Hybrid Prediction Model for Mobile Data Traffic: A Cluster-level Approach

Bethelhem S. Shawel[†], Tsegamlak T. Debella[◇],
Getinet Tesfaye[⋆], Yonas Y. Tefera[‡] and Dereje H. Woldegebreal[♮]
SECE, AAiT, Addis Ababa University, Addis Ababa, Ethiopia
Emails: {*bethelhem.seifu*[†], *tsegamlak.terefe*[◇], *yonas.yehualaeshet*[‡], *dereje.hailemariam*[♮]}@*aait.edu.et*, *getinet.tesfaye@ethiotelecom.et*[⋆]

***Abstract*—Mobile data consumption is rapidly growing following the ever-increasing bandwidth-hungry applications and improvements in network data rates. With the anticipated 5G right at the corner, operators are focusing on load-aware network dimensioning, optimization, and management, where traffic volume prediction plays a critical role. To this end, several researchers investigated different statistical and machine-learning models to exploit and predict the linear and nonlinear patterns that often arise due to the complexity of mobile networks and varying users' behaviors at different times and locations. In this paper, we propose a hybrid model composed of Double Seasonal ARIMA (D-SARIMA), which focuses on modeling the multi-seasonal nature of the data traffic and exploiting the residuals of DSARIMA via Long-Short Term Memory (LSTM)-based Networks. The residues contain the nonlinear component of the data. To incorporate the spatial dependency inherent in mobile data traffic collected from base stations, we used K-means clustering and considered the correlation among them. Experiments conducted with real-world data sets collected from 739 base stations for over four months, shows that our proposed hybrid model outperforms both D-SARIMA and LSTM models. The improvement emanates from capturing the double seasonality, non-linearity, and spatial dependency inherent in data traffic.**

***Index Terms*—Mobile traffic prediction, non-linear data traffic, Double SARIMA, Hybrid model, LSTM.**

I. Introduction

The continued evolution of mobile network technologies and the emergence of various services have led to an exponential growth of mobile data traffic [1]. From the operators' perspective, this growth is an opportunity that maximizes revenue. However, supporting such traffic demand, among others, requires availing infrastructure during network dimensioning phase as well as allocating sufficient network resources (e.g., bandwidth and energy) and network management solutions during network operation phase. The mobile data traffic demand has a dynamic nature that varies in time and space domains. Fig. 1 (a) and (b) shows the temporal and spatial variations of data traffic collected from the Universal Mobile Telecommunications Service (UMTS) network operator in the city of Addis Ababa, Ethiopia. Thus, accurately understanding the traffic dynamics in multiple dimensions and predicting the current and future demands is critical.

Researchers, in recent years, have made efforts to propose different mobile traffic prediction techniques. The prediction can be performed at different granularities: aggregate demand of an operator [2], demand on a cell level [3], per user demand [4], on packet level (e.g., packet arrival rate, the occurrence of burst, packet inter-arrival rate, flow rate) [6], and application-level (e.g., predicting applications with significant contribution to generating the traffic) [5].

Inherently, mobile traffic prediction can be treated as a time series prediction problem where models can be used to predict traffic demand based on available historical data. One way of broadly categorizing time series based models can be as linear statistical models (such as Box-Jenkins variants), machine learning-based models (such as Neural Networks), and a hybrid of the two models [6], [7]. In [8], linear auto-regressive integrated moving average (ARIMA) model, one of the Box-Jenkins variants, has been used to capture fixed temporal dependencies in network traffic and predict its yearly growth. To improve the ARIMA-based models on capturing

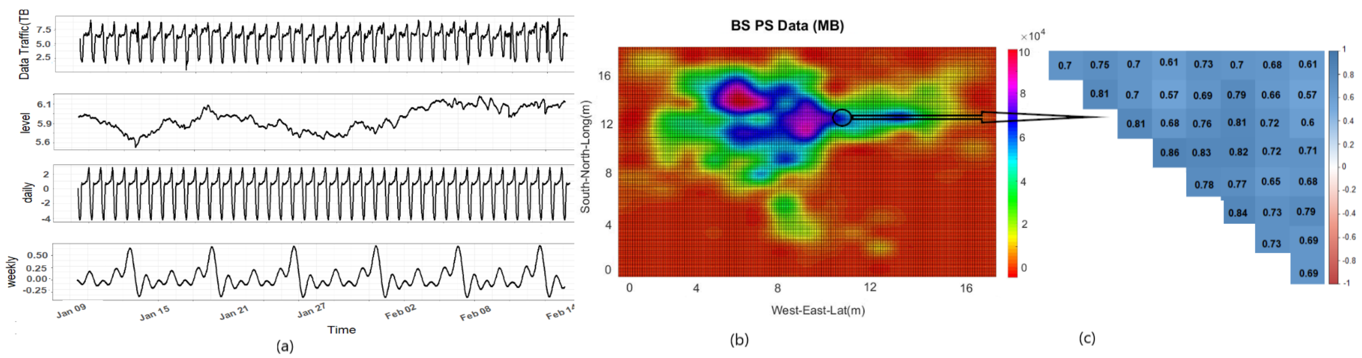

Fig. 1. (a) Temporal dynamics of data traffic decomposed to trend, and seasonal (daily and weekly) components. (b) Spatial distribution of the mobile data traffic. (c) Correlation matrix for selected neighboring base stations.

the long-term traffic repetitive patterns, [9] and [10] considered statistically decomposed components of the mobile data traffic (e.g., trends and seasonality) for prediction using Seasonal ARIMA (SARIMA) model. Generally, the linear statistical methods work well in estimating the inherent linear characteristics of the data traffic. However, due to the complexity of mobile networks and varying customers' behaviors at different times and locations, data traffic dynamics exhibits non-linear patterns and often non-stationarity which makes it difficult to be captured via those linear statistical models [3], [11].

Following advancements in machine learning techniques, prediction of mobile traffic with machine learning-based methods is proven to improve the prediction accuracy by capturing the non-linear and complex patterns inherent in data traffic [6], [11]. In [12], the authors implemented Long Short Term Memory (LSTM) network-based prediction model to capture the temporal dependencies in mobile voice and data traffic. LSTM-based deep learning model was applied in [13] to not only consider the temporal but also the spatial correlation across the entire network by analyzing traffic information from neighboring base stations. With users continuously moving within a given network, traffic patterns across neighboring base stations are correlated, and exploring both the spatial and temporal dimensions would improve the traffic prediction performance. The Double Spatiotemporal Neural Network (D-STNN) proposed in [14] used Convolutional-LSTM (ConvLSTM) and three-dimensional Convolutional Network (3D-ConvNet) structures with an encoder-decoder architecture to jointly learn the complex spatial and temporal dependencies of the mobile data traffic and provide long-term network-wide prediction. Another means to explore the spatial correlation of the data traffic by grouping together adjust and correlated base stations with similar usage patterns is considered in [3]. K-Means clustering was used in combination with Elman Neural Network and Wavelet decomposition to provide a cell-level prediction.

Another approach to explore the complex dynamics of the data traffic is by implementing the hybrid of linear statistical models (such as ARIMA) with Artificial Neural Network models [15] and Wavelet analysis [16] to capture the linear and non-linear parts, respectively, of the data and combine the two results to obtain the final mobile network traffic flow prediction. This two-step prediction approach merges the positive traits of those models and has the benefit of improving computational complexity, model interpretability, and prediction accuracy.

This paper proposes a hybrid model of Double Seasonal ARIMA (D-SARIMA), which focuses on modeling the multi-seasonal nature of the mobile data traffic, and Long-short Term memory (LSTM)-based Networks to learn non-linearities by further exploiting the residuals of D-SARIMA. To evaluate its prediction performance, a comparative analysis is done with SARIMA models and LSTM-based networks at base station level with data set collected from 739 UMTS base stations (eNodeB) in the city of Addis Ababa, Ethopia. In addition, to further capture spatial dependecy of the data, we considered a cluster-level comparison with K-means as a clustering approach.

The remainder of this paper is organized as follows. Section II presents the mobile data traffic analysis. The different traffic prediction techniques and the hybrid model are discussed in Section III. Section IV presents the data pre-processing and demonstrates the experimental results. Finally, Section V concludes this paper.

II. Mobile Data Traffic Analysis

A. Data Set Description

The mobile data traffic analyzed in this paper is obtained from a UMTS network operator in Addis Ababa, Ethiopia. The data is collected from January 2019 to April 2019 with a temporal resolution of 1 hour. Specifically, the dataset encompasses eNodeBs‘ location information, observed cell-level downlink traffic, and active users with the corresponding timestamp. Fig. 1 illustrates the temporal and spatial characteristics of mobile data traffic.

The temporal variation in aggregated data traffic can be observed in Fig. 1 (a) showing an increasing trend in traffic volume and exhibiting daily periodical behavior with relatively high demand nearly at midnight and lowest-demand during early mornings. The variation of data traffic on different days throughout the week also created new repetitive patterns on a weekly basis indicating the presence of double seasonalities in mobile traffic.

There is also significant traffic load variation on the network, as shown in Fig. 1 (b) with a significant degree of correlation among neighboring base stations illustrated by high values (greater than 0.6) of the Pearson correlation coefficient Fig. 1 (c). To further explore the spatial dependency inherent in mobile data traffic collected from base stations, we consider a clustering approach based on their temporal pattern.

B. K-means Clustering

Several techniques can be used to map a group of spatially distributed base stations with complementing traffic patterns together and also identify unique temporal patterns [3]. Among the available alternatives, we have selected K-means for main reasons as it is very fast as compared to hierarchical clustering techniques and provides less number of hyperparameters, i.e., number of cluster K, as compared to model-based clustering mechanisms [17].

The K-means approach aims at dividing a data sets into K disjoint clusters centered around their means or centroids. To obtain the clusters, K-means iteratively updates cluster members, means, or centroids, where most of the time means or centroids are initialized by randomly selecting one of the data set to be a centroid or mean. On the other hand, membership in a cluster is given based on the closeness of a data set to a cluster's mean or centroid; i.e., similarity of data traffic pattern of a particular base station to the mean traffic pattern. Afterward, means or centroids are updated by taking the average of the identified members. In this context, deploying K-means for time series data sets is faced with a critical question of identifying the optimal value of K. To this end, we have utilized inter-cluster inertia, which measures the closeness of a data sets to the cluster mean or centroid. Furthermore, following the discussion in [18], we have acknowledged the impact of temporal distortion on K-means with regards to cluster membership identification and the optimal estimation of cluster average which is still a challenge to be tackled. However, we found including the proposed solutions in [18] to our paper to be computationally costy. Thus, even if the solution provides better cluster identifications, we have avoided it in this work; and we aim to investigate the impact of the proposed solutions in [18] on prediction accuracy in future work.

With that, the optimal number of clusters, K, value is estimated by iteratively observing the inter-cluster inertia, and 5 clusters were obtained. The traffic pattern for the respective five clusters is shown in Fig. 2, indicating a presence of diversity and similarity of traffic usage among them. Similarly, the spatial correlation of the five clusters is presented in Fig. 3. In a city where the different functional areas (residential, business, or entertainment areas) are mixed, a low correlation among the clusters is expected; however, the close to zero correlation in Fig. 3 might imply peculiar usage behaviors or the sparse population distribution in the network load.

III. Traffic Prediction Techniques

A. Double Seasonal ARIMA

ARIMA is a popular statistical model used to capture the stationarity property of time-series data

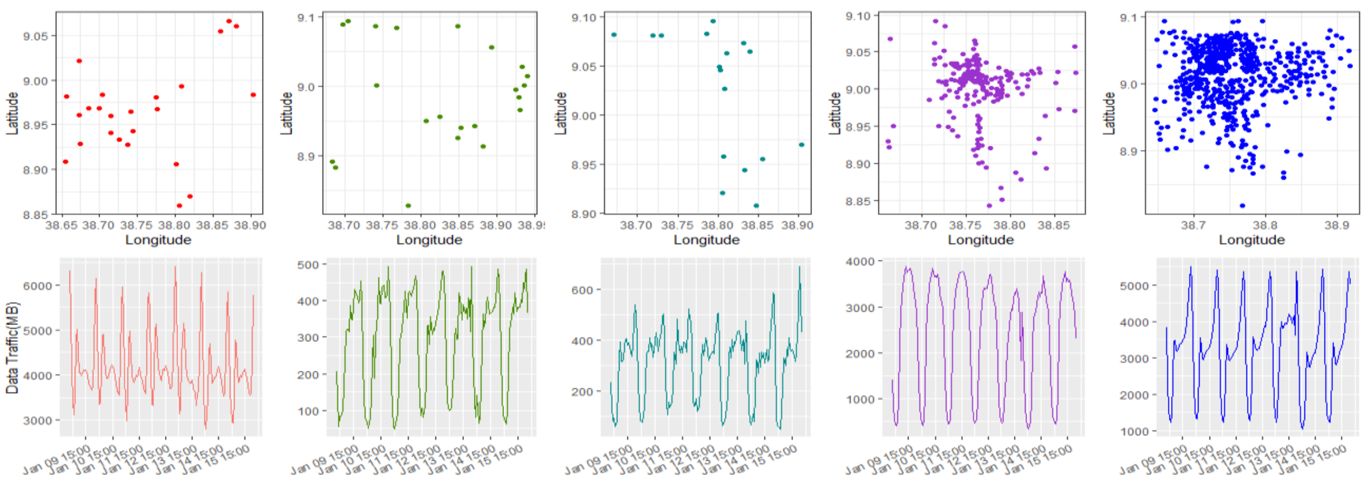

Fig. 2. Clustered base stations with their corresponding centroids.

as a function of its sequentially lagged variables as well as error terms. When there exist seasonal components in the data, it is possible to treat the seasonal and non-seasonal parts with a general multiplicative SARIMA model [19].

To capture the double (daily and weekly) seasonalities explained in Section II of the mobile data traffic, the SARIMA model can be expressed as SARIMA$(p, d, q) \times (P_1, D_1, Q_1)_{s1} \times (P_2, D_2, Q_2)_{s2}$ where the order of regression (ϕ) and moving average (Θ) coefficients for the non-seasonal and seasonal parts of the model are represented by $(p, P_{(.)})$ and $(q, Q_{(.)})$, respectively. The parameters d and D are also used to represent the differencing that can be applied one or more times to eliminate the trend, and $s_{(.)}$ seasonalities, and make the time series data stationary.

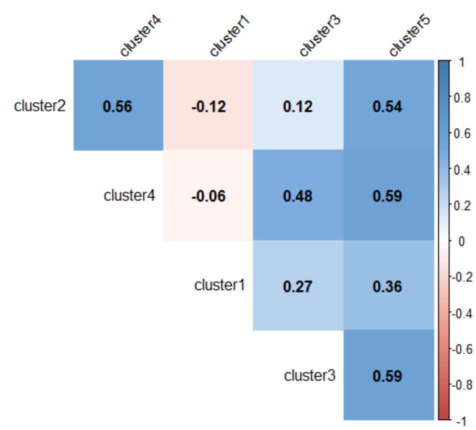

Fig. 3. Correlation matrix for the five clusters.

Assuming a polynomial that has a factor $(1 - L)$ of multiplicity, the Double SARIMA (D-SARIMA) model is formulated as [19]:

$$(1 - \sum_{i=1}^{p} \phi_i L^i)(1 - \sum_{j=1}^{P1} \phi_j L^{js1})(1 - \sum_{k=1}^{P2} \phi_k L^{ks2})$$
$$((1 - L)^d (1 - L^{s1})^{D1} (1 - L^{s2})^{D2} (X_t - \mu))$$
$$= (1 + \sum_{i=1}^{q} \theta_i L^i)(1 - \sum_{j=1}^{Q1} \Theta_j L^{js1})(1 - \sum_{k=1}^{Q2} \Theta_k L^{ks2}) \varepsilon_t \tag{1}$$

where X_t is the aggregated traffic consumption representing the non-stationary time-series and ε_t is the error term at time t.

In order to incorporate the impact of spatial dependency with SARIMA models, we can consider the aggregated traffic from different cluster as exogenous variables (independent variables). Evaluating the cross-correlation among clusters will help to identify which cluster data to be considered as external variable.

B. LSTM

Another predictor that is widely considered to learn and estimate complex multi-dimensional characteristics of the mobile data traffic is a Recurrent Neural Network (RNN). As one variant of RNN, LSTM is suitable for time series prediction and is capable of capturing the long-range temporal information by using memory cells [20]. Which input to process, whether to update the memory cell, and whether to create an output is controlled by three gates in LSTM memory block, namely the input gate, forget gate and output gate, respectively.

For a given sequential inputs p_t, hidden layer h_{t-1} and previous memory cell state c_{t-1}, the LSTM output for time step t is given as [20]

$$O_t = \sigma[W_o p_t + U_o h_{t-1} + Vo(ft \odot c_{t-1}) + (\sigma(W_i p_t + U_i h_{t-1} + V_i c_{t-1} + b_i) \odot g(W_i p_t + U_c h_{t-1} + b_i)) + b_o] \quad (2)$$

where $\odot$ denotes element-wise multiplication and forget gate f_t is equated as $\sigma(W_f p_t + U_f h_{t-1} + V_f c_{t-1} + b_f)$. For a given data traffic consumption observed at base station level for a time interval T, the input sequence is a 2-D dataset (i.e., $P = p_1, p_2, p_3, ..., p_T$), and for mobile data traffic measured over N base stations or represented in terms of K clusters, the input sequence will be 3-D (i.e., $\boldsymbol{P} = P_1, P_2, P_3, ..., P_N$). The nonlinear activation functions are represented by $g_{(.)}$ and $\sigma_{(.)}$ usually denoting the *Relu/tanh* and *sigmoid* functions, respectively. $W_{(.)}$, $U_{(.)}$, and $V_{(.)}$ are weight matrices that are adjusted during model training by minimizing the square loss function (in this paper) and $b_{(.)}$ are bias vectors.

C. Proposed Hybrid Model

To improve prediction accuracy and effectively handle the linear and non-linear dynamics of the mobile data traffic, we propose a hybridize model by leveraging the benefits of both SARIMA and LSTM. Fig. 4 illustrates the proposed model consisting the data processing part and a hybrid predictor part. The model's first part includes pre-processing part for "clean-up", clustering part as well as cluster correlation analyzing part.Whereas, hybrid predictor part *blends* the prediction output from D-SARIMA and LSTM-based network to provide combined prediction as illustrated in Fig. 4.

While conducting a prediction for a particular cluster, its correlation with the remaining four clusters is analyzed. The clusters with correlation coefficient greater than 0.5 value are assumed to be moderately correlated and are taken as external (exogenous) variables to D-SARIMA model. With that the spatial correlation among different clusters is considered as a means of improving the prediction.

The complex non-linearities that couldn't be fitted with D-SARIMA are reflected on the model residuals. For a time series data y_t the residual, r_t, from linear statistical models can be expressed as:

$$r_t = y_t - \hat{L}_t \quad (3)$$

where $\hat{L}_t$ is the estimated linear component of the data.

While keeping the temporal structure of the residual errors, LSTM-based network then used to learn additional information and provide future predictions.

Combining the outputs from the two models models depends on the different predictors considered and can be done through weighted or straightforward addition, multiplication or by ensemble averaging the prediction output or model coefficients [15]. In the proposed model, straightforward addition is considered to integrate the results from the two predictions as:

$$\hat{y}_t = \hat{r}_t + \hat{L}_t + \varepsilon_t \quad (4)$$

where $\hat{y}_t$ represented the final prediction, $\hat{r}_t$ indicating the estimation from LSTM-based network and ε_t is the error that is not captured by the hybrid model.

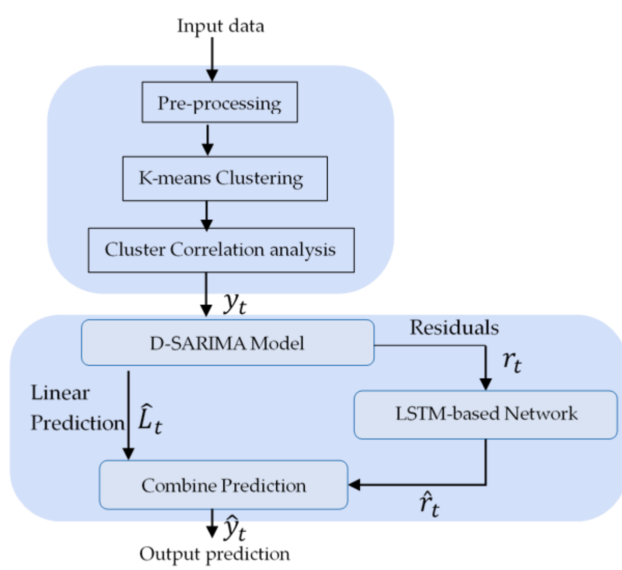

Fig. 4. Proposed hybrid model.

IV. Experimental Results

A. Data set Pre-processing and Experiment Set up

Few number of data traffic measurement values were missing due to improper data storage and retrial. Thus, we applied linear interpolation using Kalman filters prior to prediction to impute the missing values; as Kalman filter is used to find optimal estimates of the missing values by computing its conditional mean and variance up on the observed data traffic. Furthermore, to speeds up learning and faster convergence for based predictor, the traffic values are normalized into range of [0, 1] using Min-Max normalization.

To obtain the model parameters for D-SARIMA, i.e., to determine the autoregression $(p, P(.))$, moving average (q, Q(.)) and differencing orders (d, D(.)), it is necessary to investigate *Auto Correlation Function* (ACF) and *Partial Auto Correlation Function* (PACF) of the the time series. Removing the non-stationarity exhibited on the mobile data traffic as a form of increasing trend, and daily and weekly seasonalities (in Fig. 1) is essential prior to using the D-SARIMA . The ACF plot in Fig. 5 shows stationarity of the data traffic in the mean value after going through a second order differentiation process to remove the trend and seasonality. The significant sparks at lag (24, 48, ...) and at 168 on the ACF plot also confirm the daily $(s_1 = 24)$ and weekly $(s_2 = 168)$ seasonalities discussed in Section II.

After considering the correlated clusters as exogenous variables on the model, the *best-fit* model is identified as SARIMA$(1,0,2)(2,1,0)_{24}(0,1,1)_{168}$

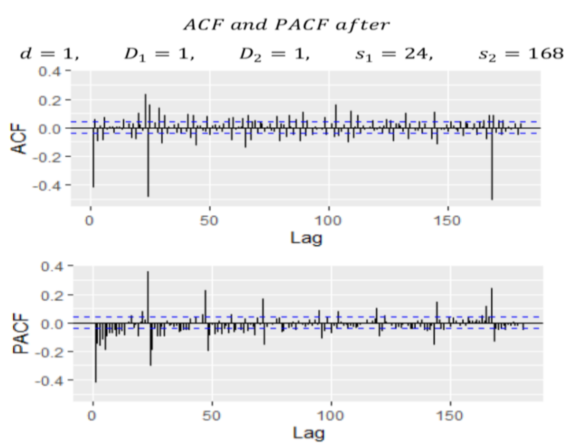

Fig. 5. ACF and PACF plot

based on minimum values of the Corrected Akaike's Information Criterion (AICc).

To capture the non-linearity on the mobile data traffic by exploiting the residuals from D-SARIMA, the LSTM-based network within the hybrid model considers 2-layers of LSTM units each with 128 and 64 hidden nodes, respectively, to form a stacked LSTM network, and *Time distributed* dense layer at the output to apply a layer to every temporal slice of an input. The $ReLU$ activation function is considered for the two LSTM layers and the $sigmoid$ activation function to restrict the prediction output within range of $[0, 1]$. For $80/10/10$ partitioned residual data for training, validation and testing, optimizing the square loss is done with *Adaptive Moment Estimation* (ADAM) optimizer, which is widely used in Neural Networks domain. the key point to note is, the values of these parameters including additional hyper-parameters such as batch size (24), epoch (100) and number of past observations $(48$ or 2 days) are determined based on experiment requirements as optimizing them was not the intent of the work. Thus, the parameters can also take different value which can impact the trade-off between the prediction accuracy and the time needed to train the network.

B. Evaluation Metrics

For the purpose of evaluating the prediction performance of the hybrid model and compare it against D-SARIMA and LSTM models, two standard prediction metrics are used: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE); calculated for prediction error e_i over n measurement points of the data traffic over space and time.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i| \tag{6}$$

C. Prediction Results and Comparison

The mobile data traffic prediction performance of aforementioned hybrid model is done considering two cases/approaches: Base Station-Level prediction approach where data traffic from a single base station is analyzed, and Cluster-level approach where the data traffic from other clusters that are moderately correlated are considered to incorporate the spatial dependency.

Table 1 shows comparison (based on average RMSE and MAE) of the hybrid, D-SARIMA, and LSTM-based models. Key observations from the base station level results are:

- The proposed hybrid model performs poorly, whereas D-SARIMA provides relatively better short-time prediction. The results also indicate the double seasonality and trend components are the dominant patterns in mobile data traffic that were better captured by the linear model.
- The LSTM model was not able to sufficiently learn the patterns (i.e., trend, seasonalities, and non-linearities) inherent both in the data or the D-SARIMA residuals to the extent that it contributes negatively in the hybrid model. This shows that linear models are good at capturing short-term dependency in the data. Possible remedies to improve the LSTM model include: increasing the data size (from the current four months), hyper-parameter optimization, or extracting additional features from other base stations are; the latter approach is used in the cluster-level prediction explained next.

As argued repeatedly, the cluster-level approach has the potential to exploit the temporal and spatial dimensions of the mobile data. The addition of the spatial dimension will, undoubtedly, add to the nonlinearity of the data but also provides more information for the LSTM to learn more. Key observations from the clusters-level results in Table 1 are:

- All three models perform better than their counterparts in base station level investigation as they exploit cluster correlation and extract multiple temporal patterns. The improvement is as high as 60%, which is significant.
- By comparison, the proposed hybrid model performs better while the D-SARIMA's performance is inferior to the two models. The LSTM captures the dynamics (the non-linearity) in the mobile data which is manifested on the improved prediction performance of the hybrid and LSTM model.

TABLE I
COMPARING THE PREDICTION PERFORMANCE ON BASE STATION-LEVEL AND CLUSTER-LEVEL APPROACH IN TERMS OF RMSE AND MAE

Approaches	Models	Evaluation Metrics	
		MAE	RMSE
Base station level	D-SARIMA	1.385	1.229
	Hybrid	1.667	1.517
	LSTM	1.408	1.237
Cluster-level	D-SARIMA	0.872	0.548
	Hybrid	0.416	0.363
	LSTM	0.617	0.548

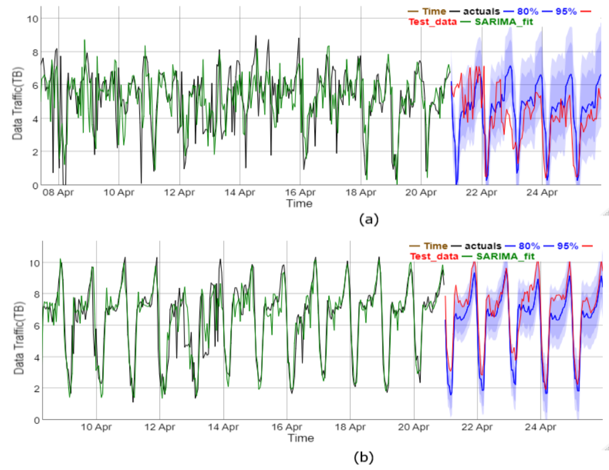

Fig. 6. Double Seasonal ARIMA model fitting and 120 hours ahead prediction considering single base station in (a) and multiple cross-correlated clusters in (b).

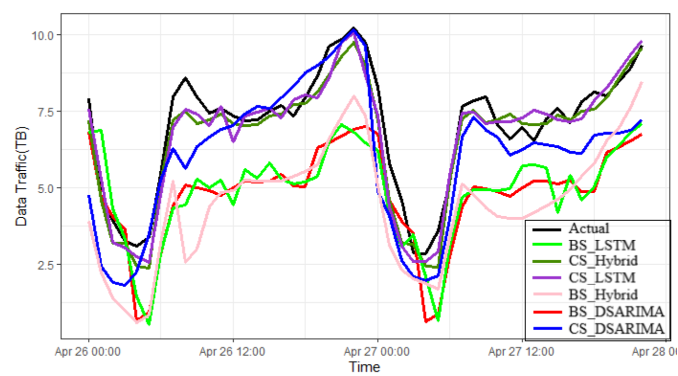

Fig. 7. 48 hours of mobile data traffic prediction performance considering base station and cluster-level approaches

The cluster-level approach benefited the linear model like D-SARIMA as the other correlated clusters' data traffic is taken as exogenous variables for the prediction of a particular cluster. See Fig. 6 to further learn the improvements in these models.

The mobile data traffic prediction for next 48 hours with the models considering both approaches is illustrated in Fig. 7. Results show that for both base station and cluster-level approaches, the prediction during low traffic load (in early morning) is comparable. However, there exists a significance difference during day-time that can associated to a user activity and spatial mobility. In addition, as the prediction time increases, despite the performance degradation of D-SARIMA, the performance hybrid model prediction remains relativity constant.

V. Conclusions

As mobile networks become more complex, and users' data traffic consumption behaviors vary over space and time, providing accurate predictions of the traffic volume gets particularly tricky. In this paper, we propose the use of hybrid model of D-SARIMA and LSTM-based network to exploit both the linear and non-linearity on the data traffic and provide accurate predictions. We considered the double (daily and weekly) seasonal patterns in our D-SARIMA and exploiting the residuals with LSTM to learn non-linearities that the linear model failed to capture. Furthermore, to explore spatial dependency inherent in mobile data traffic collected from base stations, we used k-means clustering to group base stations with complementing traffic patterns together and also identified unique temporal patterns. By evaluating the correlation among them, we considered multiple clusters together as multiple valuables to provide cluster-level prediction. The results reveal the benefits of using hybrid model and exploiting the spatial correlation. Possible extensions to this work are investigating the impact of temporal distortion on the cluster averaging, computational overhead of the hybrid model, or the impact of optimizing the hyper-parameters of the LSTM-based network on the prediction accuracy.

References

[1] Ericsson, *Sub Saharan Africa Ericsson Mobility Report*, November 2019.

[2] F. Xu et al., "Big Data Driven Mobile Traffic Understanding and Forecasting: A Time Series Approach," in IEEE Trans. Services Comput., vol. 9, no. 5, pp. 796-805, 2016. doi:10.1109/TSC.2016.2599878

[3] Y. Zang, F. Ni, Z. Feng, S. Cui, and Z. Ding, "Wavelet transform processing for cellular traffic prediction in machine learning networks.," ChinaSIP, pp. 458–462, 2015.

[4] Guangshuo Chen. "Spatiotemporal Individual Mobile Data Traffic Prediction." Technical Report. RT-0497, INRIA Saclay - Ile-de-France. 2018.

[5] D. Naboulsi, M. Fiore, S. Ribot and R. Stanica, "Large-Scale Mobile Traffic Analysis: A Survey," in IEEE Commun. Surv., vol. 18, no. 1, pp. 124-161, Firstquarter 2016. doi:10.1109/COMST.2015.2491361

[6] Azari A., Papapetrou P., Denic S., Peters G., "Cellular Traffic Prediction and Classification: A Comparative Evaluation of LSTM and ARIMA." In: Discovery Science. DS 2019. Lecture Notes in Computer Science, vol 11828. Springer, https://doi.org/10.1007/978-3-030-33778-0_11

[7] Ratnadip Adhikari, R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting", pages, 18 – 29.

[8] B. Zhou, D. He, and Z. Sun, "Traffic modeling and prediction using ARIMA/GARCH model," in Modeling and Simulation Tools for Emerging Telecommunication Networks, Boston, MA, 2006, pp. 101–121.

[9] Y. Shu, M. Yu, J. Liu, and O. W. W. Yang, "Wireless traffic modeling and prediction using seasonal ARIMA models," in IEEE International Conference on Communications (ICC), Anchorage, AK, USA, May 2003, pp. 1675–1679.

[10] Samuel Medhn, Bethelhem Seifu, Amel Salem, Dereje Hailemariam "Mobile Data Traffic Forecasting in Networks Based on SARIMA Model: The Case of Addis Ababa, Ethiopia", IEEE Africon 2017 Proceedings, pp. 285-290.

[11] Yue Xu, Feng Yin, Wenjun Xu, Jiaru Lin, Shuguang Cui,"Wireless Traffic Prediction with Scalable Gaussian Process: Framework, Algorithms, and Verification", IEEE Journal on Selected Areas in Communications,2019, PP. 1-1, 10.1109/JSAC.2019.2904330.

[12] Hoang Duy Trinh, L. Giupponi, Paolo Dini, "Mobile Traffic Prediction from Raw Data Using LSTM Networks". 10.1109/PIMRC.2018.8581000 .

[13] C. Qiu, Y. Zhang, Z. Feng, P. Zhang, S. Cui, "Spatio-temporal wireless traffic prediction with recurrent neural network," IEEE Wireless Commun. Lett., vol. 7, no. 4, pp. 554–557, August 2018

[14] Chaoyun Zhang, Paul Patras, "Long-Term Mobile Traffic Forecasting Using Deep Spatio-Temporal Neural Networks", Mobihoc '18, 2017, DOI:10.1145/3209582.3209606

[15] Dehuai Zeng, Jianmin Xu, Jianwei Gu, Liyan Liu and Gang Xu, "Short Term Traffic Flow Prediction Using Hybrid ARIMA and ANN Models," Workshop on Power Electron. and Intell. Transp. Syst., Guangzhou, 2008, pp. 621-625, doi:10.1109/PEITS.2008.135

[16] J. Li, L. Shen and Y. Tong, "Prediction of Network Flow Based on Wavelet Analysis and ARIMA Model," Inter. Conf. on Wireless Netw. and Inf. Syst., 2009, pp. 217-220, doi:10.1109/WNIS.2009.20

[17] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, Teh Ying Wah, "Time-series clustering – A decade review," Information Systems, Volume 53, 2015, Pages 16-38, https://doi.org/10.1016/j.is.2015.04.007

[18] François Petitjean, Alain Ketterlin, and Pierre Gançarski, "A global averaging method for dynamic time warping, with applications to clustering." Pattern Recogn. 44, 2011, 678–693. DOI:https://doi.org/10.1016/j.patcog.2010.09.013

[19] Box, G.E.P., G.M. Jenkins and G.C. Reinsel, "Time Series Analysis: Forecasting and Control." Fourth Edition. John Wiley and Sons, New Jersey, pp: 378.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.