

Energy Usage Forecasting for LTE: A Network-Wide Traffic Measurements Study

Endre H. Hjort Kure^{1,4,*}, Paal Engelstad³, Sabita Maharjan^{2,4}, Stein Gjessing⁴,
Xing Zhang⁵, and Yan Zhang⁴

¹Simula Research Laboratory, Norway

²Simula Metropolitan Center for Digital Engineering, Norway

³Department of Technology Systems, University of Oslo, Norway

⁴Department of Informatics, University of Oslo, Norway

⁵Wireless Signal Processing & Networks, Beijing University of Posts & Telecommunications, China

*Corresponding author: endre.hjort.kure@simula.no

Abstract—Energy usage in LTE base stations are driven by spectral efficiency and traffic. To predict the energy usage these parameters must be forecasted. In this work we analyse hourly measurements collected from more than 12000 base station cells spread across more than 3700 base stations over the course of one month. We show that the two parameters are very weakly correlated, and therefore we investigated them separately. Further, we evaluated the possible gains for advanced prediction methods using a large scale search for individually fitted time series (SARIMA models) for each base station. In total, we examined and evaluated approximately 31000 time series models from an identified group of 4 million potential models. We found that the spectral efficiency measurements can be represented fairly well with time series models, with only an average 6.5% relative error. The time series models have to be individually adapted for each base station as the unsupervised clustering showed that each cluster’s members have a wide variety of best fitted models. However, for traffic the time series models have relative high prediction errors, and we believe there is a potential for new methods to improve the forecasts.

I. INTRODUCTION

Although mobile broadband traffic is an integrated part of society, there are few available data sets for research. This is due to either privacy concerns or the operator’s commercial interests. The lack of large scale empirical data sets imply that most models are tested against small scale local measurements or simulations, which neither may reflect the average real world conditions.

Mobile broadband consumes more energy than ever, and for 5G there is a focus on energy usage in addition to increased throughput [1]. Energy models for LTE [2] and 5G [3] base stations consist of fixed and load dependent energy costs. The latter is driven by the number of resource blocks used [4], which is a product of both traffic and spectral efficiency. Forecasting of resource usage is a natural aspect of optimal network management. Thus, forecasting models need to address traffic and spectral efficiency.

Closely related to forecasting is clustering. If clusters exist, it could simplify forecasting as models could be adapted to aggregated clusters rather than individual base stations. One LTE measurement study found clusters in the base stations’ traffic patterns [5]. Similar clustering studies are yet to be done for spectral efficiency measurements. In our study we investigated clustering of both traffic and spectral efficiency.

There is a clear identified need for network-wide studies of LTE base stations. In this paper we present measurements from more than 4100 LTE bases stations, gathered over the course of 9 months on an hourly basis. The measurements constitute over 100 GB of data that contain both values for spectral efficiency and traffic.

We make the following four contributions. First, we present a new data set of LTE base station measurements that is large both in number of base stations and duration of the sampling period. The data set augments the relative small set of LTE measurement studies. Second, we show that spectral efficiency and traffic are very weakly correlated. Energy models must therefore incorporate both aspects as one cannot be derived from the other. Third, we found in this data set that clustering had limited value both for the traffic and the spectral efficiency measurements. Fourth, we provide an assessment for the potential of more advanced forecasting models both for traffic and for spectral efficiency. The base stations are individually fitted with separate time series and we derive the prediction errors for each base station. The prediction errors are used to understand the achievable improvements that other more advanced forecasting methods, such as deep learning, can provide.

The paper is structured in the following way. First, Section II presents relevant studies. Then, Section III describes the measurements in detail. The clustering approach is presented in Section IV and the time series models are presented in Section V. Then, in Section VI the time series models per base station within each cluster are evaluated. Finally, Section VII presents lessons

learned and the potential for more advanced forecasting models. The conclusion is also presented in this section.

II. PREVIOUS MEASUREMENT STUDIES

Specific LTE traffic studies are scarce. Alabaladejo et al. derived a link layer model based on measurements in Dublin City [6]. Huang et al. did an experimental setup for testing web applications' traffic and associated energy usage over LTE [7]. The analyses were extended by Foddis et al. focusing on application behaviour [8]. These studies have the problems of either being controlled experiments or limited to a few base stations, and may not reflect the average conditions in the real world. Wang et al. presented one of the largest studies on LTE traffic patterns by collecting traffic data from more than 9000 base stations in Shanghai, China over the course of one month [5]. The work was later extended by Xu et al., adapting time series models to forecast traffic [9]. Trace driven investigation of offloading from LTE to other cellular technologies was published by Li et al. based data from Shanghai Unicom [10].

III. MEASUREMENT OVERVIEW

The measurements consist of hourly values averaged over resource blocks. The sampling was done for 13008 mobile cells connected to 4153 eNodeBs during the period of January 1st to September 31st 2015. The eNodeBs are referred to as base stations since all cells connected to the same eNodeB had the same coordinates. All measurements came from LTE macro base stations operated by the same telecom operator. The majority (83%) of cells had a bandwidth of 15 MHz, while the rest operated on 20 MHz. All the base stations were located in the vicinity of Suzhou, Jiangsu, China. The number of cells per base station ranged from 1-8 cells with a median of 3 cells. The median height was 32 meters. The traffic was measured as the average downlink traffic over the Packet Data Convergence Protocol (PDCP) interface and given in Mbps, while spectral efficiency was measured as the share of different Channel Quality Indicator (CQI) index values within each hour. For compatible values, the CQI index was translated to its information bits per symbol equivalent [11], and the weighted average value was used.

Fig. 1 shows that average traffic per cell grew with 313%, while the number of new cells grew with 67% during the same period. A trough in the average traffic was observed in February, which is during the celebration of the Chinese New Year, one of the largest yearly travelling periods in China¹. The increase in number of cells over the entire period suggests that the operator was expanding its network. As measurements are aggregated values per cell, the data set is alleviated from any privacy issues as neither caller ID nor traffic content were measured.

¹<https://blogs.wsj.com/chinarealtime/2014/01/27/map-visualizes-chinese-new-year-migration/> visited 21.06.18

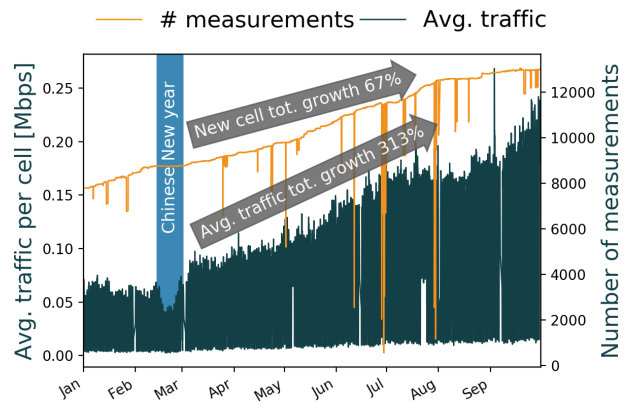


Fig. 1. Number of cells with measurements are given by the orange line, while the green line denotes average traffic per cell with measurements. The growth is calculated for the entire period. The trough in traffic that is during the Chinese New Year is marked with blue.

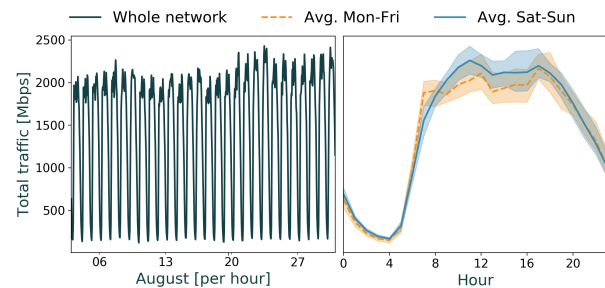


Fig. 2. The plot to the left shows the aggregated traffic for the selected period. The other plot shows the aggregated traffic for weekdays (Monday to Friday) and weekends (Saturday and Sunday) along with the interval of values per hour.

A. Sample selection and observations

We focused our analysis on a four week sampling during August (08.03.2015 - 08.30.2015) for the three following reasons. First, other comparable studies [5, 9] from areas in relative close geographical proximity had investigated the selected time interval the previous year, and by selecting a similar time frame the results would be comparable. Second, limits in available data processing capability allowed us to only work on a sub-selection of the available data. Third, the selected period was relative stable with relative low growth and few measurements missing as seen in Fig. 1.

In Fig. 2, a recurring night-day traffic pattern can be observed. The same pattern is seen in the spectral efficiency plot in Fig. 3. In the same figure a weekly pattern can be observed. Both plots show a difference between weekdays and weekends, but not at the same scale as found for traffic in [5]. The selected period contained measurements from 12411 cells distributed on 3859 base stations.

B. Traffic and spectral efficiency

The spectral efficiency and the traffic had an average traffic weighted correlation of 3.3% as given in Fig. 4.

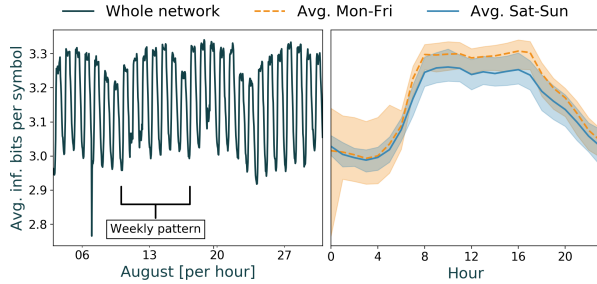


Fig. 3. The plot to the left shows the mean information bits per radio symbol for the selected period. The other plot shows the values for weekdays (Monday to Friday) and weekends (Saturday and Sunday) along with the interval of values per hour.

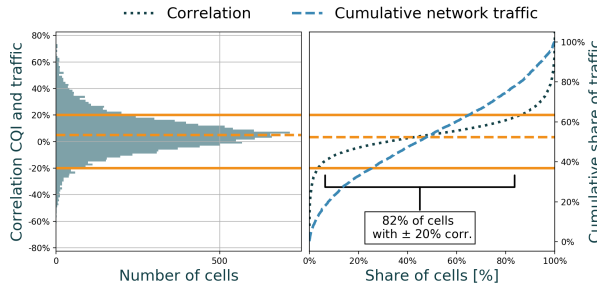


Fig. 4. The plot to the left is a histogram showing correlation between measurements per cell. The right side plot shows the correlation per cell along with the cumulative traffic carried by the respective cell over the selected period. The orange lines show the limit of $\pm 20\%$ correlation, while the dotted orange line shows the traffic weighted average of 3.3%

The correlation was calculated per cell in the selected period, and in most cells the two features are very weakly correlated. There is a bias, as the measurements were event based (measured only when resource blocks were sent) and therefore only measured when there is traffic. The two parameters might therefore be even less correlated than the measurements show, as the available spectrum efficiency might change even though traffic is zero. The lack of correlation strengthens our initial claim that spectral efficiency and traffic must be investigated together to understand resource usage in the network. In some cells, the spectral efficiency and traffic were highly correlated, both negatively and positively, but these cells represent only a small fraction of the cells as the traffic weighted majority (82 %) had correlation within $\pm 20\%$.

IV. UNSUPERVISED CLUSTERING

Researchers have found that base stations can be categorized based on traffic patterns [5]. These patterns describe the base stations' environment, such as residential and office areas. Therefore we investigated if it was possible to cluster the base stations based on traffic behaviour. Since Fig. 4 showed that the traffic and the spectral efficiency were very weakly correlated, the clustering was done separately in both dimensions. The measurements were preprocessed, enabling the use

of agglomerative hierarchical clustering. The clustering extracted the generic patterns, making it possible to bundle together the base stations that have similarities in their measurements.

A. Data preprocessing for clustering

The measured traffic x_{it} is given per base station $i \in \mathcal{I}$ per time $t \in \mathcal{T}$ with $X_i =: [x_{it} \ t \in \mathcal{T}]$ denoting the time series of each base station. As seen in Fig. 1, some hours on certain dates lack measurements. These missing values, x_{it}^{int} , were interpolated as given by Eq. (1). With the apparent 24 hour cyclic pattern, shown both in Fig. 2 and Fig. 3, the interpolated values were based on four measurements. The four measurements were the hours before and after the missing value, along with the measured values the day before and after. The interpolation was done only on measurements leaving out any approximated values. Hence, two consecutive missing values are both interpolated based only on three values.

$$x_{it}^{int} = \frac{x_{i(t-24)} + x_{i(t-1)} + x_{i(t+1)} + x_{i(t+24)}}{4} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \quad (1)$$

B. Agglomerative hierarchical clustering

Each time series was normalized with z-score as in (2) for equal comparison, giving z_{it} . This allows for clustering on the relative traffic pattern instead of the absolute one. $\sigma(X_i)$ is the variation, and $\mu(X_i)$ is the mean of the time series. Feature scaling was also tested, but did not lead to meaningful clustering.

$$z_{it} = \frac{x_{it} - \mu(X_i)}{\sigma(X_i)} \quad \forall t \in \mathcal{T}, i \in \mathcal{I} \quad (2)$$

All the base stations with missing values after interpolation were removed so that the distance measures could be calculated equally for all base stations. This resulted in that 2.12% of the base stations were discarded, leaving a total of 3777 base stations to be clustered. Agglomerative hierarchical clustering was used to cluster the base stations both on the traffic and the spectral efficiency measurements. A similar approach have provided good results in previous work on traffic clustering [5]. Euclidean distance was used as a distance metric, but other Minkowski metrics such as Manhattan was also tested, without resulting in meaningful clustering. This suggests that the base stations follow the same overall pattern, as the Euclidean distance punishes deviation on a single value stricter than the Manhattan distance. Several linking methods were tested, such as nearest, average and farthest linking. Only the latter lead to meaningful clustering. The same approach was used for the spectrum efficiency measurements. This resulted in removal of 2.25% of the sample's base stations, clustering a total of 3772 base stations.

To evaluate the cluster size, both the Davies-Bouldin Index (DBI) and the Silhouette Index (SI) were used [12]. Both evaluation criteria measure the relative difference between the intra- and inter-distance of the clusters,

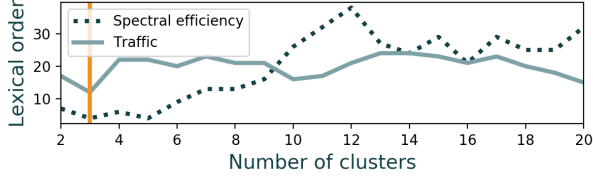


Fig. 5. The lexical order for optimal cluster size for the traffic and the spectral efficiency. The lexical order is the sum of the relative positions of both DBI and SI. The orange vertical line denotes the optimal cluster size.

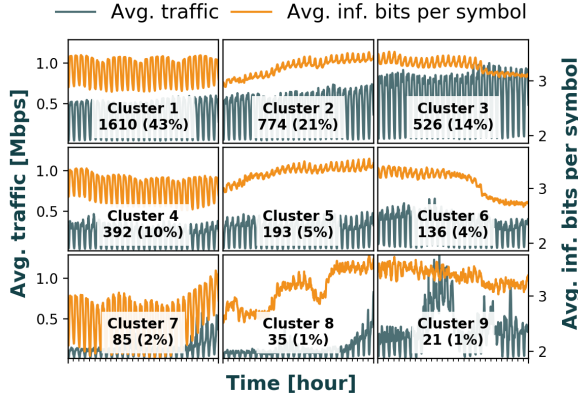


Fig. 6. The various sub-clusters for traffic and spectral efficiency. All the base stations in each row belong to the same traffic cluster (Clusters 1 to 3 from the top to bottom), while all the base stations in the same column are in the same spectral efficiency cluster (Clusters 1 to 3 from left to right). Only mean values of base stations in a cluster are plotted. The cluster's size and its size's relative share of the total number of base stations are stated for each sub-cluster.

but in a different manner [12]. We used the combined lexical order of the two criteria to determine the optimal cluster size as shown in Fig. 5. As can be seen, the optimal cluster size is three clusters for traffic, while for spectral efficiency both three and five clusters are equally good candidates. As it is better to work with fewer categories, three clusters were chosen for spectral efficiency.

The optimal identified clusters created a set of sub-clusters as shown in Fig. 6. The sub-clusters were created by combining the two clustering results of traffic and spectral efficiency. Since time series were removed if data could not be interpolated, the sub-clusters lack five time series that could not be mapped back to clusters in both dimensions.

V. TIME SERIES MODELS

A time series model family such as Seasonal Autoregressive Integrated Moving Average (SARIMA) has proven to be effective on forecasting CQI values for satellite based LTE [13] and terrestrial based LTE traffic [9]. The first study was based on simulations while the latter was done only for two base stations. Hence, it is yet to be determined to what degree a SARIMA model is applicable to different base stations.

A. Data preprocessing for SARIMA models

The time series of each base station has to be stationary in both variance and mean, in order for a model to be fitted to the measurements and used for forecasting [14]. The heteroscedasticity (variance in variance) was removed with a BoxCox power transformation [14] as shown in (3). The transformation requires that the time series is strictly positive, and since some measurements were zero, a small constant C was added to x_{it} , such that $x_{it}^* = x_{it} + C$. X_i^{trans} is the power transformed time series of X_i . The value of λ_i was found by maximum likelihood estimation. Various approaches can be used to set C , but we chose $C=1$ as it maps zero back to zero.

$$x_{it}^{trans} = \begin{cases} \log(x_{it}^*) & \text{if } \lambda_i = 0 \\ \frac{(x_{it}^*)^{\lambda_i} - 1}{\lambda_i} & \text{else} \end{cases} \quad (3)$$

$$\forall i \in \mathcal{I}, t \in \mathcal{T}$$

Power transformation may result stationarity of the mean [14], but it must be tested. We used the Augmented Dickey-Fuller test to determine if the mean was stationary. If the transformed time series was not stationary in its mean, it was differenced. Based on the observations in Section III differencing lags of 24 hours and 168 hours (one week) were tried. The Autoregressive Correlation Function (ACF) and Partial ACF (PACF) were used to identify possible models of the moving average (MA) and the autoregressive (AR) model parts of the ARIMA share of SARIMA. The MA part is given by the ACF lags, and the PACF gives the number of lags for the AR. Also, in case of dampening behaviour in the ACF, the time series differenced by one lag was also tested. The SARIMA model is given in the form $(p,d,q) \times (P,D,Q,S)$, where capital letter is the models equal seasonal part. The p denotes the size of the AR model, q the size of the MA model, d the number of lags the time series is differenced, and S the season used. Refer to [14] for a full description of the SARIMA model.

B. Searching and fitting candidate models

For a SARIMA model to be used for forecasting, the AR part must be stationary, and the MA part must be invertible [14]. Only models fulfilling these requirements were accepted as candidate models. To limit the model search, only permutations of the four largest values over the 95% confidence level threshold for both MA and AR, given by the corresponding ACF and PCF functions, were investigated. Only seasonal differencing was used, excluding seasonal AR and MA from the model. The time series were differenced with three difference lags [1,24,168], and within each lag the MA and AR parts were adapted for 3-4 different lags, giving a total of approximately 690 possible models per base station. Only a small fraction of the time series models fulfilled the described requirements, and a small set of models in the range of 1-6 (on average 5) were rated based on the Akaike Information Criterion (AIC). The best rated model was used per base station. As data was

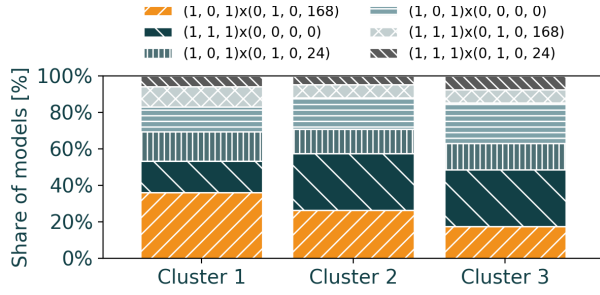


Fig. 7. The split of the various best fit models as percentile of total number of base stations in the clusters.

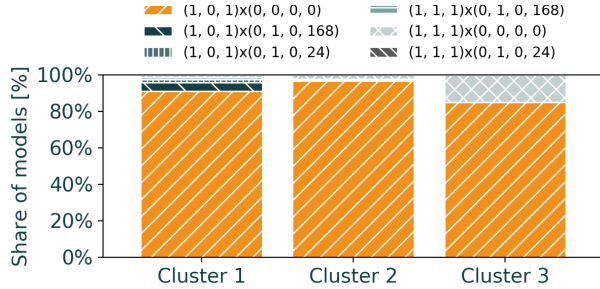


Fig. 8. The split of the various best fit models as percentile of total number of base stations in the clusters.

not interpolated, 9.7 % of the base stations that were clustered were removed, so that 3407 base stations got time series models fitted for the spectral efficiency. Restrictions in available computational power and time, limited the analyses to 2877 base stations for the traffic dimension. This means that 90% of the base stations that got a time series models fitted for spectral efficiency also had a traffic time series fitted. In total, we identified over 4 million possible models and fitted over 31000 models to the measurements.

VI. MODEL EVALUATION

As expected, for clusters based on the spectral efficiency, the majority of base stations were best fitted with a model of 1 week (168 hours) season. This model was the majority for Cluster 1. For the other clusters, the ARIMA model of (1,1,1) was the dominating one. The low presence of models with 24 hours season is surprising as they would be expected based on the day-night pattern seen in Fig. 3. As shown in Fig. 6, there are also larger shifts in the growth throughout the period, suggesting that it could be an explanatory feature.

Across all the clusters of traffic, the ARIMA model of (1,1,1) with no season is the dominating model. Fig. 8 shows the distribution of the model types. The dominance of a single model type suggests that the traffic patterns follow the same general structure, although they are clustered in different clusters.

In-sample and out-of-sample predictions are two ways to test a time series model. The latter tests how good the model is at predicting, while the former tests how good

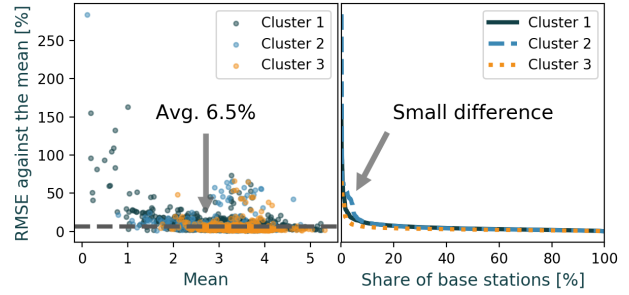


Fig. 9. The RSME against the mean for the spectral efficiency time series models per cluster.

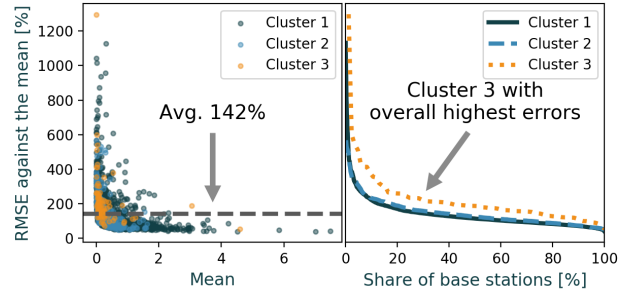


Fig. 10. The RSME against the mean for the traffic time series models per cluster.

the model is at describing the historic measurements. We started with an in-sample testing, leaving out-of-sample for future work. We did an in-sample testing for the last week in the period, comparing the root mean square deviation (RMSE) with the mean, calculating the relative prediction error. Fig. 9 presents this for the spectral efficiency time series models, while Fig. 10 shows the same calculations for the traffic time series models.

The individual fitted time series models provide a good prediction for the spectral efficiency measurements, with an average relative error of 6.5% (Fig. 9). As shown on the right side of Fig. 9, there are some variations in the 20% worst performing base stations per cluster. Fig. 10 shows that the time series models are poor models for the traffic measurements, with an average relative prediction error of 142%. There are large variations between the clusters, with Cluster 3 consistently underperforming compared to the other two. The best relative prediction error is 39%, suggesting that using time series models is a poor way to describe the traffic behaviour.

VII. CONCLUSION AND FUTURE WORK

For spectral efficiency the forecasting potential is limited as long as each base station is fitted with its own individually time series model. The poor fit of time series models to traffic patterns indicates that it could be fruitful to pursue other methods. With the large variations between the base stations, it is important that other methods are validated over large data sets. The overall spectral efficiency and traffic are very weakly

correlated, but for a small set of base station cells (18%), the correlation is larger $\pm 20\%$. The results and key learnings can be summarized as:

- 1) The time series models describe the spectral efficiency measurements in a good way. On average only a 6.5% gain can be acquired with more advanced models. Thus, the possible gain of using other prediction methods is limited.
- 2) Time series models are not good at describing the traffic, as the average relative error was 142% with the best performing base station having a relative error of 39%. Our findings are in line with results in [9], where there were also large errors in the predictions.
- 3) Within the spectral efficiency clusters there were little commonality with the forecasting models of each base station. For the traffic clusters, one time series model type dominated all the clusters, but the relative errors were too high to conclude on the benefits of clustering.
- 4) We collected data from 3859 base stations, but analysed only 3407 of them, discarding 12% of the samples due to missing data. Therefore, it is important that forecasting methods robustly handle missing values.

As mentioned earlier the measurements include substantial growth, and represent a dynamic behaviour. As previously discussed this was accounted for in the time series analyses. One could argue that the measurements are not stationary and therefore not a good representation of the typical base station. We believe that our data is actually a good representation since growth is expected, both in number of cells and cell traffic².

In our future work we will do out-of-sample predictions and test other methods for predicting traffic. The clustering was only done for August. We intend to investigate the cluster stability over longer time horizons. This would provide valuable insights into the time horizon for which measurements must be gathered to be representative.

ACKNOWLEDGEMENT

This research is supported by the project 240079/F20, funded by the Research Council of Norway.

REFERENCES

- [1] J. J. G. Andrews, S. Buzzi, W. Choi, S. V. S. Hanly, A. Lozano, A. A. C. K. Soong, and J. J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [2] G. Auer, V. Giannini, C. Desset, I. Gódor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wirel. Commun.*, vol. 18, no. 5, pp. 40–49, 2011.
- [3] A. Fisusi, D. Grace, and P. Mitchell, "Energy saving in a 5G separation architecture under different power model assumptions," *Comput. Commun.*, vol. 105, pp. 89–104, 2017.
- [4] J. Gong, J. S. Thompson, S. Zhou, and Z. Niu, "Base Station Sleeping and Resource Allocation in Renewable Energy Powered Cellular Networks," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3801–3813, 2014.
- [5] H. Wang, F. Xu, Y. Li, P. Zhang, and D. Jin, "Understanding Mobile Traffic Patterns of Large Scale Cellular Towers in Urban Environment," *Internet Meas. Conf.*, pp. 225–238, 2015.
- [6] M. B. Albaladejo, D. J. Leith, and P. Manzoni, "Measurement-Based Modelling of LTE Performance in Dublin City," in *IEEE Int. Symp. Indoor Mob. Radio Commun.*, 2016.
- [7] J. Huang, F. Quian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A Close Examination of Performance and Power Characteristics of 4G LTE Networks," *MobiSys*, pp. 225–238, 2012.
- [8] G. Foddis, R. G. Garroppo, S. Giordano, G. Proccissi, S. Roma, and S. Topazzi, "LTE Traffic Analysis and Application Behavior Characterization," in *Eur. Conf. Networks Commun.*, 2014, pp. 2–6.
- [9] F. Xu, Y. Lin, J. Huang, D. Wu, H. Shi, and Y. Li, "Big Data Driven Mobile Traffic Understanding and Forecasting: A Time Series Approach," *IEEE Trans. Serv. Comput.*, vol. 9, no. 5, pp. 1–1, 2016.
- [10] L. H. Li, W. Zhang, S. Yao, S. Xia, and X. Ji, "Analysis of LTE Traffic Offloading based on Big Data," in *Int. Symp. Commun. Inf. Technol.*, 2016.
- [11] C. Cox, *An Introduction to LTE*. Chennai, India: John Wiley & Sons Ltd, 2012.
- [12] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, "Understanding of internal clustering validation measures," in *IEEE International Conf. Data Min.*, 2010, pp. 911–916.
- [13] Y. Zheng, S. Ren, X. Xu, Y. Si, M. Dong, and J. Wu, "A modified ARIMA model for CQI prediction in LTE-based mobile satellite communications," in *Proc. 2012 IEEE Int. Conf. Inf. Sci. Technol. ICIST 2012*, 2012, pp. 822–826.
- [14] W. W. Wei, *Time Series Analysis Univariate and Multivariate Methods*, 2nd ed. United States of America: Pearson Addison Wesley, 2006.
- [15] Cisco VNI Mobile, "Cisco Visual Networking Index (VNI) Update Global Mobile Data Traffic Forecast 2016 - 2021," Cisco Public Information, Tech. Rep., 2017.

²Cisco predicts that global mobile traffic will double every other year [15]