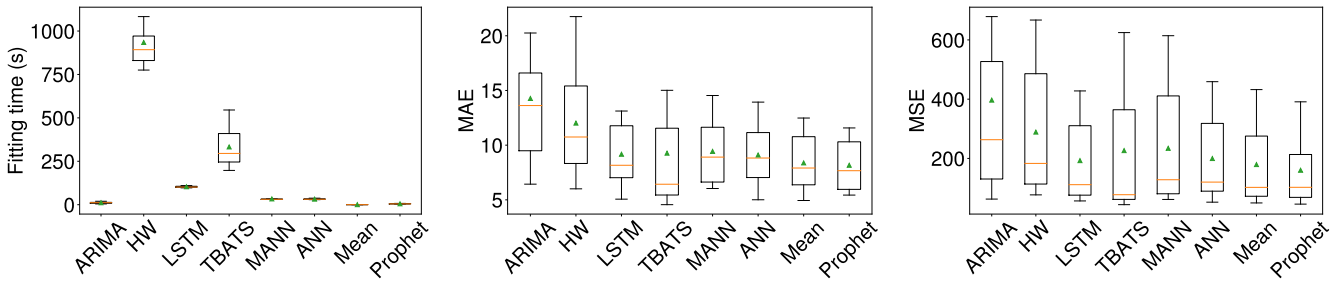
Therefore, for the real trace and the expanded trace, at each data point there is a pair of values SM_X and $\widehat{SM_X}$ corresponding to the percentage of time spent in each sleep mode ($X \in \{1, 2, 3, 4\}$). We used the paired t-test to test whether the percentages for each sleep mode for the 37 pairs have different mean values or not. Table II shows the p-values of the paired t-test for each SM where the null hypothesis is that both the real and expanded traces have the same mean time spent in each SM. The table clearly shows that the p-values are higher than 1% and cannot reject the null hypothesis. This indicates that the expanded trace results in a similar distribution of sleep modes and thus, a similar distribution of time durations as in the real TTI trace. Note that certain signaling periodicity values prevent a cell from entering deeper sleep modes; for instance, a cell cannot enter SM3 or SM4 when the SS is 5 ms. Accordingly, the use of certain sleep modes are denoted as N/A in Table II when the cell cannot enter these modes.

Next, Figure 6b shows the CDF of the PRB utilization for the real TTI trace as well as the expanded one. The CDF shows that the distribution of the values in the expanded and real trace have similar patterns, with a small variation in about 10% of the samples. We observe from the histogram (Figure 6a) that the real trace exhibits slightly higher occurrence of utilization values between 20% to 50%. However, this difference is minimal (as seen in the CDF) and we find our expanding approach is suitable for evaluating the energy savings[3].

IV. Forecasting Utilization to Save Energy

This section addresses the planning aspects related to saving energy in a cellular network. This corresponds to the forecasting and power management policy modules in the proposed energy saving system (recall Figure 2). First, we analyze the performance of state-of-the-art time series forecasters in predicting the utilization in traces obtained from real LTE networks. Next, we describe how a power management policy can define energy savings plans based on the predicted utilization as well as control strategies to overcome inaccurate predictions.

[3]Our evaluation uses the utilization values aggregated at 10-second intervals (see Section IV-B).



(a) (b) (c)

Fig. 7: Box plots with mean (denoted by triangles in each box) of (a) fitting time, (b) MAE and (c) MSE using different forecasting methods.

A. *Evaluation of different forecasting methods*

We evaluate the suitability of several state-of-the-art forecasting methods to predict the PRB utilization in pairs of capacity-coverage cells. Particularly, we choose *automated* forecasting methods that can be easily used by domain experts without detailed knowledge about the model itself. For instance, such methods require only the seasonality (weekly or daily) of the time series to be defined. Thus, they can be applied by operators to networks with unknown and possibly very different traffic patterns.

To this end, we first study the accuracy and fitting time of state-of-the-art forecasting methods, described next.

`Auto ARIMA` fits a range of AutoRegressive Integrated Moving Average (ARIMA) models and automatically selects the best one [37]. We use an open source Python implementation[4] of `Auto ARIMA`. We use a non-seasonal ARIMA model due to the extremely long fitting times when seasonality is included [38].

`Holt-Winters` (HW) [39, 40] decomposes time series into three components: a level, a trend, and a seasonal effect. In our experiments, we use `statsmodels`, a widely used Python implementation[5], with Holt-Winters's additive method and the seasonality set to one day.

`TBATS` [41] is a time series forecasting model based on exponential smoothing, and it supports different seasonalities – including daily, weekly, and yearly. We use an open source implementation[6] of TBATS in Python for our experiments.

`LSTM` is a type of Recurrent Neural Network (RNN) which is designed for sequential data and has been adopted in various domains including time series forecasting [11, 27, 42]. We implement LSTM in Tensorflow [43] with the hyperparameters provided by Trinh et al. [27] for forecasting traffic at base stations.

`ANN` uses feedforward Artificial Neural Networks (ANNs) trained to predict the utilization at time interval t of day d with the following[7] set of features [11] – $U_{d-1,t}$ (the utilization at t on the previous day), $U_{d-2,t}$ (the utilization at time t two days before d) and $U_{d-1,t-1}$ (the utilization at previous time interval t on the previous day).

`MANN` [11] is similar to `ANN`, but uses the mean of the values observed in the past. Specifically, the features used as input to the `ANN` are the difference between the mean utilization at time interval t and $U_{d-1,t}$, $U_{d-2,t}$ and $U_{d-1,t-1}$.

`Prophet` [38] is a regression model for time series forecasting developed at Facebook. It is designed to allow domain experts to add components (trend, seasonality, and holidays) to the model and easily make adjustments as needed. We use the open source implementation[8] released by Facebook Data Science team.

`Mean` is a baseline forecasting approach which simply utilizes the mean of the values observed in the past. It is used for comparison purposes only [38].

We evaluate the forecasting methods in terms of the fitting time and accuracy of forecasts on a dataset of PRB utilization for 11 pairs of cells over a period of 7 days obtained from an LTE network provider. The utilization of each cell is aggregated in 15-minute intervals, which results in 96 data points for each day. We used the utilization history of the first 6 days for fitting each forecasting model, and used the last day for evaluating performance in terms of mean absolute error (MAE), mean squared error (MSE) and fitting time. We set the seasonality in both `TBATS` and `Prophet` to one day.

Figure 7 reports the results from our experiments. We observe that `Mean` and `Prophet` have the shortest fitting time within 10 seconds. `ANN` and `MANN` achieve a fitting time of less then 50 seconds, whereas `LSTM`, `HW` and `TBATS` have a fitting time in the order of minutes. In terms of MAE and MSE, `Prophet` and `Mean` provide an overall better performance compared to the other methods. For instance, the average MAE values achieved by `Prophet` and `Mean` are 8.17 and 8.4 respectively followed by `ANN` (9.10), `LSTM` (9.17) and `MANN` (9.43). Similarly, `Prophet` provides the best performance for MSE (160.17) followed by `Mean` (179.28), `LSTM` (192.57), `ANN` (199.65) and `MANN` (234.17). The low spread of the accuracy metrics for `Prophet`, `Mean`, `LSTM` and `ANN` show that these methods have lower variability compared to the others. Thus, we consider `Prophet`, `Mean`, `ANN`, `MANN` and `LSTM` as

[4]https://github.com/alkaline-ml/pmdarima

[5]https://www.statsmodels.org/

[6]https://github.com/intive-DataScience/tbats/

[7]While Vallero et al. [11] additionally use feature $U_{d,t-1}$, we do not include it in our analysis as we forecast the utilization of the next day at once – the utilization values for day d are not available at the time of forecasting.

[8]https://facebook.github.io/prophet/

suitable candidates for forecasting PRB utilization at base stations in the next section.

B. Determining a power management policy

Once the predicted PRB utilization for a pair of capacity-coverage cells is obtained, the next step is to devise an energy savings plan. Specifically, an energy savings plan comprises the time periods during which a capacity cell may be switched off. Such time periods are chosen based on an operator-defined utilization *threshold* δ. Thus, the plan consists of the periods corresponding to when the predicted utilization does not exceed δ.

Next, switching off a cell typically requires migrating users from the capacity cell to the active coverage cell. An important consideration for determining the off-periods is the trade-off between energy savings and the overhead in moving users between cells. Frequently migrating users between cells is undesired [17, 18], as it may result in increased (signaling) load in the cells [44]. Moreover, there is a risk of service degradation due to potential back-and-forth signaling [45] when users are moved between cells [46].

A simple approach to reduce user migrations is to use a low δ value. However, this would result in smaller energy savings and under-utilized capacity cells. Another approach to reduce user migrations is to make the decision to switch off a cell as *robust* as possible. For instance, if the actual utilization is higher than predicted, our method should not switch off the capacity cell and unnecessarily migrate users to the coverage cell. To this end, instead of using the predicted values directly, we can use the upper bound of the forecasted values to decide when to switch off the cell. Figure 8 shows that the output from `Prophet` includes the predicted utilization (i.e., solid blue line in the figure) and an uncertainty interval[9] (*UI*) for each predicted value (for instance, $UI = 95\%$ for the lower and upper predictions in the figure). We evaluate the impact of choosing the time periods based on the uncertainty interval in the next section.

Once a plan has been derived, the capacity cell is gracefully shut down during the off-period and users in that cell are moved to the coverage cell. However, the execution of the plan requires the load in affected cells to be monitored in real-time to react to any unexpected behavior. For instance, the PRB utilization may be unexpectedly high during an off-period, in which case the capacity cell should not be switched off. On the other hand, if the traffic surges once a cell is switched off, recovery actions are required to switch on the capacity cell again. To this end, in practice, base stations employ a *control* strategy (recall from Section II that this strategy applies the power management policy in real-time at the base station) that maintains PRB utilization between a target maximum ($\text{PRB}_{\max}$) and minimum ($\text{PRB}_{\min}$) values. Given these target levels, during a configured off-period: a capacity cell is switched off only if the average utilization of both coverage and capacity cell for every ten seconds during the first five minutes of the planned off-period does not exceed $\text{PRB}_{\min}$; and the capacity cell is switched on immediately if the average PRB utilization of both coverage and capacity cells (measured every ten seconds) exceeds $\text{PRB}_{\max}$ continuously for two minutes. The time periods of five minutes and two minutes are used in existing energy saving features in base stations [8, p. 8]. Our solution is designed to be implemented in live mobile networks, and, thus, incorporates the monitoring intervals supported by base stations today.

[9]https://facebook.github.io/prophet/docs/uncertainty_intervals.html

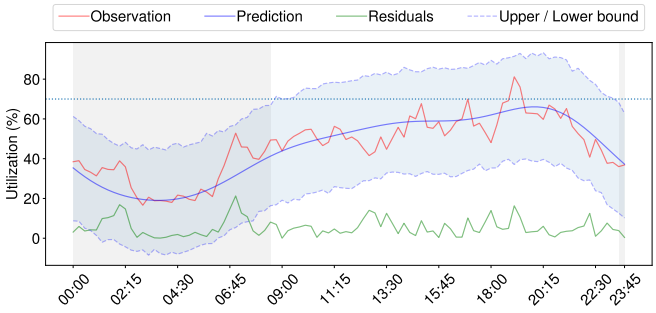


Fig. 8: The solid blue line is the forecasted utilization while the two blue dotted lines nearby are the related bounds of the 95% prediction interval. Only the time periods (shaded in gray) where the upper bound of the forecasted utilization is lower than an operator-defined threshold (e.g., 70%) are selected in an energy savings plan.

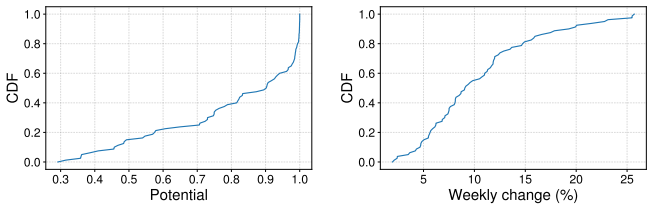
Even once an energy savings plan is obtained, the control strategy allows the base station to react to traffic changes in real-time and to, accordingly, decide whether to switch on or off the capacity cell. Alternatively, such a control strategy can be applied throughout the day, as a reactive approach, in coverage-capacity cells that are under-utilized (i.e., the PRB utilization is low). In such pairs, the utilization rarely exceeds δ and thus, energy savings are possible through the control strategy by setting $\text{PRB}_{\min}$ to δ. We evaluate the reactive approach in combination with the forecasting methods later in Section V.

V. Evaluation

We study the opportunities for energy savings through analysis of a real-world dataset comprising 80 pairs of cells. We first describe the dataset and its properties, then discuss some preliminary processing required for the evaluation setup. We then present a systematic study of the settings for forecasters and how they impact energy savings that can be achieved in a real network. We also present a thorough investigation of the trade-offs between energy savings and increased user migrations between inactive and active cells. The analysis provides insights into implementing the energy savings solution in real cellular networks, taking into account the network operator concerns.

A. Setup

1) Dataset: We used real cell utilization data obtained from an LTE network provider. The dataset contains the PRB utilization data and the number of connected users with a granularity of 15 minutes. The capacity-coverage relationship



(a) (b)

Fig. 9: CDF of dataset dimensions: (a) potential and (b) weekly change.

TABLE III: Range of values of potential and weekly change (with semi-open intervals) dimensions for each percentile value (k). For instance, $k = 20$ denotes the range of values between the 10th and the 20th percentiles in the dataset.

k	Potential	Weekly change
10	(0.292 - 0.457]	(1.945 - 4.653]
20	(0.457 - 0.575]	(4.653 - 5.646]
30	(0.575 - 0.730]	(5.646 - 7.084]
40	(0.730 - 0.816]	(7.084 - 8.090]
50	(0.816 - 0.902]	(8.090 - 9.056]
60	(0.902 - 0.941]	(9.056 - 11.011]
70	(0.941 - 0.985]	(11.011 - 11.980]
80	(0.985 - 0.994]	(11.980 - 14.731]
90	(0.994 - 0.999]	(14.731 - 19.270]
100	(0.999 - 1.000]	(19.270 - 25.705]

is also included in the dataset. From this dataset, we identified 80 pairs of cells where data was available throughout the two-month period; the related utilization pattern distribution was representative of the larger dataset. To establish the latter, we analyzed the *trace pairs* (i.e., the utilization traces of the paired capacity and coverage cells) and characterized them according to the following dimensions.

Potential, as the fraction of time that the sum of the coverage and capacity cell utilization was below 70%. This directly corresponds to the fraction of time that the capacity cell could be switched off, at least with perfect knowledge of future utilization. Therefore, it gives a measure of *potential energy savings*. For instance, a potential of 1 indicates that the capacity cell can be switched off permanently, while a potential of 0 indicates that it is never safe to switch off the capacity cell.

Weekly change, as the maximum discrepancy between the weekly mean utilizations over the considered time period, which captures the temporal variability of the cell. Intuitively, cells with high weekly change should be more difficult to predict than others.

Figure 9 characterizes the considered trace pairs according to the dimensions described above. We ensured that for each dimension, the 80 trace pairs are representative of the larger dataset. Figure 9a shows that about 40% of the cells have a high potential for energy savings. Most cells (Figure 9b) experience a weekly variation in mean under 15%, but there are many cases in the dataset with an increase or decrease in mean weekly utilization beyond this value.

For each dimension we created buckets according to the k-th to $(k-10)$th percentile of the values of the dimension, for $k \in [10, 20, 30, \ldots, 100]$. These are described by Table III and later used in presenting the evaluation results.

2) Methodology: We used a trace-driven approach to evaluate our solution. We would like to recall that the dataset described above contains PRB utilization metrics for each cell reported every 15 minutes. However, the control strategy to switch off a cell requires data at a smaller time interval, in the range of seconds (Section IV-B). To this end, we first generated such a trace containing the PRB utilization values at 1 ms intervals for the test period from the averaged 15-minute values (see Section III-B). This resulted in TTI-level traces for each pair of capacity and coverage cells. Specifically, each utilization data point from the original dataset is expanded to 900,000 data points at the millisecond level in the TTI-level trace. Once the expanded traces are obtained, we simulated the implementation of the energy savings plan including the control strategy for the time intervals specified in the plan. We evaluated the energy consumed in the traces through the widely adopted Greentouch model [3], which characterizes the energy consumed by a base station at different load levels as in Table IV. The model also includes advanced *sleep modes* of a base station, wherein a subset of its components is deactivated for further energy savings. In our evaluation, a cell automatically enters the sleep modes if the utilization is 0% for a duration longer than the time required to enter the sleep mode (i.e., the transition time in Table IV). However, an LTE base station sends synchronization signaling every 5 ms – even when the cell is empty [20] – preventing to employ sleep modes 3 and 4.

Our evaluation did not account for the potential increase in PRB utilization due to the additional messaging required for moving users from one cell to another. However, signaling messages are negligible with respect to typical traffic, and thus, the related impact on overall energy consumption is insignificant. Moreover, we have evaluated the impact of user transitions in a conservative manner by assuming that *all* users are migrated from a cell to another upon switching events. In practice, when a capacity cell is switched on, only a fraction of users needs to be migrated to it, as the coverage cell can continue serving at least part of those already connected to it.

3) Considered methods and metrics: We evaluate the performance of a subset of the forecasters[10] introduced in Section IV, namely, `Prophet`, `ANN`, `MANN`, `LSTM` and `MEAN`. Each forecaster utilizes the data from the past seven days to forecast the combined capacity-coverage utilization for the next day[11]. The value of the operator-defined utilization threshold δ (Section IV-B) was set to 70%[12] and accordingly, PRB_{max} was set to 70% and PRB_{min} to 65% in the control strategy to implement the plan. The test period consisted of 32 days.

[10]We do not consider `TBATS` and `HW` due to their long fitting time and `ARIMA` due to the large average MSE.

[11]We experimented with different settings for the number of days (7, 14, 21 and 42) included as the training data. We found no significant differences in the energy consumed; thus, we chose 7 as this results in smallest amount of data that needs to be sent to and processed by the forecaster.

[12]In fact, a PRB utilization over 70% may result in poor user experience and potential drop of user sessions [47].

TABLE IV: Power consumption and transition times in different states for a base station [3].

Metric	Active mode		Sleep mode			
	Full load	No load	1	2	3	4
Power (W)	702.6	114.5	76.5	8.6	6.0	5.3
Transition time (ms)	–	–	0.0714	1.0	10.0	1000.0

We evaluated the forecasters with different settings. Specifically, `Prophet` can be configured to use the predicted value directly, referred as `P0` or with a 95% uncertainty level, referred as `P95`. Since `Mean` does not provide a prediction interval, we use the 95th percentile of the past utilization values (for a fair comparison) and refer to it as `Stats`. `ANN`, `MANN` and `LSTM` do not support uncertainty intervals, and thus, their point predictions are used directly in determining an energy savings plan, similar to [11]. The forecasters with different settings are used to determine the time periods in which a cell can be switched off (according to Section IV-B).

In addition to the forecasting methods, we consider the following two approaches for comparison purposes.

`Oracle`. This method assumes the actual utilization of both the capacity and coverage cells at a 15-minute granularity are known in advance for the next day. This perfect forecast is then used to select the time periods when the capacity cell can be switched off, i.e., when the combined utilization of both cells does not exceed δ percent. `Oracle` represents the optimal savings possible when using a forecasting method to plan the intervals when the capacity cell is switched off.

`Reactive`. This represents an opportunistic energy savings solution that tries to greedily save energy by switching off the capacity cell whenever possible. This is the same as using the control strategy throughout the day. Specifically, the capacity cell is switched off whenever the combined utilization of the capacity and coverage cell goes below 70% for at least 5 minutes. It then remains off as long as the combined utilization does not exceed 70% for 2 minutes. This approach is similar to the one proposed in [16], wherein the cells are switched off whenever[13] the combined utilization is below a certain threshold. We refer to this as `React` throughout the article, for conciseness. Since this method operates at smaller time intervals than 15 minutes, it may achieve higher energy savings than `Oracle` or any other forecasting method.

We evaluate the performance on the basis of the following metrics.

Energy savings (ES). We measure the difference in the energy consumed when the capacity cell is switched off (and traffic offloaded to the coverage cell) compared to that when both cells are on, expressed in kilowatt hours (kWh).

Number of UE transitions. We measure the number of users affected whenever a capacity cell is switched on or off. This is calculated by summing the number of user equipment (UEs) associated with the capacity cell at each time the cell is switched on or off. This metric represents the overhead in switching off cells.

[13]A strategy to switch on the cells in case of excess load is not defined in [16], which is crucial in a real network.

Energy savings per UE transition. We calculate the difference in the energy consumed when the capacity cell is switched off compared to that when both cells are on (in kWh) divided by the total number of UE transitions. This metric captures both the absolute energy savings and the impact on the users: a higher value indicates that the energy savings are obtained with lower impact on the users.

Occurrence of excess load. This metric denotes the number of times the PRB utilization of the coverage cell exceeds 70% (aggregated every 10 seconds) during the periods when the capacity cell is off. A high PRB utilization indicates that the cell is congested when handling the extra traffic from the capacity cell that is off.

Each reported metric is calculated per test day and pair of cells, and thus, the values are reported as average over the pairs of cells and over the test period of 32 days. Accordingly, we also report the standard deviation in the obtained results, which captures the variation over cell pairs.

B. Obtained results

1) Comparison of methods across dataset dimensions: We performed a segmentation analysis of results by grouping cell (or trace) pairs according to the percentile value, k, of dimensions reported in Table III. The most interesting dimension is potential (Figure 10), as it exhibits a clear dichotomy between the trace pairs in terms of both the number of UE transitions by the different strategies, as well as the energy savings achieved by them. Figure 10a shows that, rather intuitively, energy savings increase with the potential. Both `React` and `Oracle` achieve the highest energy savings, with `React` performing slightly better than `Oracle`. The reason for this is that `Oracle` is based on a pre-planning schedule computed using data aggregated to 15 minute intervals. As a consequence, the aggressive switching off policy provided by `React`, which operates at a finer timescale, achieves slightly more energy savings. However, `Oracle` does not incur any excess PRB utilization (Figure 10d), whereas `React` is the worst method with respect to this metric. Next, `ANN`, `MANN`, `LSTM` and `P0` achieve similar energy savings, whereas `Stats` and `P95` have lower average savings as they are more conservative (i.e., the capacity cell is switched off based on the 95th percentile of observed values or the upper bound of the 95% prediction interval, respectively). However, this conservative approach of `P95` allows it to maintain a lower number of user migrations (Figure 10b).

Figure 10b shows the average number of UEs transitioned by each strategy per day. The number of UE transitions is less than 30 for cells with potential values in the 10th percentile and starts increasing until the 30th percentile range, and then decreases afterwards. As expected, when the potential is low,

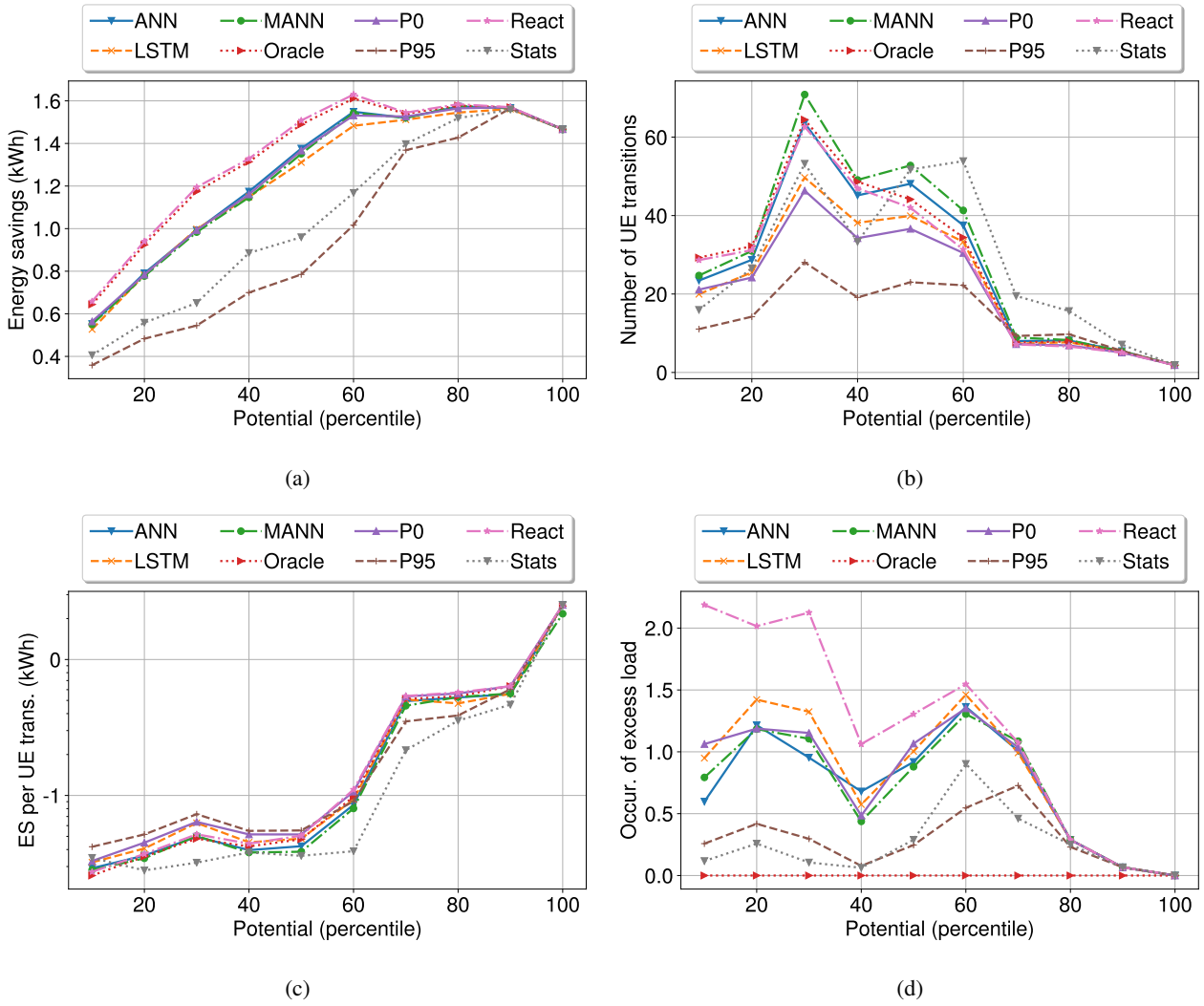


Fig. 10: (a) Energy savings (kWh) per day (on log scale), (b) number of UE transitions per day, (c) energy savings (kWh) per UE transition (on log scale), and (d) occurrence of excess load per day, averaged over all trace pairs as a function of potential. Each bucket (i.e., x tick) on the x-axis corresponds to 10% of the trace pairs.

the cells tend to be highly utilized and the opportunity to switch off the capacity cell is limited, which in turn limits the number of UE transitions. The peak of the number of UE transitions occurs for potential in the 30th percentile range (specifically, $0.575 - 0.730$) as cells in this range have a utilization level around δ. This results in more transitions if point predictions are in the range of δ and could result in more cycling between on and off states if there is a variation in actual observed traffic. Thus, `P95`, with its conservative approach, has far fewer UE transitions ($10-29$) than the other strategies (when the potential is less than 0.902). `P0` follows a similar pattern though with more UE transitions ($19 - 48$). `Oracle` and `React` exhibit higher and similar numbers of UE transitions ($27 - 67$), while `Stats` is more inconsistent and occasionally obtains the maximum value (i.e., for potential in the 50th percentile range). Interestingly, on the right side of the figure (with potential at least 0.902), the number of UE transitions significantly reduces for most strategies. In detail, `P0` and `React` require the fewest UE transitions in this region. As the cells are less utilized, there are fewer users that need to be moved between cells for all strategies.

Next, in terms of energy saved per UE transition, Figure 10c indicates that `P95` significantly outperforms other methods (with a p-value< 0.05) for potential less than 0.902. However, in cells with higher potential, `LSTM`, `ANN`, `MANN` and `P0`, being point predictors, are able to achieve a higher energy savings. This is because the negative effects of under-estimating delivery ratio are more rare as the cells are under-utilized in any case. Finally, `P95` and `Stats` result in the lowest occurrence of excess load (Figure 10d), whereas a reactive approach such as `React` incurs more in excess load.

Finally, for the weekly change dimension, we do not observe a significant difference between the point predictors or those with uncertainty intervals. However, Figure 11 shows that mean occurrence of excess load increases as the week-to-week-change values grows. Here again, `P95` is able to maintain a lower occurrence of excess load conditions due to its conservative approach.

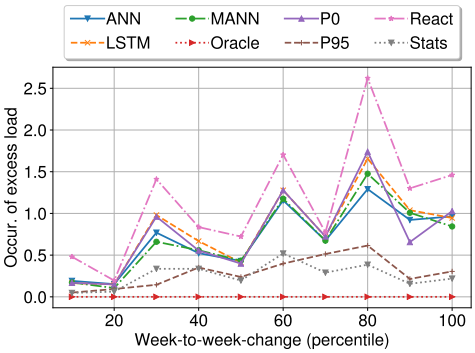


Fig. 11: Average occurrence of excess PRB utilization per day as a function of the week-to-week-change.

TABLE V: Overview of results averaged for pairs of cells with *low* potential.

	ES/UE trans. (kWh)	**ES** (kWh)	**Occur. excess load** (#)
React	0.042 (±0.032)	1.126 (±0.346)	1.739 (±1.709)
Stats	0.034 (±0.025)	0.692 (±0.246)	0.166 (±0.262)
Oracle	0.040 (±0.030)	1.109 (±0.344)	0.000 (±0.000)
LSTM	0.046 (±0.039)	0.954 (±0.320)	1.055 (±1.107)
ANN	0.039 (±0.032)	0.977 (±0.330)	0.873 (±0.920)
MANN	0.038 (±0.035)	0.962 (±0.322)	0.880 (±0.995)
P0	0.049 (±0.039)	0.974 (±0.319)	0.991 (±1.092)
P95	0.055 (±0.046)	0.574 (±0.207)	0.259 (±0.398)

2) Difference between cells with high and low potential: Table V presents the overall results for the 40 pairs of cells with a low potential. The highest energy savings per UE transition is achieved by P95. Interestingly, Oracle achieves a lower value in this respect. This indicates that the prediction interval is more important than the accuracy of forecasting to ensure that few users are impacted when designing energy savings solutions. For instance, Figure 13 shows the average number of UE transitions in detail over the 32 days in the test period. As we can see from the figure, P95 is able to maintain a lower number of UE transitions on all days as compared to all other methods. Here, we also draw a distinction between the UE transitions as a consequence of using a forecasting method and React. Transitions can be better planned with forecasters, as the off-periods are known in advance. For instance, the operator can implement a graceful shutdown of the capacity cell according to the schedule, by gradually powering down the cell and migrating traffic gracefully to the active cell. On the other hand, all decisions to switch on or off the cell with React are unplanned and may result in service degradation for users upon frequent switches.

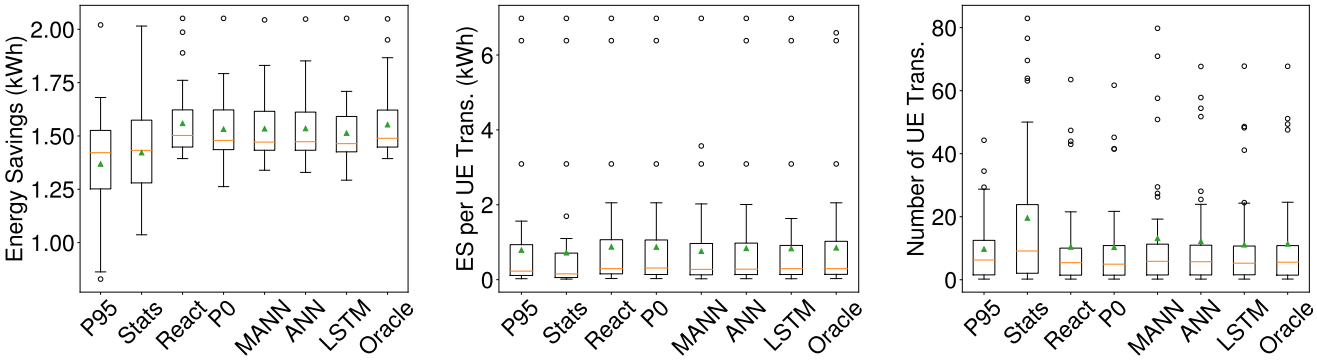
Next, Figure 12 presents a summary of the metrics for the 40 pairs of cells with a high potential. We observe that in such cells, the highest energy savings are obtained by React and the forecasters that use point predictions when compared to P95. For example, React provides the highest energy savings (1.56kWh) which is significantly better than 1.37kWh achieved with P95. However, there is no significant difference between React and the point predictors (P0, LSTM, ANN or MANN) in terms of the energy saved per UE transition. Thus, it is reasonable to use any of the point predictors or even React in such cells to achieve an overall higher energy savings. Also, React is as good as the point predictors (P0, LSTM, ANN or MANN) in terms of the energy saved per UE transition (i.e., no significant difference). An operator would prefer React for high potential pairs of cells for several reasons. First, the daily pattern of a high potential pair of cells may not ideal for forecasting methods (e.g., see Figure 14). Secondly, a forecasting method works exactly the same as React when its pre-plan switches off the capacity cell for the whole day.

3) Stability of the energy savings system: Finally, we evaluate how often the capacity cell is switched on and off in succession, commonly referred to as ping-pong effect [48]. Such ping-pong effect is undesirable, as the users are migrated from the capacity cell to the coverage cell for only a short duration (i.e., 10 minutes) before being moved back again to the capacity cell that is switched on in the next time interval. We find that the control strategy at the base station (Section IV-B) prevents the cell from switching on and off the cells in consecutive 15-minute intervals. Specifically, a capacity cell is switched off only after observing whether the actual PRB utilization (of both the cells) is below PRB_{min} for five minutes. In the test period of 32 days, there are 233 intervals where the planned duration is 15 minutes when P95 is used to forecast the off periods. Nevertheless, over all 80 pairs of cells, a capacity cell is switched off for 10 minutes (excluding the 5-minute monitoring interval) only 185 times over the 32-day test period. Thus, the average occurrence of this effect is less than 0.07% over all pairs of cells. This demonstrates the importance of having a control strategy to respond to real-time traffic. However, since the planned off-periods are known in advance, an operator may also choose to remove such 15-minute intervals from the planned off-periods. In contrast, React switches off cells whenever the utilization goes below PRB_{min} and is thus suitable only for high potential cells, where the cells are typically less utilized.

VI. Discussion

a) Summary and main insights: The above analysis of cell pairs provides a framework for operators to select methods appropriate for *their network* based on: the potential of the cell pairs, and; what they deem most important in their network. For instance, given our dataset, using P95 for pairs with potential less than 0.902, and React for the rest is the best strategy for minimizing the number of UE transitions. We note that the potential can be calculated using the previous few weeks of utilization history (e.g., aggregated to 15 minute intervals), which is likely to be available to an operator.

Finally, the energy savings when expressed as a percent show greater variation between high and low potential cell



(a) (b) (c)

Fig. 12: (a) Energy saved (ES) per day (kWh) (b) Energy savings (kWh) per UE transition, and (c) Number of UE transitions per day for 40 cells with *high* potential.

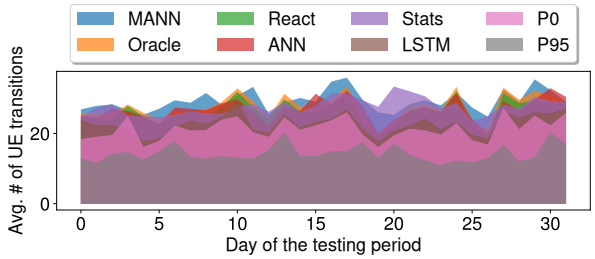


Fig. 13: Average number of UE transitions for pairs of cells with *potential* below 0.902 over 32 days.

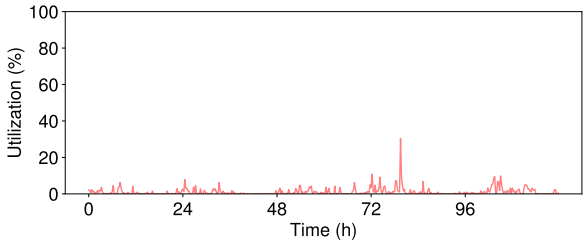


Fig. 14: Utilization trace for a sample cell pair with high potential where there is no daily pattern, which is not ideal for using forecasting methods.

pairs. For instance, as a percentage of the energy consumed when both cells are on, `Oracle` can achieve 22%–33% energy savings for high potential cases and around 6% for low values of potential (although in terms of absolute numbers they vary between 1 to 1.5 kWh). The large difference in the energy savings achievable for cell pairs with high and low potential indicate that the composition of cells with different utilization levels used for different studies may result in different levels of energy savings. This can be one of possible reasons for the different degrees of energy savings claimed by different companies [49, 50] or in the literature [21]. Nevertheless, in our network, an average of 10.24% energy savings is achieved by utilizing `P95` in the low potential cells and `React` in high potential cell pairs. This is in line with the energy savings of approximately 6% reported in [51] for scenarios where cells are shutdown.

b) Extensions to other scenarios: Our solution determines energy saving plans for pairs of capacity-coverage cells at each base station. However, there is typically more than one capacity cell per base station. In this case, it is possible to extend our approach by prioritizing capacity cells and then switching them off according to the assigned priority (e.g., by increasing average utilization). The forecasting of utilization still takes place over pairs of cells, as capacity cells are switched on or off one at a time. Moreover, the O-RAN architecture allows network operators to collect and analyze data from multiple cells that are not necessarily located at the same base station. In fact, our solution is applicable as long as a coverage-capacity relationship exists, such as in heterogeneous networks with macro coverage cells and small cell base stations for added capacity [15].

c) Energy savings in 5G networks: Our results focused on an LTE network, as the dataset we had access to is based on such a technology. However, the proposed approach could be applied to 5G networks as well. In fact, our evaluation incorporates the periodic synchronization signaling in LTE that takes place every 5 ms, thereby preventing a cell from entering deeper sleep modes [20]. In contrast, the periodicity of such signaling can be configured to as large as 160 ms in 5G [35, 36]. As a consequence, base stations are able to enter deeper sleep modes when synchronization is more sporadic. Thus, instead of switching off the cell, the energy savings plan can configure a larger periodicity of signaling during the off-periods. Our preliminary results have shown that considerable energy savings are possible in underutilized capacity cells by increasing the signaling periodicity to 160 ms. An advantage of this approach is that users need not be moved between cells. However, such an approach would need to be aware that the periodicity of signaling is affected by the type of services – for instance, low-latency services cannot tolerate an increase in delay due to less frequent signaling.

VII. Related Work

Reducing energy consumption in mobile networks is an active research area [6, 52, 53]. More specifically, several

works target switching off base stations during periods of low utilization [11, 14, 17–20, 54]. In this context, Han et al. [55] survey several load-aware optimization problems that switch off under-utilized base stations. However, many works focus on the impact of switching off cells on interference and re-association of users to the active base stations [19, 54]. For instance, Beitelmal et al. [19] analytically evaluate energy savings of switching off cells for the special case where base stations are evenly distributed to cover a certain geographical area. Feng et al. [54] propose an optimization problem to switch off under-utilized base stations to improve energy efficiency through power control. Their solution also determines the optimal association of users to active base stations. Rached et al. [18] propose an algorithm that dynamically switches off base stations while satisfying a given power budget and a minimum percentage of successfully served users. Their solution determines the time periods during which base stations are switched on or off based on an operator-defined risk level and by observing the traffic in real-time. In contrast to the approaches above, our solution considers pairs of coverage-capacity cells within a single base station. Consequently, it does not require coordination between separate base stations. Moreover, the articles described above address the problem of energy-efficient mobile networks by using either analytical modeling or simulations. Instead, we focus on a solution that can be easily applied to real cellular networks.

Some works in the literature are grounded on data available from cellular network operators. Among them, Vallero et al. [11] predict the traffic in each cell and use this forecast to switch off cells when underutilized. Next, Parzysz and Gourhant [20] propose the use of flexible duty cycles, so that the base stations can be put in deeper sleep modes in periods of low utilization. Their solution requires knowing the average utilization in advance; however, they do not focus on forecasting utilization in their article. Dalmasso et al. [16] present a simple heuristic that monitors the utilization at the beginning of each half hour and reactively switches off underutilized base stations for that duration. We evaluate such a reactive approach in this article and find that it works best for low utilized cells, whereas, many users need to be migrated between cells in busier cells. Peng et al. [14] present a traffic-aware algorithm to select active base stations in 3G networks. Their solution divides the network area into grid cells and uses the traffic profile information to determine active sets of base stations from those with overlapping coverage. However, determining such a grid may be complex, especially for scenarios with a high density of base stations [15], such as in urban environments. In contrast, we target switching off cells within a single base station to save energy, without requiring global information about the network. As a consequence, our approach scales to cellular networks with a large number of base stations. To the best of our knowledge, this article presents the first comprehensive data-driven study of energy savings for modern cellular networks based on forecasting PRB utilization.

Traffic prediction in cellular networks is an active research area as well [11, 12, 21, 56]. The Holt-Winters (HW) model is used to predict the hourly utilization at base stations in [21, 56]. However, the forecasting performance is not compared to other alternatives. Instead, we evaluate the HW model and show that it is not as accurate as other forecasting methods for our traces with a 15-minute granularity. Vallero et al. [11] propose feed-forward artificial neural networks to predict the hourly traffic in cells and compare their approach to Long Short Term Memory (LSTM) networks [57]. We evaluate both the neural networks and LSTM in this article and find that the absence of an uncertainty interval in the predictions limits their use in busy cells for energy savings. Xu et al. [12] employ a distributed Gaussian process to predict hourly utilization and evaluate their solution on a dataset collected from a real network in China. Their solution also includes a measure of uncertainty in the predicted results. However, their proposed solution has a high computational complexity and would need to be tuned for each network in which it is deployed. In contrast, we focus on automated forecasting methods that require only minimal configuration. Azari et al. [58] show that LSTM outperforms ARIMA in predicting user traffic at a timescale between 2 to 60 seconds. However, the evaluation was carried out on a traffic trace generated by the authors. Nevertheless, we also find that LSTM performs better than ARIMA when forecasting traffic at 15-minute intervals in our dataset. Finally, Albanna and Yousefi'zadeh [59] observe that certain cells are able to handle more users before the PRB utilization reaches a certain threshold (80% in their article). Accordingly, they propose a deep neural network that predicts the average number of connected UEs that cross a utilization threshold of 80% . Such a predictor can help set different cell-specific utilization thresholds in our energy savings solution to maximize the energy saved per UE transition. A few works have addressed the prediction of network utilization and traffic at shorter time intervals. Trinh et al. [27] propose the use of LSTM networks to predict traffic (in Mbps) at the TTI-level (1 ms) and find that LSTM outperforms ARIMA models. More recently, Rostami et al. [28] present an LSTM network to predict traffic (packet arrival times) in the next 1 to 30 TTIs. They find that the prediction accuracy is high when predicting arrivals up to 15 TTIs ahead. However, their evaluation considers only lightly loaded base stations with the dataset captured between 1 a.m. and 6 a.m. In contrast, we find that the utilization at TTI-level is a stationary time series in our dataset, and thus, does not have a predictable pattern in the long-term. In a different scenario of forecasting data center load, Mozo et al. [60] observe that the utilization trace at one second granularity contains a lot of noise. Specifically, they find that at a one-second resolution, the forecaster (using neural networks) does not outperform a naive approach that simply forecasts the last observed value.

Finally, an analysis of traffic utilization patterns in TTIs has not been explored in the literature. The closest relevant work is presented by Peng et al. [61]. The authors propose compound probability distributions to model the packet arrivals and their lengths in a TTI. Specifically, the distributions model the arrival of a heavy-tailed number of heavy-tailed packet lengths. In this article, we model the distribution of PRB utilization in TTIs with a beta distribution to represent the U-shaped pattern of utilization values.

VIII. Conclusion

In this article, we proposed a data-driven approach to forecast the time periods during which an under-utilized cell can be switched off and save energy in cellular networks. To this end, we evaluated several state-of-the-art forecasting tools in predicting future utilization in traces obtained from a live LTE network. Our evaluation indicates that the design of the energy saving solution does not depend only on the accuracy of the forecasting tool. This is because switching off a capacity cell at a base station impacts the UEs (end users), who then have to be migrated to the active coverage cell. In this context, a network operator typically aims to minimize the number of user transitions. Thus, a forecasting tool that provides prediction or uncertainty intervals is crucial to plan the durations when a cell can be switched off. Specifically, a conservative approach that utilizes the upper bound of the 95% prediction interval is able to achieve a higher energy savings per migrated UE. Nevertheless, in cells that are usually under-utilized (for instance, those that can be switched off during the whole day), the uncertainty interval is not as important, and a point predictor is able to achieve higher energy savings. Overall, our evaluation showed that by switching off capacity cells based on predicted load, an energy savings of 10.24% is realistically possible. The analysis considered only switching off one capacity cell at a time, and thus, more promising energy savings are expected when more capacity cells are turned off. Due to the encouraging results, we aim to implement such a feature in a real cellular network. In this article, we have also proposed a method to generate PRB utilization traces at the TTI-level from aggregated data. Thus, researchers can evaluate the TTI trace generation with their own datasets. Indeed, a Poisson model may also be used for certain ranges of average utilization values [61]. Although we find the expanded TTI trace suitable for our evaluation, further modifications of the algorithm are required to obtain a better fit for PRB utilization at the TTI-level. Evaluating other such models remains as future work, and our analysis provides a valuable step in this direction.

IX. Acknowledgment

The authors would like to thank Oliver Blume for guidance with the *Green Touch* power model, and Balakrishnan Chandrasekaran for his feedback on the article.

References

[1] "Ericsson Mobility Report, June 2019," Tech. Rep., 2019, available at https://www.ericsson.com/4acd7e/assets/local/mobility-report/documents/2019/emr-november-2019.pdf.

[2] F. Richter, A. J. Fehske, and G. P. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks," in *2009 IEEE 70th Vehicular Technology Conference Fall*. IEEE, 2009, pp. 1–5.

[3] B. Debaillie, C. Desset, and F. Louagie, "A flexible and future-proof power model for cellular base stations," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*. IEEE, 2015, pp. 1–7.

[4] NGMN Alliance, "5G white paper-executive version," 2014.

[5] R. Clark, "Operators Starting to Face Up to 5G Power Cost," 2019. [Online]. Available: https://www.lightreading.com/asia-pacific/operators-starting-to-face-up-to-5g-power-cost-/d/d-id/755255

[6] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5g networks," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 72–80, 2017.

[7] GSM Association, "2019 mobile industry impact report: Sustainable development goals, executive summary," https://www.gsma.com/betterfuture/2019sdgimpactreport/wp-content/uploads/2019/09/SDG_Report_2019_ExecSummary_Web_Singles.pdf, 2019.

[8] Nokia, "LTE1203 - Load-based power saving with Tx path switching off," 2016, accessed 20/10/2020. [Online]. Available: https://telecomfiles.com/lte1203-load-based-power-saving-with-tx-path

[9] A. Mathur, "5g densification and network power efficiency," 2018, accessed 20/10/2010. [Online]. Available: https://futurenetworks.ieee.org/images/files/pdf/SantaClaraTutorial2018/5-5G_EET_5G_Densification_Power_Efficiency_Apurv_Mathur_Nokia-FINAL-MATHUR-NOKIA-PDF-TO-SHARE-9-28-18.pdf

[10] M. Polese, R. Jana, V. Kounev, K. Zhang, S. Deb, and M. Zorzi, "Machine learning at the edge: A data-driven architecture with applications to 5g cellular networks," *IEEE Transactions on Mobile Computing*, 2020.

[11] G. Vallero, D. Renga, M. Meo, and M. A. Marsan, "Greener ran operation through machine learning," *IEEE Transactions on Network and Service Management*, vol. 16, no. 3, pp. 896–908, 2019.

[12] Y. Xu, F. Yin, W. Xu, J. Lin, and S. Cui, "Wireless traffic prediction with scalable gaussian process: Framework, algorithms, and verification," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1291–1306, 2019.

[13] O-RAN Alliance, "O-RAN: Towards an Open and Smart RAN," Tech. Rep. October, 2018.

[14] C. Peng, S.-B. Lee, S. Lu, H. Luo, and H. Li, "Traffic-driven power saving in operational 3g cellular networks," in *Proceedings of the 17th annual international conference on Mobile computing and networking*, 2011, pp. 121–132.

[15] M. Feng, S. Mao, and T. Jiang, "Base station on-off switching in 5g wireless networks: Approaches and challenges," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 46–54, 2017.

[16] M. Dalmasso, M. Meo, and D. Renga, "Radio resource management for improving energy self-sufficiency of green mobile networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 44, no. 2, pp. 82–87, 2016.

[17] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE transactions on wireless communications*, vol. 12, no. 5, pp. 2126–2136, 2013.

[18] N. B. Rached, H. Ghazzai, A. Kadri, and M.-S. Alouini, "A time-varied probabilistic on/off switching algorithm for cellular networks," *IEEE Communications Letters*, vol. 22, no. 3, pp. 634–637, 2018.

[19] T. Beitelmal, S. S. Szyszkowicz, D. González, and H. Yanikomeroglu, "Sector and site switch-off regular patterns for energy saving in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 2932–2945, 2018.

[20] F. Parzysz and Y. Gourhant, "Drastic energy reduction with gdtx in low cost 5g networks," *IEEE Access*, vol. 6, pp. 58 171–58 181, 2018.

[21] D. Clemente, L. Ferreira, G. Soares, N. Valente, and P. Sebastião, "Implementation of a cloud-based, traffic aware and energy efficient management of base stations' activity," in *2018 21st International Symposium on Wireless Personal Multimedia Communications (WPMC)*. IEEE, 2018, pp. 600–605.

[22] Z. Wang and J. Huang, "Research of power supply and monitoring mode for small sites under 5g network architecture," in *2018 IEEE International Telecommunications Energy Conference (INTELEC)*. IEEE, 2018, pp. 1–4.

[23] Huawei, "5G Spectrum - Public policy position," 2017, accessed 20/10/2020. [Online]. Available: https://www-file.huawei.com/-/media/CORPORATE/PDF/public-policy/public_policy_position_5g_spectrum.pdf

[24] A. Gatherer, P. Dent, S. Bhadra, and R. Vedantham, "Self-optimizing networks (SON): doing more with less," *White paper*, 2009.

[25] 3GPP, "Management and orchestration;5G performance measurements," 3rd Generation Partnership Project (3GPP), Technical

Specification (TS) 28.552, 2018, version 15.0.0.
[26] ——, "Evolved Universal Terrestrial Radio Access (E-UTRA); Layer 2 - Measurements," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.314, 2018, version 15.2.0.
[27] H. D. Trinh, L. Giupponi, and P. Dini, "Mobile traffic prediction from raw data using lstm networks," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2018, pp. 1827–1832.
[28] S. Rostami, H. D. Trinh, S. Lagen, M. Costa, M. Valkama, and P. Dini, "Wake-up scheduling for energy-efficient mobile devices," *IEEE Transactions on Wireless Communications*, 2020.
[29] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice, 2nd edition*. OTexts Melbourne, Australia. OTexts.com/fpp2, 2018.
[30] S. E. Said and D. A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, vol. 71, no. 3, pp. 599–607, 1984.
[31] A. P. Iyer, L. E. Li, and I. Stoica, "Automating diagnosis of cellular radio access network problems," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 2017, pp. 79–87.
[32] E. W. Weisstein, "Beta distribution," 2003.
[33] A. K. Gupta and S. Nadarajah, *Handbook of beta distribution and its applications*. CRC press, 2004.
[34] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, ser. Continuous Univariate Distributions. Wiley & Sons, 1994, no. v. 2.
[35] P. Frenger and R. Tano, "More capacity and less power: How 5g nr can reduce network energy consumption," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*. IEEE, 2019, pp. 1–5.
[36] S. Ahmadi, *5G NR: Architecture, Technology, Implementation, and Operation of 3GPP New Radio Standards*. Academic Press, 2019.
[37] R. J. Hyndman, Y. Khandakar *et al.*, *Automatic time series for forecasting: the forecast package for R*, 2007, no. 6/07.
[38] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.
[39] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," *International journal of forecasting*, vol. 20, no. 1, pp. 5–10, 2004.
[40] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Management science*, vol. 6, no. 3, pp. 324–342, 1960.
[41] A. M. De Livera, R. J. Hyndman, and R. D. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1513–1527, 2011.
[42] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, "An overview and comparative analysis of recurrent neural networks for short term load forecasting," *arXiv preprint arXiv:1705.04378*, 2017.
[43] Tensorflow, "An end-to-end open source machine learning platform," "https://www.tensorflow.org/", 2019.
[44] B. Jeong, S. Shin, I. Jang, N. W. Sung, and H. Yoon, "A smart handover decision algorithm using location prediction for hierarchical macro/femto-cell networks," in *2011 IEEE Vehicular Technology Conference (VTC Fall)*. IEEE, 2011, pp. 1–5.
[45] T. Bilen, B. Canberk, and K. R. Chowdhury, "Handover management in software-defined ultra-dense 5g networks," *IEEE Network*, vol. 31, no. 4, pp. 49–55, 2017.
[46] S. Xu, A. Nikravesh, and Z. M. Mao, "Leveraging context-triggered measurements to characterize lte handover performance," in *International Conference on Passive and Active Network Measurement*. Springer, 2019, pp. 3–17.
[47] LTE capacity monitoring. [Online]. Available: https://www.slideshare.net/KlajdiHusi/lte-capacity-monitoring?from_action=save
[48] L. Saker and S. E. Elayoubi, "Sleep mode implementation issues in green base stations," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2010, pp. 1683–1688.
[49] Huawei unveils PowerStar solution at HAS 2018 to help reduce carbon footprint of mobile communications networks. [Online]. Available: https://www.huawei.com/en/press-events/news/2018/4/Huawei-PowerStar-Solution
[50] Ericsson radio system solutions energy efficiency. [Online]. Available: https://www.ericsson.com/en/portfolio/networks/ericsson-radio-system/radio-system-solutions/energy-efficiency
[51] R. Lee, D. Pinner, K. Somers, and S. Tunuguntla. The case for committing to greener telecom networks. [Online]. Available: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-case-for-committing-to-greener-telecom-networks
[52] Y. Li, Y. Zhang, K. Luo, T. Jiang, Z. Li, and W. Peng, "Ultra-dense hetnets meet big data: Green frameworks, techniques, and approaches," *IEEE Communications Magazine*, vol. 56, no. 6, pp. 56–63, 2018.
[53] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE communications surveys & tutorials*, vol. 17, no. 2, pp. 803–826, 2015.
[54] M. Feng, S. Mao, and T. Jiang, "Boost: Base station on-off switching strategy for green massive mimo hetnets," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7319–7332, 2017.
[55] F. Han, S. Zhao, L. Zhang, and J. Wu, "Survey of strategies for switching off base stations in heterogeneous networks for greener 5g systems," *IEEE Access*, vol. 4, pp. 4959–4973, 2016.
[56] S. Morosi, P. Piunti, and E. Del Re, "A forecasting driven technique enabling power saving in lte cellular networks," in *2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2013, pp. 217–222.
[57] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735
[58] A. Azari, P. Papapetrou, S. Denic, and G. Peters, "Cellular traffic prediction and classification: a comparative evaluation of lstm and arima," in *International Conference on Discovery Science*. Springer, 2019, pp. 129–144.
[59] A. Albanna and H. Yousefi'zadeh, "Congestion minimization of lte networks: A deep learning approach," *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 347–359, 2020.
[60] A. Mozo, B. Ordozgoiti, and S. Gómez-Canaval, "Forecasting short-term data center network traffic load with convolutional neural networks," *PLOS one*, vol. 13, no. 2, p. e0191939, 2018.
[61] X. Peng, B. Bai, G. Zhang, Y. Lan, H. Qi, and D. Towsley, "Bit-level power-law queueing theory with applications in lte networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

Gopika Premsankar received the Ph.D. degree in Computer Science from Aalto University, Espoo, Finland in 2020. She is a postdoctoral researcher at Ivey Business School in Western University, Canada. Her research is supported by a post doc-pooli grant from the Finnish Cultural Foundation. Her research interests include edge computing, wireless communications and energy-efficient networking.

Guangyuan Piao received the Ph.D. degree in Engineering and Informatics from the National University of Ireland Galway in 2018. He is a Lecturer with the Department of Computer Science, Maynooth University, Ireland. Before joining Maynooth University, he was a research scientist at Nokia Bell Labs. His main research interests include User Modeling, Recommender Systems, Knowledge Graphs, and AIOps.

Patrick K. Nicholson received the Ph.D. degree in computer science from the David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Canada, in August 2013. He was a Post-Doctoral Researcher with the Algorithms and Complexity Group, Max-Planck-Institut für Informatik in Saarbrücken, Germany, from September 2013 to February 2015. He joined Nokia Bell Labs, Dublin, in March 2015. His research interests span the areas of highly optimized data structures, machine learning, graph algorithms, and the applications of techniques in his research areas to solve problems in scalable distributed systems.

Mario Di Francesco received the Ph.D. degree in information engineering from the University of Pisa, Pisa, Italy, in 2009. He is an Associate Professor with the Department of Computer Science, Aalto University, Espoo, Finland. His current research interests include IoT, pervasive computing, mobile networking, and performance evaluation. Dr. Di Francesco was a recipient of the Best Paper Award at the UbiComp Conference in 2014 and at the International Conference on the IoT in 2015. He is an Area Editor of the Pervasive and Mobile Computing Journal. He was the Co-Chair of the Third IEEE Workshop on the IoT:Smart Objects and Services in 2014.

Diego Lugones (Member, IEEE) is the Head of the Application Platforms and Software Systems Department at Nokia Bell Labs in Dublin, Ireland, where he leads the research on software systems and technology prototyping of emerging cloud infrastructures with a focus on artificial intelligence (AI) enabled management for predictive and multi-variable orchestration. Before joining Nokia Bell Labs, he worked with the IBM Exascale Research team, on topics related to data center interconnects, specifically on hybrid optical/electronic-circuit/packet-switching network architectures. Diego received the M.S. and Ph.D. degrees in High Performance Computing from the Autonomous University of Barcelona, Spain, and the B.S. degree in engineering from La Plata National University, Buenos Aires, Argentina. From 2006 to 2009, he was a Professor of computer architecture at the Autonomous University of Barcelona. He also worked at the Exascale Computing Lab, HP Labs, Barcelona, in topics related to simulation frameworks for communication analysis.