Change Detection Based Segmentation and Modeling of LTE Spectrum Tenancy

Rui Zou and Wenye Wang
Department of Electrical and Computer Engineering
NC State University
Raleigh, NC 27606, USA
Emails: {rzou, wwang}@ncsu.edu

***Abstract*—The mainstay of current spectrum access grants exclusive rights to proprietary occupants who exhibit tidal traffic patterns, leading to low usage of valuable spectrum resources. To remedy this situation, Dynamic Spectrum Access (DSA) is proposed to allow Secondary Users (SUs) to opportunistically exploit idle spectrum slices left by Primary Users (PUs). The key to the success of DSA lies in SUs' knowledge on radio activities of PUs. To enhance the understanding of PU spectrum tenancy patterns, various mathematical models have been proposed to describe spectrum occupancy dynamics. However, there are still two overlooked aspects in existing studies on spectrum tenancy modeling, i.e., time-varying spectrum tenancy patterns and multiple channels within the same Radio Access Technology (RAT). To address the two issues, we apply a change detection algorithm to discover time points where spectrum tenancy patterns vary, and propose to characterize spectrum usage in a multi-channel RAT by the Vector Autoregressive (VAR) model. Through analyzing LTE spectrum tenancy data with the algorithm and the model, we validate that the segment size discovered by the online change detection method coincides with the one obtained by brute force, and VAR outperforms the widely adopted on/off model.**

I. INTRODUCTION

Driven by increasing demands on mobile data rates in various applications and network topologies [1]–[3], tremendous efforts have been made to improve spectral efficiency, such as high-order modulation and assorted MIMO (Multiple-Input and Multiple-Output) technologies [4]. Though wireless spectrum has been painstakingly sought and loaded with as much data as possible, the mainstay of current spectrum access mechanism has been proved to keep spectrum utilization low. Spectrum bands auctioned for exclusive access may have staggeringly low utilization of below 50% [5]. The unoccupied spectrum resources, or spectrum holes which are caused by tidal traffic of Primary Users (PUs), can be utilized by Secondary Users (SUs) in Dynamic Spectrum Access (DSA) systems where SUs are allowed to opportunistically access idle PU spectrum slices [6].

DSA is a fundamental improvement to current spectrum access paradigm, and the key to its success lies in SUs' knowledge on spectrum tenancy of PUs. To gain insight into PU spectrum occupancy characteristics, a plenitude of modeling studies have been conducted [7]–[9]. The statistics of occupancy time and transition time are studied based on measurements of spectrum tenancy in a large frequency span [7]. Another wide-range spectrum study proposes to model duty cycle lengths with beta distributions [10]. Besides works on measuring spectrum occupancy and analyzing distributions of time statistics, the underlying stochastic processes of spectrum tenancy have also been investigated. Channel occupancy of multiple PUs is characterized as a discrete time Markov chain to assess link qualities in heterogeneous networks [11]. An autoregressive (AR) model is adopted to describe and predict spectrum tenancy in multiple bands and locations based on three days' measurement results [12].

The aforementioned spectrum tenancy models enhance our understanding of PU radio activities, but they do not consider how spectrum occupancy models should change with time-varying spectrum usage patterns. When modeling a time series, such as spectrum occupancy data, we need to consider the time-varying nature of the data-generating process, which is especially true when measurements last for a long time [13]. Initial efforts in this regard have been shown in [14], where the occupancy model of a WiFi spectrum band considers the changes of radio activity patterns, namely data transmissions, short interframe space waits, acknowledgement transmissions, and idle periods. Since WiFi contains only one data channel that can be occupied by a single device at a time, the channel usage states underlying different spectrum tenancy patterns can be enumerated and captured by a tractable Markov chain. In frequency bands with complex tenancy, however, it still remains unanswered that *what is the way to identify and model changing spectrum occupancy patterns*. Additionally, existing spectrum tenancy models mostly treat each radio channel independently, apart from some studies on occupancy correlations of channels in the same Radio Access Technology (RAT) at different locations [15]. Thus, it is worth studying *what spectrum tenancy model is able to characterize the correlations of multiple co-located channels*.

To showcase our solutions to these research questions, we target PU spectrum tenancy in LTE systems where versatile assignments of spectrum resources in multiple channels are conducted. For a PU system with complex scheduling of multiple channels whose state space is too large for Markovian models that described WiFi channel activities [14], we adopt a change detection method to identify data points where variations of occupancy patterns occur [16]. Since spectrum tenancy prediction is one of the most important applications of occupancy modeling, the adopted change detection method

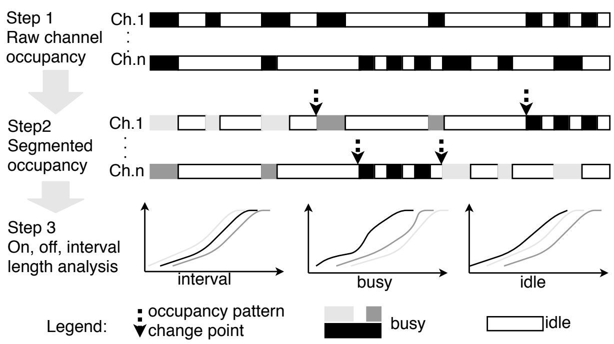

Fig. 1. An overview of the analysis process.

identifies variations of spectrum activity patterns based on parameter changes of prediction models. Having identified the data points where spectrum occupancy patterns change, we then segment the time series of spectrum tenancy into shorter lengths, each of which is governed by a stable occupancy pattern. The segment length distribution and its impact on spectrum tenancy analysis are investigated. After the segmentation, Vector Autoregression (VAR) is proposed to model the spectrum tenancy of multiple channels in an LTE cell. To demonstrate the superiority of VAR over single-channel models, we compare the fitting performance of VAR with that of the widely adopted on/off model. The comparisons consider how well the models characterize time lengths of channel occupancy, such as distributions of idle and busy periods. The overall work-flow of this paper is also illustrated in Fig. 1, and the major findings are listed below.

- The fitting accuracy of both on/off and VAR models are sensitive to the choice of the segment size of spectrum occupancy. Both the change detection algorithm and the brute force method suggest the segment size of around 100 millisecond (ms) for LTE spectrum tenancy.
- Judging from three aspects of fitting accuracy, correlations among adjacent channels, and the correlation between adjacent idle and busy period lengths, VAR outperforms on/off model in LTE spectrum tenancy analysis.

II. SEGMENTATION OF SPECTRUM OCCUPANCY

As the spectrum tenancy is decided by highly volatile radio activities in LTE systems, we first segment time series of spectrum occupancy to avoid fitting one model to tenancy data with time-varying channel usage patterns. Martingale Test (MT) algorithm is adopted, because it is a one-pass method with validated change detection accuracy [16].

A. The adopted segmentation method

For self-containment purpose, MT algorithm is briefly explained. To detect changes in time series using MT, first the features $\mathbf{x}_i$ to predict spectrum tenancy y_i should be identified, so a data set $T = \{(\mathbf{x_1}, y_1), \cdots, (\mathbf{x_n}, y_n)\}$ is constructed by pairing the prediction inputs and the occupancy labels y_i. Then, a strangeness measure s_i is computed for the ith prediction input. The strangeness measure is

$$s_i = \frac{\sum_{j=1}^{k} d_{ij}^{y}}{\sum_{j=1}^{k} d_{ij}^{\neg y}}, \tag{1}$$

where d_{ij}^{y} is the jth shortest distance of input i to other inputs with the same occupancy label, and $d_{ij}^{\neg y}$ is the jth shortest distance of $\mathbf{x_i}$ to others with a different tenancy. The strangeness measure is then applied to compute two other statistics, $\hat{p}_i$ and $M_n^{(\epsilon)}$, where $\hat{p}_i$ is

$$\begin{aligned}&\hat{p}_i(\{(\mathbf{x_1}, y_1), \ldots, (\mathbf{x}_i, y_i)\}, \theta_i) \\ &\quad = \frac{\#\{j : s_j > s_i\} + \theta_i \#\{j : s_j = s_i\}}{i},\end{aligned} \tag{2}$$

and $M_n^{(\epsilon)}$ is

$$M_n^{(\epsilon)} = \prod_{i=1}^{n} \left(\epsilon \hat{p}_i^{\epsilon-1}\right). \tag{3}$$

The tune parameter ϵ is in range $[0, 1]$, and θ_i is a uniform random variable in $[0, 1]$. The number of elements in a set is denoted by $\#\{\cdot\}$. Finally, decisions on occupancy pattern changes are made based on $M_n^{(\epsilon)}$. The null hypothesis H_0 that no occupancy pattern change happens at time instance n is accepted if $M_n^{(\epsilon)} < \lambda$, where λ is chosen based on an acceptable false alarm rate $1/\lambda$.

B. Chosen features for spectrum prediction

To drive the aforementioned segmentation method, we propose to predict LTE spectrum tenancy using k-nearest neighbor (k-NN) algorithm, and choose the features to compute the distances for strangeness in (1). Consider a training data set T and an input $\mathbf{x_0}$ for which the corresponding output y_0 needs to be predicted. K-NN prediction algorithm is essentially a majority vote, meaning that the forecast $\hat{y_0}$ is the arithmetic mean of the occupancy of k nearest neighbors in the training set.

$$\hat{y_0} = \frac{1}{k} \sum_{\mathbf{x_i} \in N_k(\mathbf{x_0})} y_i, \tag{4}$$

where $N_k(\mathbf{x_0}) = \{\mathbf{x_j} | \sum_{i \in \{1,\cdots,n\}} \mathbb{1}_{\{||\mathbf{x_j}-\mathbf{x_0}|| \geq ||\mathbf{x_i}-\mathbf{x_0}||\}} \leq k\}$. The indicator function is written as $\mathbb{1}_{\{\cdot\}}$, and $||\cdot||$ stands for the Euclidean distance.

According to existing works on spectrum prediction, the inputs have been unanimously chosen as history tenancy [17]. For example, the inputs to predict channel occupancy are considered as the history tenancy in last 15 time steps in [18], and occupancy in 4 previous time slots in [19]. We choose the features for the k-NN algorithm through trial and error to obtain the inputs that achieve the most accurate predictions. Since there are multiple channels in an LTE system, history occupancy in adjacent channels are also considered.

The spectrum tenancy data to train the prediction algorithm is a subset of the downlink spectrum occupancy measurement of a commercial LTE cell in 24 hours. The cell works in Frequency Division Duplex (FDD) mode with 10 MHz bandwidth, so there are 50 data channels and each of them is scheduled every millisecond [20]. Correspondingly, the entire data set comprises the binary tenancy of the 50 channels in all the 8.64×10^7 ms of a day, where idle and busy slots are represented by zero and one, respectively.

To choose the features for the strangeness computation in (1), we compare the prediction accuracies of 30 groups of inputs, considering three choices of history lengths, 4, 8, and 12 ms, and two options for channels, single channel or 3 adjacent channels, and five choices for the number of neighbors, ranging from 1 to 5. The training and validation data sets are the occupancy of the first 10^4 ms in each hour of 3 channels in the center, i.e., channel 24, 25, and 26. The first 8000 ms of the tenancy data in each hour is used for model training, while the rest 2000 ms is reserved for validation.

For most of the hours, the prediction accuracies of the 30 groups of features are very similar, and all of them are over 90%. Hour 9 is the only hour when relatively inaccurate predictions and large accuracy differences are observed for different groups of inputs, so the features are selected based on the accuracy in this hour. The prediction accuracy in hour 9 is listed in Table I. The first row shows the number of channels and the history length, while the number of neighbors are presented in the first column. The other cells contain the prediction accuracy of k-NN algorithms with the corresponding features and the number of neighbors. Thus, the chosen features are the occupancy of the 3 closest channels in the past 4 ms, and the number of nearest neighbors who vote the prediction result is 4, because the accuracy achieved by this combination of parameters $c = 3, h = 4, k = 4$ is the highest in hour 9, which is highlighted in the table.

Then, the changing points of spectrum occupancy of channel 25 are detected based on M values computed from (3), where $\lambda = 100$, and $\epsilon = 0.9$ as recommended in [16]. The distribution of segment lengths is illustrated in Fig. 2. Because long segments are scarce, the horizontal axis is plotted in log scale. As shown in the figure, most of the segment lengths fall in the range of 80 to 120 ms, and the mean time of a segment discovered by MT change detection algorithm is 105 ms. The impact of segment length on analysis accuracy will be further explained in section IV.

TABLE I
PREDICTION ACCURACIES OF DIFFERENT FEATURES IN HOUR 9

c,h	1,4	1,8	1,12	3,4	3,8	3,12
k=1	0.5040	0.6255	0.5020	0.5235	0.6250	0.5130
k=2	0.5040	0.6255	0.6045	0.5220	0.6350	0.6120
k=3	0.5040	0.6255	0.6045	0.5220	0.6450	0.6120
k=4	0.6495	0.6255	0.6045	0.6565	0.6355	0.6120
k=5	0.5040	0.6255	0.6045	0.5220	0.6345	0.6120

III. SPECTRUM OCCUPANCY MODELS

In this section, the ways to fit on/off and VAR models to measurement results are explained. Since there are multiple channels in LTE, we propose to model the multi-channel spectrum occupancy with VAR. As reviewed earlier, the state-of-the-art spectrum occupancy models are mostly targeting a single channel [21], so we choose the widely used on/off model as a representative for performance comparison. The way to compare on/off and VAR models is described, and three metrics are proposed for performance comparisons.

A. *Single-channel spectrum occupancy model*

Among the single-channel occupancy models surveyed in [21], we choose the on/off model due to its wide usage. Our occupancy data comprises 0 and 1 to indicate whether an LTE channel is idle or occupied. Assume the time lengths of idle or busy periods of one channel to be independent and identically distributed (i.i.d.). Define the vectors $(Y_n, Z_n), n \in \mathbb{N}^+$ where Y_n and Z_n are samples of i.i.d. random variables representing time lengths of idle or busy periods. We fit five widely used distributions, exponential, Weibull, Lognormal, Generalized Pareto, and Gamma distributions to the observed samples, and the parameters are estimated using Maximum Likelihood Estimation (MLE). The goodness-of-fit is obtained by conducting Kolmogorov Smirnov (K-S) test, also used in [14] for the same purpose. K-S test is a tool for comparing the closeness of two distributions. We employ K-S test to compare the empirical distributions of the time statistics of measurement data with the distributions obtained by the on/off model fitting. The empirical distribution $F_e(x)$ of a random variable X that has n observed samples X_i is

$$F_e(x) = \mathbb{P}(X < x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i < x\}}. \tag{5}$$

The upper bound of the difference between the empirical distribution and the fitting model distribution is D,

$$D = sup|F_e(x) - F_0(x)|. \tag{6}$$

If empirical distribution F_e and fitting model distribution F_0 are identical, the distribution of the random variable D in this case, denoted as D^*, is independent of the fitting distribution. Let G be the cumulative distribution function of D^*. The p value is defined as $p = 1 - G(D)$, so the larger the p value, the more likely D obeys the distribution of D^*, meaning that $F_e(x)$ and $F_0(x)$ are more likely to be the same. A threshold value $p = 0.05$ is chosen, so the null hypothesis that the samples follow the distribution F_0 is accepted when $p \geq 0.05$.

B. *Multi-channel spectrum occupancy model*

VAR extends the AR model for single-channel occupancy to multiple channels [12]. VAR model treats spectrum occupancy of multiple channels at each time slot as a sample of a multivariate normal random variable that results from the sum of a constant, white noise, and multivariate normal random variables representing the data in previous time slots. To model the spectrum usage of multiple channels, we fit VAR models with different time lags to spectrum usage measurements, and the parameters are estimated using MLE. The spectrum occupancy at time slot n are random vectors, denoted as $\mathbf{V}_n$,

$$\mathbf{V}_n = \mathbf{c} + \sum_{i=1}^{k} \boldsymbol{\phi}_i \mathbf{V}_{n-i} + \boldsymbol{\varepsilon}_n. \tag{7}$$

The constant vector is $\mathbf{c}$, and $\boldsymbol{\varepsilon}_n$ is the noise term. $\mathbf{V}_{n-i}$, where $1 \leq i \leq k$, is the channel usage in a previous time slot

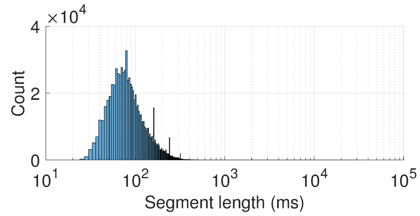

Fig. 2. The distribution of segment lengths in channel 25.

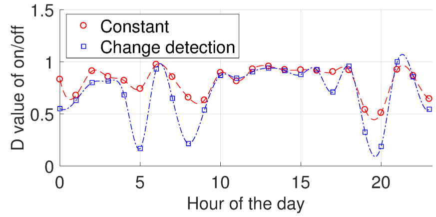

Fig. 3. Off time D value comparisons of on/off models fitted to different segment sizes.

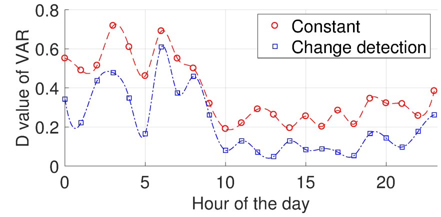

Fig. 4. Off time D value comparisons of VAR models fitted to different segment sizes.

no earlier than the time lag k, and its linear relations with $\mathbf{V}_n$ are described by the matrix ϕ_i.

In VAR model, the time lag k is a design parameter. In order to compare the performance of VAR models with different time lags, we adopt the Akaike Information Criterion (AIC), defined as

$$AIC = 2n_p - 2\log(L), \tag{8}$$

where L is the optimized scalar value of log-likelihood objective function, and n_p is the number of parameters that are estimated for the model. AIC measures the relative qualities of statistical models fitted to a data set, and models with small AIC values are preferred because they better capture statistical features of the data and have fewer parameters to estimate.

C. Comparisons between on/off and VAR models

We compare on/off model with VAR from three aspects. They are the D value of K-S test, the correlation coefficients of spectrum occupancy among adjacent channels, and the correlation coefficient between adjacent off time and on time. Specifically, we extract the spectrum occupancy in 10^4 time slots of 10 LTE channels from each hour of the day, and then fit the single-channel and the multi-channel models to the data. For the VAR model, fitting multiple channels requires only adjusting the number of elements in the vectors $\mathbf{V}_n$, $\mathbf{c}$, ε, and the matrices ϕ in (7). For the on/off model, we fit 10 on/off models to each of the 10 channels independently, and obtain 10 sets of parameters. Using two types of models, we produce synthetic spectrum occupancy of the same size with the measurement data. To compare how close the synthetic data resembles the measurements, the D values are calculated between the measurement data and the two sets of synthetic data using (5) and (6), where $F_0(x)$ becomes the empirical distribution of synthetic data. As indicated in previous studies, spectrum occupancy of the same radio access technology are correlated even in different locations [15], we study how the different channels in the same cell are correlated and whether the correlations are captured by the models. The correlation coefficient is Pearson correlation coefficient. Moreover, since the correlations between adjacent off time and on time are suggested in previous studies [8], we investigate this correlation reflected by measurement and synthetic data for LTE channels. The results of the comparisons in the three aspects are presented in Subsection IV-C.

IV. Fitting Models to Measurement Results

We fit on/off and VAR models to the measurement data, and provide the results in three perspectives. First, the impact of segment lengths on fitting accuracy is investigated, and we show that the change detection method is able to find the segment length that coincides with the one obtained by brute force. Then, fitting results of on/off model are presented, and we observe the preferred distribution for three occupancy time metrics. Guidelines for fitting on/off models to spectrum tenancy data are proposed. Lastly, we identify the time lag parameter for the VAR model, and compare the performance of VAR and on/off models in three aspects introduced earlier, i.e. the D value of K-S test, correlation coefficients of spectrum occupancy among adjacent channels, and the correlation coefficient between adjacent off time and on time. Based on the comparisons, we conclude that VAR outperforms on/off model in characterizing LTE spectrum occupancy.

A. The impact of segment length

To demonstrate the impact of segment lengths on goodness-of-fit, we fit on/off and VAR models to spectrum occupancy of channel 25. The spectrum tenancy data in each hour is segmented into groups for analysis in two different ways. In one way, the segments are obtained by cutting the time series of spectrum occupancy before data points where pattern changes are discovered by MT change detection algorithm. In the other way, the segments are obtained by cutting the binary channel occupancy series into groups of the constant size of 500 ms. The two models are fitted to the segmented occupancy, and the D values between the distributions of measured idle periods and those of synthetic data are presented in Fig. 3 to 4. As shown by the lower blue lines, both models have smaller D values, i.e. better goodness-of-fit, when fitted to tenancy segments discovered by MT change detection.

The segment size obtained by change detection is further validated by brute force method. To rediscover the segment length through trial and error, we fit on/off models to time metrics of channel occupancy in various segment sizes. We fit five different distributions, exponential, Weibull, Lognormal, Generalized Pareto, and Gamma, to the time length of idle and busy periods, and intervals which are the sum of idle and the following busy periods. The fitting rate is the number of fittings accepted by K-S tests over the total number of fittings attempted. The fitting rates for segment lengths of 10, 30, 50, 100, 200, 500, 1000, and 2000 ms are illustrated in Fig. 5. For each segment length value, the tenancy of channel 25 in 3.6×10^6 ms of each hour are first partitioned into segments of the designated length, and the three occupancy time metrics in every segment are fitted by the five distributions. According to Fig. 5, all the five distributions can fit the time lengths well

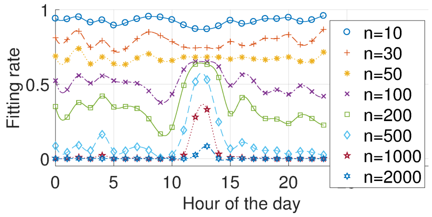

Fig. 5. Fitting rates of on/off models fitted to different segment lengths.

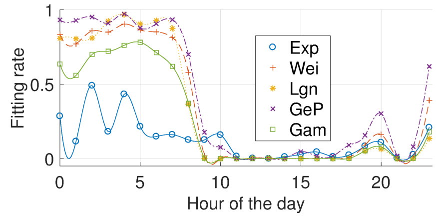

Fig. 6. Off time fitting rates achieved by on/off models with different distributions.

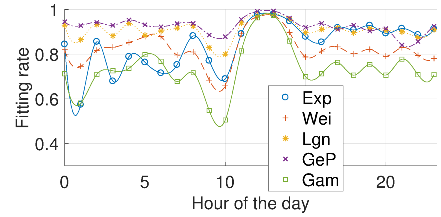

Fig. 7. On time fitting rates achieved by on/off models with different distributions.

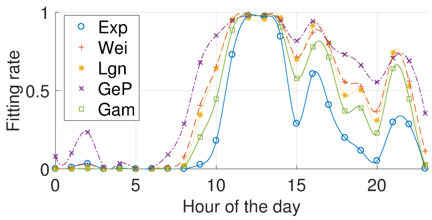

Fig. 8. Interval length fitting rates achieved by on/off models with different distributions.

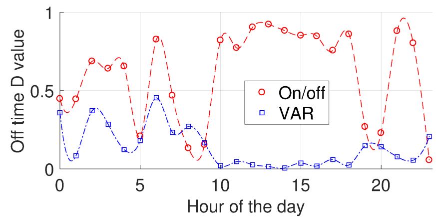

Fig. 9. D values between off time distributions of synthetic data and measurements.

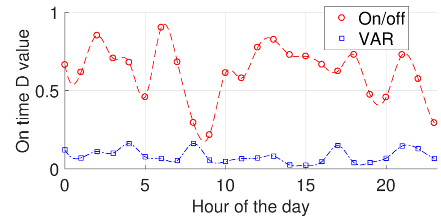

Fig. 10. D values between on time distributions of synthetic data and measurements.

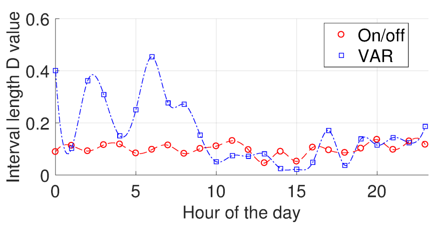

Fig. 11. D values between interval length distributions of synthetic data and measurements.

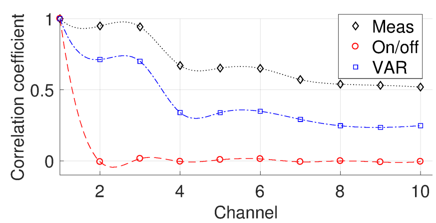

Fig. 12. Spectrum tenancy correlations among adjacent channels.

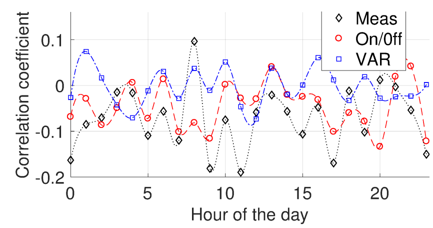

Fig. 13. Correlations between on time and the following off time.

when segments are short, e.g. 10 or 30 ms. In such cases, the distributions do not reflect the statistical features of the time lengths due to small numbers of samples. When the segments are long, such as 1000 or 2000 ms, fitting the time lengths using on/off model is impossible due to low fitting rates close to zero. Since five distributions are employed, the reasonable choice is the segment length of 100 ms whose fitting rate is around 0.5, meaning that the idle, busy, and interval time lengths can be fitted by two or three of the five distributions on average. According to the distribution shown in Fig. 2, the mean size of segments obtained by MT algorithm is 105 ms, which is close to the 100 ms segment length achieved by brute force. As the segment size of 100 ms is suggested by both the change detection algorithm and brute force trials, the segment length considered in later analysis is 100 ms.

B. Fit on/off model to measurement

With the chosen segment size of 100 ms, the single-channel fitting rates of the five distributions are compared in Fig. 6 to 8 in terms of the three metrics, idle, busy, and interval lengths, respectively. We observe from Fig. 6 that generalized Pareto distribution has the highest fitting rates for off time in most hours. The off time cannot be well fitted by any of the five distributions from the 9th hour to the last. The on time can be best fitted by generalized Pareto distribution with fitting rates over 90% all day, as shown in Fig. 7. Fig. 8 illustrates that the interval length suffers from poor fitting rates during hours 23 to 8. Generalized Pareto also achieves the most top fitting hours for interval lengths. Thus, we conclude that analyzing LTE downlink spectrum occupancy with on/off models should use different strategies based on the hours. In daytime, on time and interval lengths are fitted accurately by the generalized Pareto distribution. In the night, on time and off time should be fitted with generalized Pareto distribution.

C. Comparisons between on/off and VAR models

To obtain the time lag parameter k, we fit VAR models with different k values, 1, 2, 3, 4, 8, and 12, to the measurement data, and the AIC values are calculated for the fitted models. Because the differences among AIC values for the six models with various time lags are negligible, the time lag is chosen to be 1, which has the fewest parameters to estimate. Other parameters of VAR are obtained by fitting the model to measurement results in the way described in Subsection III-B.

After achieving the segment length, the on/off model fitting strategy, and the time lag of VAR, we set out to compare on/off and VAR models from the aspects of D values of K-S tests, correlations among adjacent channels, and correlations between adjacent off time and on time. Specifically, the D values are the statistics of K-S test, and they are calculated according to (5) and (6). To compare D values, on/off and VAR models are fitted to occupancy data of channel 25, and synthetic data is generated by the models. The D values between the distributions of measurement and synthetic spectrum occupancy are analyzed and presented in Fig. 9 to 11. Fig. 9 compares the differences between idle time distributions of synthetic data and that of the measurement data. In most of the hours, VAR model produces synthetic data whose off time distributions achieve smaller D values. In terms of the similarity comparison of on time distributions, VAR model has smaller D values in all hours, meaning that the busy periods of synthetic occupancy generated by VAR models resemble those in measurements, as revealed by Fig. 10. Fig. 11 exhibits that the interval length distributions of the synthetic data produced by on/off and VAR models have similar D values. Thus, the synthetic data produced by VAR model has more similar occupancy time distributions to those of measurement data than the data synthesized by on/off model.

Obtaining the multi-channel fitting results as described in Subsection III-C, we investigate the occupancy correlations among adjacent LTE channels. Fig. 12 presents the pairwise correlation coefficients between the spectrum tenancy of the first channel and all the ten channels. The black line shows the channel tenancy correlations of measurements versus frequency-wise distance. The measured spectrum tenancy in the first channel shows very high correlations with those of the three nearest channels, and the correlations decrease as the frequency distance grows. This trend is captured very well by the data generated by the VAR model, though the correlations in blue are lower than those of measurements. Since on/off model is a single-channel model, its synthetic tenancy has zero correlations among adjacent channels, plotted in red.

Though some previous studies suggest that off times and the following on times are negatively correlated [8], LTE spectrum tenancy does not show this phenomenon as demonstrated by Fig. 13. In this figure, the correlations between idle periods and the following busy periods are studied for measurements and synthetic data. The correlations among adjacent idle and busy periods in the three groups of data are very similar and bounded within $[-0.2, 0.1]$, meaning that off time and the following on time are not correlated. The phenomenon is due to the fact that LTE systems schedule spectrum resources every 1 ms, so the off time and the next on time many scheduling intervals away are unrelated.

In conclusion, VAR achieves better performance in LTE spectrum tenancy modeling, due to its superior capabilities to capture the statistical features of busy and idle time lengths, and occupancy correlations among adjacent channels.

V. Conclusion

To model LTE spectrum occupancy accurately, data series of channel tenancy are segmented by a change detection algorithm to avoid fitting models to occupancy data produced by radio activities of time-varying patterns. We observe the impact of data series lengths on fitting accuracy, and verify that the segment size discovered by the change detection coincides with the one obtained by brute force. LTE spectrum tenancy measurements are characterized by the existing on/off model and the proposed VAR model. The guidelines for fitting LTE spectrum tenancy with on/off models are introduced. The performance of on/off and VAR models is compared from three perspectives, and VAR captures the characteristics of LTE spectrum occupancy more precisely overall.

References

[1] L. An, R. Zou, Z. Liu, Z. Hu, and Q. Niu, "An analytical model for tdma-based mac protocols in vanets," in *2014 International Symposium on Wireless Personal Multimedia Communications (WPMC)*. IEEE, 2014, pp. 607–612.

[2] T. Zhang, J. Zhao, L. An, and D. Liu, "Energy efficiency of base station deployment in ultra dense hetnets: A stochastic geometry analysis," *IEEE Wireless Communications Letters*, vol. 5, no. 2, pp. 184–187, 2016.

[3] M. Wei, J. Wang, Z. Lu, and W. Wang, "On studying information dissemination in social-physical interdependent networks," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.

[4] R. Zou and W. Wang, "A flow rule timeout assignment algorithm for sdn-assisted network mimo systems," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.

[5] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *Proceedings of the first international workshop on Technology and policy for accessing spectrum*. ACM, 2006, p. 1.

[6] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for distributed dynamic spectrum access," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 310–323, 2019.

[7] T. Harrold, R. Cepeda, and M. Beach, "Long-term measurements of spectrum occupancy characteristics," in *2011 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2011, pp. 83–89.

[8] T. A. Hall, A. Sahoo, C. Hagwood, and S. Streett, "Exploiting lte white space using dynamic spectrum access algorithms based on survival analysis," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–7.

[9] B. Khalfi, B. Hamdaoui, M. Guizani, and N. Zorba, "Efficient spectrum availability information recovery for wideband dsa networks: A weighted compressive sampling approach," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2162–2172, 2018.

[10] M. Wellens and P. Mähönen, "Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model," *Mobile networks and applications*, vol. 15, no. 3, pp. 461–474, 2010.

[11] B. Li, W. Guo, H. Zhang, C. Zhao, S. Li, and A. Nallanathan, "Spectrum detection and link quality assessment for heterogeneous shared access networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1431–1445, 2019.

[12] A. Gorcin, H. Celebi, K. A. Qaraqe, and H. Arslan, "An autoregressive approach for spectrum occupancy modeling and prediction based on synchronous measurements," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2011, pp. 705–709.

[13] G. Lu, Y. Zhou, C. Lu, and X. Li, "A novel framework of change-point detection for machine monitoring," *Mechanical Systems and Signal Processing*, vol. 83, pp. 533–548, 2017.

[14] S. Geirhofer, L. Tong, and B. Sadler, "A measurement-based model for dynamic spectrum access in wlan channels," 2006.

[15] S. Yin, D. Chen, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: a large-scale spectrum measurement study," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 1033–1046, 2012.

[16] S.-S. Ho and H. Wechsler, "A martingale framework for detecting changes in data streams by testing exchangeability," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2113–2127, 2010.

[17] G. Ding, Y. Jiao, J. Wang, Y. Zou, Q. Wu, Y.-D. Yao, and L. Hanzo, "Spectrum inference in cognitive radio networks: Algorithms and applications," *IEEE Communications Surveys & Tutorials*, 2018.

[18] V. K. Tumuluru, P. Wang, and D. Niyato, "A neural network based spectrum prediction scheme for cognitive radio," in *2010 IEEE International Conference on Communications*. IEEE, 2010, pp. 1–5.

[19] L. Yin, S. Yin, W. Hong, and S. Li, "Spectrum behavior learning in cognitive radio based on artificial neural network," in *2011-MILCOM 2011 Military Communications Conference*. IEEE, 2011, pp. 25–30.

[20] E. U. T. R. Access, "Physical channels and modulation, 3gpp ts 36.211," *V10*, vol. 2, 2009.

[21] Y. Saleem and M. H. Rehmani, "Primary radio user activity models for cognitive radio networks: A survey," *Journal of Network and Computer Applications*, vol. 43, pp. 1–16, 2014.