

A Framework for Statistical Characterization of Indoor Data Traffic for Efficient Dynamic Spectrum Access in the 2.4 GHz ISM Band

Muhammad Khurram Ehsan and Dirk Dahlhaus

Communications Laboratory, University of Kassel, Germany

Email: khurram.ehsan@student.uni-kassel.de, dahlhaus@uni-kassel.de

ABSTRACT

The key for efficient dynamic spectrum access (DSA) is to model the spectral resources accurately. A large number of measurement campaigns have been performed to estimate the spectrum usage in outdoor and indoor scenarios. This spectrum usage estimation helps policy makers to optimize the spectrum management methodologies. The spectrum usage studies also assist researchers to constitute a way for efficient DSA using prior knowledge of the distribution of the observed data traffic in cognitive radio (CR) systems. In this paper we extend our previous work which statistically modeled the observed data traffic in the industrial, scientific and medical (ISM) band at 2.4-GHz in two neighboring frequency subbands and time slots, respectively, to *three* neighboring frequency subbands and time slots, respectively. As before, the frequency and time correlation functions of the observed data traffic are modeled by an exponentially decaying function. The multivariate Gaussian mixture (MGM) is validated as a good candidate to model the joint distribution of measured data and also to estimate the correlation between the measured data in neighboring frequency subbands and as well as in time domain samples.

KEYWORDS

Wi-Spy Dual Band Spectrum Analyzer (WSDSA), Dynamic Spectrum Access (DSA), multivariate Gaussian mixture (MGM), k-means clustering, ISM data.

1 INTRODUCTION

It is a challenge for the industry as well as academia to optimize spectral resources in wireless communications. The numbers of customers are growing rapidly in the

telecommunication industry due to innovative wireless technologies and high-speed data services. On the contrary, extensive measurement campaigns in different parts of the world [1] have claimed the under-utilization of the already allocated spectrum. The under-utilization has arisen due to the static allocation policies e.g. of the Federal Communications Commission (FCC). The advantages of static spectrum allocation policies are interference reduction and simple system hardware. On the other hand, the major disadvantage of these static policies is the inefficient utilization of radio spectral resources.

To address the issues between spectrum under-utilization and demand, cognitive radio (CR) technology stands as a promising candidate. CR provides the mechanism in which the available spatial and spectral resources are utilized by intelligently adapting parameters in the radio environment [2]. The switching between different operating frequencies, time duration to access specific frequency bands, variable operating frequencies and different spectrum ranges are representatives of intelligent adaptations and are collectively termed as dynamic spectrum access (DSA).

CR systems operate in three different ways, namely in so-called overlay, underlay and interweave modes. In an interweave system, secondary users (SUs) can access the spectrum in an opportunistic way only in the absence of a primary user (PU) and will stop transmission as soon as the PU resumes transmission. As a result, an efficient DSA with non-interfering SU signals arises from a correspondingly accurate statistical model for the existing data traffic observed in the radio environment. At the same time, the model should be simple enough to limit the complexity of resulting DSA approaches.

A lot of field measurement campaigns have been conducted in licensed (global system for mobile communication (GSM), universal mobile telecommunication system (UMTS)) and unlicensed (industrial, scientific and medical (ISM)) bands in both outdoor and indoor scenarios. The time-dependent spectrum occupancy using a four-state Markov model is discussed in [3]. The time-variant power spectrum at a CR receiver is measured in real-time and presented graphically in [4] and [5]. Interference temperature model based on real time measurements is discussed in [6]. Spectrum occupancy measurements for ultra high frequency (UHF) television (TV) bands are presented in [7]. The duty cycle models based on real time measurements are described for CR systems in [8] and [9]. In [10], real time measurement campaigns are conducted in different spectrum bands and also the average signal channel power in TV bands is modeled as a Gaussian random variable. The quantitative study of spectrum occupancy based on field measurements is presented in [11]. In [12], the average spectrum occupancies for UHF and very high frequency (VHF) bands are investigated. In [13], the authors assume that the signal power distribution of the PU in each subband is Gaussian. In [14], the autocorrelation function (ACF) of a wide-sense stationary (WSS) or short-range dependent process is modeled as a decaying exponential. The drawbacks in [3]-[11] are:

1. Simple occupancy matrices are used to estimate the data traffic in measured bands.
2. The data traffic is modeled using Markov chains following a Gaussian distribution under the assumption that the data traffic is independent and identically distributed (i.i.d) as well as unimodal, which clearly contradicts the properties of the data traffic observed in real-world scenarios.

In this paper, to model the data traffic measured in realistic scenarios more accurately, we have to deal with the correlation of the observed data traffic and the multimodality of the traffic data distribution. Here, we extend our previous work [15], which statistically modeled the observed data traffic in the ISM band at 2.4 GHz in two neighboring frequency subbands and time slots, respectively, to three neighboring frequency

subbands and time slots, respectively. The observed data traffic in an indoor environment in the aforementioned ISM band is again modeled as a stochastic field in time and frequency. The process in time and frequency is assumed to be WSS with an autocorrelation function approximated by a decaying exponential. The multivariate Gaussian mixture (MGM) is considered as a suitable distribution model for the observed data traffic in this band. While there is no prioritization for the data traffic in the ISM band, we consider WLAN data traffic as a PU for simplicity.

The organization of the paper is as follows. The measurement scenarios and set-up are discussed in section 2. In section 3, the statistical characterization of ISM data traffic is detailed. The parametric modeling of the data traffic is described in section 4. In section 5, the MGM model is presented and validated using correlation functions. Conclusions are drawn in section 6.

2 MEASUREMENT SCENARIOS AND SET-UP

2.1 Measurement Scenarios

Indoor measurements are conducted at the second floor of the engineering campus of the university of Kassel, Germany. The engineering campus is located in a residential area. We choose two different locations for conducting the measurement campaigns. At first, the measurements are taken in a computer laboratory. In order to analyze different user traffic profiles, measurements are also taken in an office room. At each location, measurements are taken for 8 hours per day for the duration of two weeks. In Figures 1 and 2, the snapshots of the measurements for an 8-hour day can be seen in an office room and a laboratory, respectively.

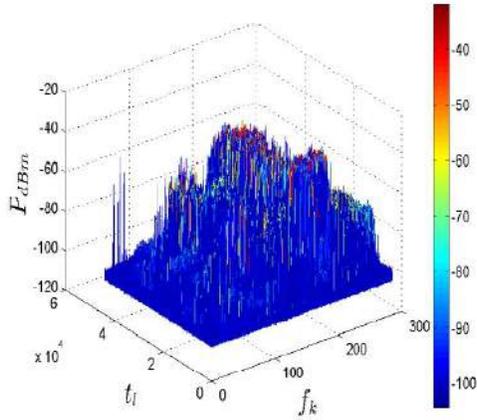


Figure 1. ISM data traffic measured in an office room.

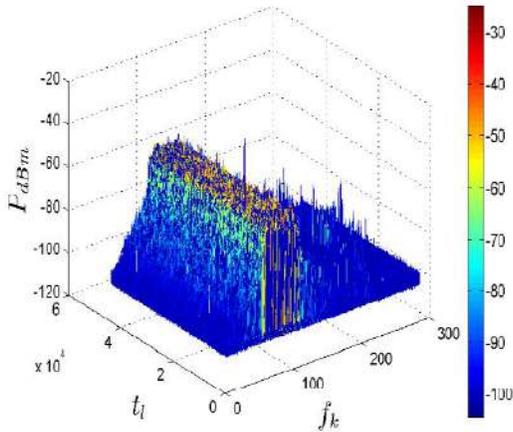


Figure 2. ISM data traffic measured in a computer lab.

Here, the traffic is characterized by the instantaneous power P in dBm (P_{dBm}).

2.2 The Testbed

A Wi-Spy Dual Band Spectrum Analyzer (WSDSA) with an omnidirectional antenna, as shown in Figure 3, is used to measure the ISM band activity.

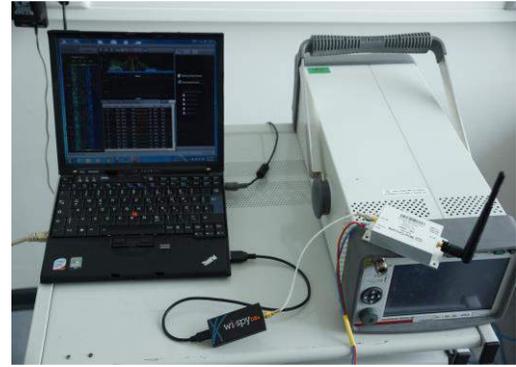


Figure 3. Testbed including WSDSA and omnidirectional antenna.

The term *activity* refers to the time-variant power spectral density (PSD) of the measured data, and it is this PSD as mentioned above that is considered to be traffic in the following. In this study, we only have the measurements at the 2.4 GHz band. However, the WSDSA can be used in both 2.4 GHz and 5 GHz bands to analyze the ISM data traffic.

The frequency span of the WSDSA is 83 MHz and it has a sweep time of 560 ms. A low noise amplifier (LNA) ZLR-3500+ is also introduced in the set-up, not only to have a better detection of weak signals, but also to improve the sensitivity of the measurement system. The frequency range to be scanned by the WSDSA is 2.4...2.483 GHz in our measurements. More of the measurement specifications are detailed in Table 1.

Table 1. Measurement Specifications

| Parameters | Values |
|----------------------|-----------------------|
| Frequency Range | 2.4 GHz ... 2.483 GHz |
| Frequency Span | 83 MHz |
| Frequency Resolution | 333 KHz |
| Measurement Duration | 8 Hours |
| Sweep Time | 560 ms |
| Preamplifier | 21 dB |

The testbed is interfaced with the Chanalyzer Pro® software that helps in the detailed visualization of the ISM data traffic observed by the WSDSA. The spectral view of the observed data given by Chanalyzer Pro® shows the spectrum usage over time which looks like a scrolling waterfall plot as shown in Figure 4.

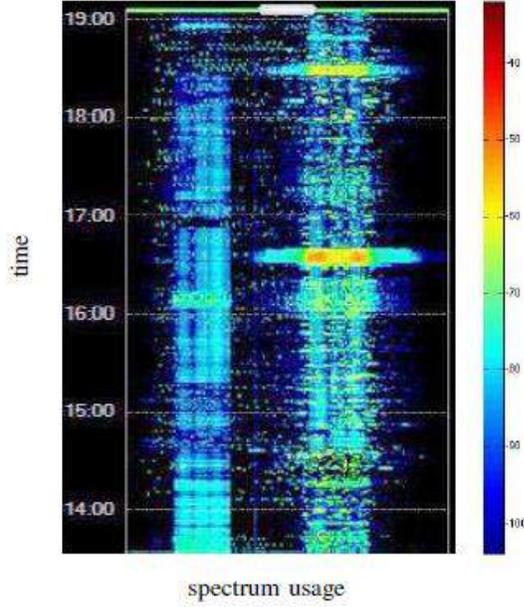


Figure 4. Waterfall plot showing spectrum usage over time.

2.3 Data Matrix

The real-time observed data traffic is saved in a comma-separated values (CSV) format by Chanalyzer Pro® to analyze and process it further. The observed data is then represented as a matrix where each row denotes the time instances having a resolution of $\Delta t = 560$ ms, while each column denotes the frequency subbands with a bandwidth of $\Delta f = 333$ kHz. The obtained data matrix normalized to unit average power is represented as

$$\mathbb{E} = \begin{pmatrix} x(t_1, f_1) & \dots & x(t_1, f_K) \\ \vdots & \ddots & \vdots \\ x(t_L, f_1) & \dots & x(t_L, f_K) \end{pmatrix}, \quad (1)$$

where $x(t_l, f_k)$ in the data matrix is recognized as an observation of the data traffic process random field $X(t_l, f_k)$ to denote the instantaneous power in dBm for each frequency subband $f_k = k\Delta f$, where $k = 1, 2, \dots, K$ and the time instants are defined by $t_l = l\Delta t$ with $l = 1, \dots, L$ and the upper limits $K = 285$ and $L = 1000$.

3 STATISTICAL CHARACTERIZATION OF ISM DATA TRAFFIC

It is assumed that the ISM data traffic $X(t_l, f_k)$ is a WSS process in both time and frequency. This is why the ACF of $X(t_l, f_k)$ is a function of the time and frequency differences, respectively, for a given value of the frequency and time, respectively.

3.1 Frequency Correlation Function

It is known that the frequency correlation function (FCF) for the ISM data traffic $X(t_l, f_k)$ which is dependent on Δk , is defined as

$$\phi_f(\Delta k) = E(X^*(t_l, f_k)X(t_l, f_{k+\Delta k})) \quad (2)$$

where the estimate of $\phi_f(\Delta k)$ is defined by the empirical FCF $\hat{\phi}_f(\Delta k)$ using the measured ISM data traffic according to

$$\hat{\phi}_f(\Delta k) = \frac{1}{L(K - \Delta k)} \sum_{k=1}^{K-\Delta k} \sum_{l=1}^L x^*(t_l, f_k)x(t_l, f_{k+\Delta k}), \quad (3)$$

where $\Delta k = 0, \dots, 284$. The FCF is modeled as a decaying exponential $\phi_f(\Delta k) = \rho_f^{\Delta k}$ due to the short-range dependency as described in [14]. Here, ρ_f represents the correlation between the frequency subbands with $\rho_f \in [0, 1]$. A standard numerical method is employed to find the estimate $\hat{\phi}_f(\Delta k)$. The value of $\rho_f^{\Delta k}$ is computed for $r_f = 0, 0.01, 0.02, \dots, 1$ and is chosen as the least squares estimate (LSE) between the modeled FCF $\phi_f(\Delta k)$ and the estimated FCF $\hat{\phi}_f(\Delta k)$. Both, the estimated FCF $\hat{\phi}_f(\Delta k)$ and the modeled FCF $\phi_f(\Delta k)$ are shown in Figure 5. The estimated value of $\hat{\rho}_{f, \text{LSE}}^{\Delta k}$ using the aforementioned approach is 0.56 when $\Delta k = 1$ and 0.31 for $\Delta k = 2$.

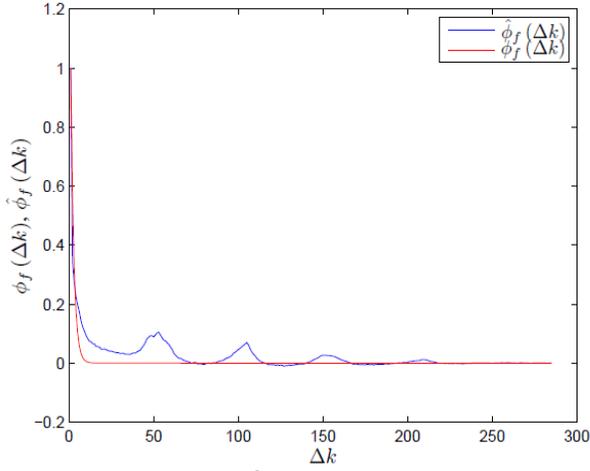


Figure 5. Empirical $\hat{\phi}_f(\Delta k)$ and modeled $\phi_f(\Delta k)$ Frequency Correlation Functions.

3.2 Time Correlation Function

The time correlation function (TCF) for the ISM data traffic $X(t_l, f_k)$, which is dependent on Δl , is expressed as

$$\phi_t(\Delta l) = E(X^*(t_l, f_k)X(t_{l+\Delta l}, f_k)) \quad (4)$$

where the estimation of $\phi_t(\Delta l)$ is obtained by the empirical TCF $\hat{\phi}_t(\Delta l)$ using the measured ISM data traffic as

$$\hat{\phi}_t(\Delta l) = \frac{1}{K(L-\Delta l)} \sum_{l=1}^{L-\Delta l} \sum_{k=1}^K x^*(t_l, f_k)x(t_{l+\Delta l}, f_k) \quad (5)$$

with $\Delta l = 0, \dots, 999$. As in the case of FCF, the TCF is also modeled as $\phi_t(\Delta l) = \rho_t^{\Delta l}$ as done in [14], where ρ_t denotes the correlation between the ISM data traffic observed at different time instances and $\rho_t \in [0, 1]$. The modeled TCF $\phi_t(\Delta l)$ and the empirical TCF $\hat{\phi}_t(\Delta l)$ according to the above model are shown in Figure 6.

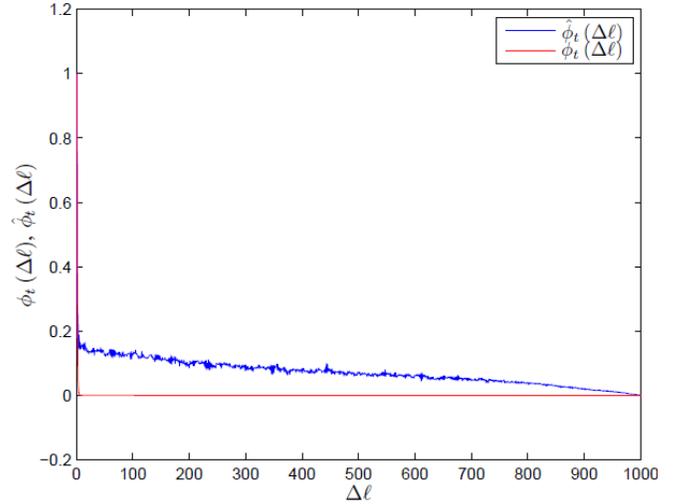


Figure 6. Empirical $\hat{\phi}_t(\Delta l)$ and modeled $\phi_t(\Delta l)$ Time Correlation Functions.

By using the LSE approach, the estimated value of $\hat{\rho}_{t, \text{LSE}}^{\Delta l}$ is 0.23 for $\Delta l = 1$ and 0.06 for $\Delta l = 2$. The aforementioned characterization of ISM data traffic $X(t_l, f_k)$ is used to validate its model in Section 5.

4 PARAMETRIC MODELING

The second-order moments of the process $X(t_l, f_k)$ are estimated as described in Section 3, but the joint PDF is also required for the complete description of the process $X(t_l, f_k)$. The joint PDF is estimated by adapting the parametric modeling approach, which is recognized as a powerful statistical tool for estimating the parameters of the presumed distribution.

In [10], the authors accept that their work to model the average signal channel power in television bands as a Gaussian is not always in agreement with the observed data. The reason for the disagreement between the observed and modeled data is the missing consideration of the multimodality of the signal power distribution. The Gaussian mixture (GM) is a suitable option to address the multimodality of the distribution of the measured signal as described in [16].

In this work, we separate the estimation of the joint PDF of the process $X(t_l, f_k)$ into the estimation of the PDF of the neighboring process

samples in frequency and time. After the PDFs of the aforementioned vectors are estimated, the joint PDF of $X(t_l, f_k)$ can be obtained by considering the independence of the components within the GM. Consider $\mathbf{X} = [X(t_{l_1}, f_{k_1}), X(t_{l_2}, f_{k_2})]^T$. At first, we consider the samples being neighbored in frequency (NIF), where $k_1 = k = k_2 - 1$ and $l_1 = l_2 = \lambda l_0$. The obtained vector is $\mathbf{X}_\lambda = \mathbf{X}_{\text{NIF}, \lambda}$ with $l_0 \geq 1$, where k is chosen arbitrarily within the set $k \in \{1, \dots, K-1\}$. The vector of samples neighbored in time (NIT) is denoted by $\mathbf{X}_\kappa = \mathbf{X}_{\text{NIT}, \kappa}$ where $k_1 = k_2 = \kappa k_0$ and $l_1 = l = l_2 - 1$. Similar to the case of NIF, the choice of l is given by $l \in \{1, \dots, L-1\}$ and $k_0 \geq 1$.

It is assumed that the distribution of the data vector \mathbf{X} is a multivariate Gaussian mixture (MGM) in both NIF and NIT scenarios. The sum of N Gaussian densities can be expressed as in [17] and [18] as

$$p(\mathbf{X} | \Theta) = \sum_{n=1}^N \pi_n f(\mathbf{X}_n | \mu_n, \Sigma_n), \quad (6)$$

where $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ and each θ_n represents the parameter set $\theta_n = \{\pi_n, \mu_n, \Sigma_n\}$ and $f(\mathbf{X}_n | \mu_n, \Sigma_n)$ denotes the Gaussian density of each component given as

$$f(\mathbf{X}_n | \mu_n, \Sigma_n) = \frac{1}{2\pi(\det \Sigma_n)^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}(\mathbf{X} - \mu_n)^T \Sigma_n^{-1}(\mathbf{X} - \mu_n)\right\}. \quad (7)$$

In (7), μ_n denotes the mean vector, Σ_n represents the covariance matrix and π_n is the prior probability of the n th component.

4.1 Expectation-Maximization Algorithm for Estimating the Parameters of ISM Data Traffic

The Expectation-Maximization (EM) is considered as a suitable technique for estimating the parameters of the PDF of the observed data. This algorithm finds its applications in other fields including data clustering in machine learning, reconstruction of medical images and also in computer vision. The EM scheme is performed iteratively for different number of components of a MGM with $n=1, 2, \dots, N$. Initial prior probabilities π_n are assumed to be identical among the components. The covariance matrices Σ_n are initially identity matrices, while the choice of mean vectors μ_n is made by a k -means clustering, where k -means clustering chooses the mean vectors randomly [19].

The objective here is to estimate the parameter set Θ for a MGM having N components, which maximizes the log-likelihood function (LLF)

$$\Lambda_N = \log p(\mathbf{X}_1, \dots, \mathbf{X}_M | \Theta) = \sum_{m=1}^M \log \sum_{n=1}^N \frac{\pi_n}{2\pi(\det \Sigma_n)^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}(\mathbf{X}_m - \mu_n)^T \Sigma_n^{-1}(\mathbf{X}_m - \mu_n)\right\}, \quad (8)$$

where \mathbf{X}_m represents the observation of \mathbf{X}_m for both NIF and NIT scenarios. The description of the conditional probability density of the m th observation $\mathbf{X}_m = \mathbf{x}_m$ given the n th component of the MGM distribution is given by

$$p(\mathbf{X}_m | n, \Theta) = \frac{1}{2\pi(\det \Sigma_n)^{\frac{1}{2}}} \times \exp\left\{-\frac{1}{2}(\mathbf{X}_m - \mu_n)^T \Sigma_n^{-1}(\mathbf{X}_m - \mu_n)\right\}. \quad (9)$$

The EM algorithm works iteratively where it alternates between the expectation step (E-step) to update the posterior probabilities and maximization step (M-step) to update the parameters of the components of the MGM.

E-Step: Assume that we have an estimate of $\hat{\Theta}^{(j-1)}$ as the $(j-1)$ th iteration of EM completed. In the E-Step, the conditional distribution of the n th component of the MGM given the observation $\mathbf{X}_m = \mathbf{x}_m$ is determined using the conditional probabilities $p(\mathbf{x}_m | n, \hat{\Theta}^{(j-1)})$ and $\hat{\Theta}^{(j-1)}$ according to

$$p(n | \mathbf{X}_m, \hat{\Theta}^{(j-1)}) = \frac{\pi_n p(\mathbf{X}_m | n, \hat{\Theta}^{(j-1)})}{\sum_{n=1}^N \pi_n p(\mathbf{X}_m | n, \hat{\Theta}^{(j-1)})}. \quad (10)$$

M-Step: In this step, updated estimates of the parameter set Θ are obtained using the posterior probabilities $p(n | \mathbf{x}_m, \hat{\Theta}^{(j-1)})$ determined in the E-step according to

$$\hat{\pi}_n^{(j)} = \frac{1}{M} \sum_{m=1}^M p(n | \mathbf{x}_m, \hat{\Theta}^{(j-1)}) \quad (11)$$

$$\hat{\mu}_n^{(j)} = \frac{\sum_{m=1}^M \mathbf{x}_m p(n | \mathbf{x}_m, \hat{\Theta}^{(j-1)})}{\sum_{m=1}^M p(n | \mathbf{x}_m, \hat{\Theta}^{(j-1)})} \quad (12)$$

$$\hat{\Sigma}_n^{(j)} = \frac{\sum_{m=1}^M (\mathbf{x}_m - \hat{\mu}_n^{(j-1)})(\mathbf{x}_m - \hat{\mu}_n^{(j-1)})^T p(n | \mathbf{x}_m, \hat{\Theta}^{(j-1)})}{\sum_{m=1}^M p(n | \mathbf{x}_m, \hat{\Theta}^{(j-1)})}. \quad (13)$$

4.2 Model Selection Criteria

It is well-known that the LLF in (8) is an increasing function of the number of components. There are several information-theoretic approaches for the selection of a suitable value of N to minimize the corresponding cost function. In this work, we adopt a heuristic approach in which we choose the minimum value of N for which the relative increase in the LLF is below a given threshold. It can also be expressed alternatively as

$$N = \min_{\hat{N}} \left| \frac{L_n - L_{n-1}}{L_{n-1}} \right| \leq \tau \quad (14)$$

where the threshold is chosen to be $\tau = 0.003$.

5 MGM VALIDATION

For efficient DSA, it is valuable to have the distribution of the observed data traffic. In perspective of CR, the FCF describes the frequency behavior of the ISM data traffic. The joint distribution of the neighboring frequency subbands assists secondary users to determine the temporal occupancy of the observed frequency subbands. The TCF illustrates the temporal behavior of the ISM data traffic. It is also helpful for secondary users to have information about the bandwidth occupation for the specific time duration if the joint distribution of neighboring time slots is already known.

In this section, the joint distributions of neighboring frequency subbands and neighboring time domain signals of ISM data traffic are modeled parametrically using MGMs as mentioned in Section 4. Moreover, MGM is also validated using FCF and TCF for neighboring frequency subbands and the neighboring time domain signals, respectively.

5.1 MGM Validation using Frequency Correlation

5.1.1 Two Neighboring Frequency Subbands

In order to validate the MGM using FCF, the estimated parameters of MGM are required. The analytical expression of the correlation ρ_f is obtained from the MGM and yields the MGM estimator

$$\hat{\rho}_{f,\text{MGM}}^{\Delta k} = \rho_f(\hat{\Theta}) = \sum_{n=1}^N \hat{\pi}_n (\hat{\varepsilon}_n + \hat{\mu}_{1n} \hat{\mu}_{2n}) \quad (15)$$

where $\hat{\varepsilon}_n$ represents the estimated cross-covariance between the frequency subbands, $\hat{\mu}_n$ denotes the estimated mean vector of the components and $\hat{\pi}_n$ are the estimated prior probabilities of the MGM model. In the NIF case, the frequency subbands are denoted by the vector $\mathbf{X}_\lambda = \mathbf{X}_{\text{NIF},\lambda}$ as mentioned in Section 4. For the

two neighboring frequency subbands where $l_0 = 1$, $M = L = 1000$ and $k = 99$, the number of components $N = 9$ is calculated by (14). The scatter plot of the measured data and modeled joint distribution of the data is shown in Figure 7.

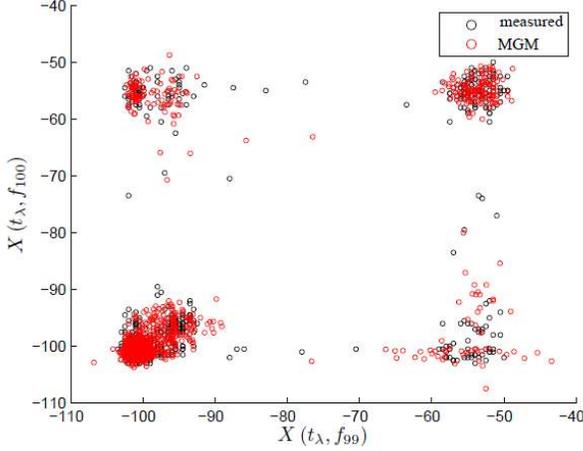


Figure 7. Joint distribution of $\mathbf{X}_{\text{NIF},\lambda}$ with $l_0 = 1$, $M = L = 1000$, $k = 99$ and $N = 9$ modeled using MGM.

Figure 7 clearly shows the close agreement between the L measurements in neighboring frequency subbands and M samples from the MGM model.

The MGM estimator for the two neighboring frequency subbands when $\Delta k = 1$ is expressed as

$$\hat{\rho}_{f,\text{MGM}}^1 = \rho_f(\hat{\Theta}) = \sum_{n=1}^N \hat{\pi}_n (\hat{\varepsilon}_n + \hat{\mu}_{1n} \hat{\mu}_{2n}). \quad (16)$$

The values of the components' mean vector are negligible. The estimated values of priors and cross-covariances are given in Table 2.

Table 2. Estimated parameters of MGM for two neighboring frequency subbands with $N = 9$

| | | | | | | | | | |
|-----------------------|------------|-----------------|------------|------------|------------|------------|------------|------------|------------|
| $\hat{\pi}_n$ | 0.02 91 | 0.03 03 | 0.18 46 | 0.11 60 | 0.00 66 | 0.04 13 | 0.40 12 | 0.16 31 | 0.02 77 |
| $\hat{\varepsilon}_n$ | 0.46 73 | - 0.78 24 | 0.44 03 | 0.59 28 | 52.5 53 | 0.00 51 | 0.01 62 | 0.09 03 | 0.79 95 |

We substitute these values in (16), which yields a value of $\hat{\rho}_{f,\text{MGM}}^1 = 0.54$ and is in close agreement to $\hat{\rho}_{f,\text{LSE}}^1 = 0.56$ calculated from the modeled FCF in Section 3.

5.1.2 Three Neighboring Frequency Subbands

Here we extend our validation of the MGM model to three neighboring frequency subbands where $l_0 = 1$, $M = L = 1000$ and $k = 106$. In this case, the number of components for $\mathbf{X}_\lambda = \mathbf{X}_{\text{NIF},\lambda}$ is estimated to be $N = 10$. The scatter plot of the measured data and modeled joint distribution of the data is shown in Figure 8.

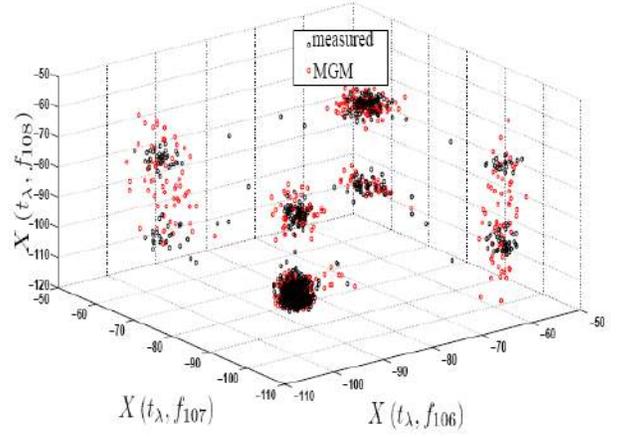


Figure 8. Joint distribution of $\mathbf{X}_{\text{NIF},\lambda}$ with $l_0 = 1$, $M = L = 1000$, $k = 106$ and $N = 10$ modeled using MGM.

As was the case in Figure 7, the scatter plot in Figure 8 also shows a good fit between the measured and modeled data.

The MGM estimator for the three neighboring frequency subbands with $\Delta k = 1$ and $\Delta k = 2$ is given by (17) and (18) respectively,

$$\hat{\rho}_{f,\text{MGM}}^1 = \rho_f(\hat{\Theta}) = \sum_{n=1}^N \hat{\pi}_n (\hat{\varepsilon}_n + \hat{\mu}_{1n} \hat{\mu}_{2n} \hat{\mu}_{3n}) \quad (17)$$

$$\hat{r}_{f,\text{MGM}}^2 = r_f(\hat{\Theta}) = \hat{\mathbf{a}} \hat{\rho}_n (\hat{\varepsilon}_n + \hat{m}_{1n} \hat{m}_{2n} \hat{m}_{3n}). \quad (18)$$

After substituting the estimated parameters from MGM in (17) we have the value of $\hat{\rho}_{f,\text{MGM}}^1 = 0.55$ between the frequency subbands f_{106} and f_{107} whereas $\hat{\rho}_{f,\text{MGM}}^1 = 0.51$ is obtained between frequency subbands f_{107} and f_{108} , both the estimated values of $\hat{\rho}_{f,\text{MGM}}^1$ are very close to $\hat{\rho}_{f,\text{LSE}}^1 = 0.56$ estimated in Section 3. Moreover, $\hat{\rho}_{f,\text{MGM}}^2 = 0.24$ is obtained after substituting the

estimated parameters in (18) also approximately in agreement with $\hat{\rho}_{f,\text{LSE}}^2 = 0.31$.

5.2 MGM validation using Time Correlation

5.2.1 In Two Neighboring Time Slots

To validate the MGM model of the ISM data traffic using TCF, the required MGM estimator is

$$\hat{r}_{t,\text{MGM}}^{DI} = r_t(\hat{\Theta}) = \hat{\mathbf{a}} \hat{\rho}_n (\hat{\varepsilon}_n + \hat{m}_{1n} \hat{m}_{2n}). \quad (19)$$

In the NIT case, the vectors of neighboring time domain signals are denoted as $\mathbf{X}_\kappa = \mathbf{X}_{\text{NIT},\kappa}$.

For the case of two neighboring time domain signals where $k_0 = 1$, $M = K = 285$ and $l = 477$, the number of components estimated by (14) is to $N = 8$. A close agreement between the measured data and the modeled data based on the estimated parameter set $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ as shown in Figure 9 validates our assumption for the NIT case.

The modeled MGM is also validated for two neighboring time domain signals using the given estimator. The estimated values of priors and cross-covariances are given in Table 3.

Table 3. Estimated parameters of MGM for two neighboring time domain signals with $N = 8$

| | | | | | | | | |
|-----------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| $\hat{\pi}_n$ | 0.253 | 0.017 | 0.018 | 0.450 | 0.024 | 0.020 | 0.124 | 0.090 |
| | 7 | 5 | 3 | 2 | 3 | 6 | 5 | 8 |
| $\hat{\varepsilon}_n$ | 0.373 | 17.29 | 17.69 | 0.004 | 2.782 | 0.615 | 0.023 | 0.612 |
| | 8 | 0 | 3 | 3 | 2 | 5 | 0 | 2 |

After inserting the values of estimated parameters from Table 3 in the MGM estimator

$$\hat{r}_{t,\text{MGM}}^1 = r_t(\hat{\Theta}) = \hat{\mathbf{a}} \hat{\rho}_n (\hat{\varepsilon}_n + \hat{m}_{1n} \hat{m}_{2n}). \quad (20)$$

We obtain $\hat{\rho}_{t,\text{MGM}}^1 = 0.2037$ with $\Delta l = 1$ which is clearly close to the estimated value of $\hat{\rho}_{f,\text{LSE}}^1 = 0.23$ using TCF in Section 3.

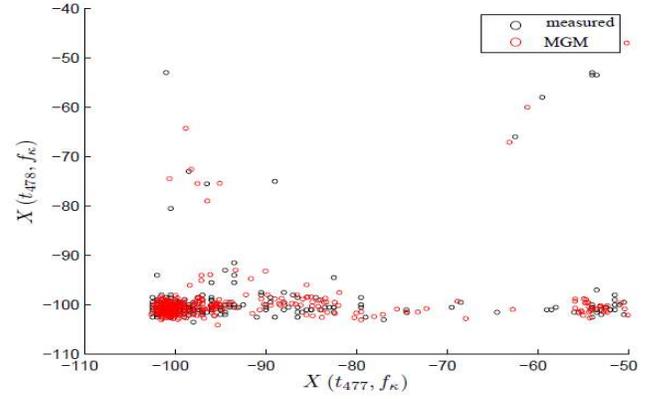


Figure 9. Joint distribution of $\mathbf{X}_{\text{NIT},\kappa}$ with $k_0 = 1$, $M = K = 285$, $l = 477$ and $N = 8$ modeled using MGM.

5.2.2 Three Neighboring Time Slots

In case of the extension of MGM for three neighboring time domain signals where $k_0 = 1$, $M = K = 285$ and $l = 777$, the estimated number of components for three neighboring time slots using (14) is $N = 10$. In Figure 10, the scatter plots of the measured data and MGM samples show a good fit.

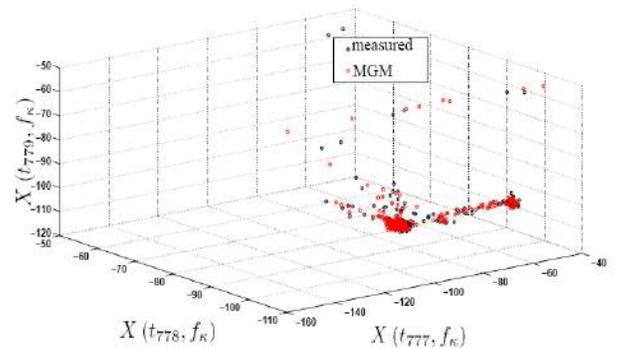


Figure 10. Joint distribution of $\mathbf{X}_{\text{NIT},\kappa}$ with $k_0 = 1$, $M = K = 285$, $l = 777$ and $N = 10$ modeled using MGM.

The MGM estimator for three neighboring time domain signals with $\Delta l = 1$ and $\Delta l = 2$ is formulated as

$$\hat{\rho}_{t,\text{MGM}}^1 = \rho_t(\hat{\Theta}) = \sum_{n=1}^N \hat{\pi}_n (\hat{\varepsilon}_n + \hat{\mu}_{1n} \hat{\mu}_{2n} \hat{\mu}_{3n}) \quad (21)$$

$$\hat{\rho}_{t,\text{MGM}}^2 = \rho_t(\hat{\Theta}) = \sum_{n=1}^N \hat{\pi}_n(\hat{\varepsilon}_n + \hat{\mu}_{1n}\hat{\mu}_{2n}\hat{\mu}_{3n}). \quad (22)$$

After substituting the estimated parameters from MGM in (21) we find that the value of $\hat{\rho}_{t,\text{MGM}}^1 = 0.19$ estimated between the signals at time slots t_{777} and t_{778} with $\hat{\rho}_{t,\text{MGM}}^1 = 0.25$ estimated between the signals at time slots t_{778} and t_{779} are in close agreement with $\hat{\rho}_{t,\text{LSE}}^1 = 0.23$. The value of $\hat{\rho}_{t,\text{MGM}}^2 = 0.08$ obtained after substituting the estimated parameters in (22) is also nearly equal to the $\hat{\rho}_{t,\text{LSE}}^2 = 0.06$ estimated using TCF in Section 3.

6 CONCLUSIONS

We conduct indoor measurement campaigns to analyze the RF activity. The frequency and time correlation functions of measured data traffic are modeled by decaying exponentials. MGM is considered to model the observed data traffic in neighboring frequency subbands and also at neighboring time instances by adopting a parametric approach. In order to estimate the parameters of MGM, the EM algorithm is used. The selection of a suitable initial mean vector is done using k -means clustering in a heuristic way. It is found that using real time measurements in CR, a suitable model for the observed data in the two neighboring frequency subbands is MGM with $N = 9$ and for the extended case of three neighboring frequency subbands with $N = 10$. For the observed data traffic in two neighboring time slots, MGM is selected with $N = 8$ and for the case of three neighboring time domain signals, the suitable choice is MGM with $N = 10$. It is also validated that the MGM estimator provides an accurate correlation between the frequency subbands and time domain signals.

REFERENCES

[1] J. Xue, Z. Feng, and K. Chen, "Beijing spectrum survey for cognitive radio applications," in Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th, Sept 2013, pp. 1–5.
 [2] A. A. Ayeni, N. Faruk, N. T. Surajudeen-Bakinde, R. A. Okanlawon, and Y. A. Adediran, "Spatial spectrum

utilization efficiency metric for spectrum sharing system," International Journal of Digital Information and Wireless Communications (IJDIWC), vol. 5, no. 1, pp. 44–51, 2015.
 [3] M. Biggs, A. Henley, and T. Clarkson, "Occupancy Analysis of the 2.4 GHz ISM Band," Communications, IEE Proceedings-, vol. 151, no. 5, pp. 481–488, Oct 2004.
 [4] V. N. Q. Bao, L. Q. Cuong, L. Q. Phu, T. D. Thuan, N. T. Quy, and L. M. Trung, "Vietnam spectrum occupancy measurements and analysis for cognitive radio applications," in Advanced Technologies for Communications (ATC), 2011 International Conference on, Aug 2011, pp. 135–143.
 [5] D. Roberson, C. Hood, J. L. LoCicero, and J. MacDonald, "Spectral occupancy and interference studies in support of cognitive radio technology deployment," in Networking Technologies for Software Defined Radio Networks, 2006. SDR '06. 1st IEEE Workshop on, Sept 2006, pp. 26–35.
 [6] S. Pagadarai and A. M. Wyglinski, "A quantitative assessment of wireless spectrum measurements for dynamic spectrum access," in Cognitive Radio Oriented Wireless Networks and Communications, 2009. CROWNCOM '09. 4th International Conference on, June 2009, pp. 1–5.
 [7] M. Dzulkifli, M. Kamarudin, and T. Rahman, "Spectrum occupancy at uhf tv band for cognitive radio applications," in RF and Microwave Conference (RFM), 2011 IEEE International, Dec 2011, pp. 111–114.
 [8] K. Patil, K. Skouby, and R. Prasad, "Stochastic duty cycle model based on measurement for cognitive radio," in Wireless Personal Multimedia Communications (WPMC), 2012, 15th International Symposium on, Sept 2012, pp. 128–132.
 [9] M. Dzulkifli, M. Kamarudin, and T. Rahman, "Spectrum occupancy of malaysia radio environment for cognitive radio application," in Wireless Communications and Applications (ICWCA 2012), IET International Conference on, Oct 2012, pp. 1–6.
 [10] F. C. Miguel Lopez-Benitez, "Spectrum occupancy in realistic scenarios and duty cycle model for cognitive radio," Advances in Electronics and Telecommunications, vol. 1, no. 1, pp. 26–34, April 2010.
 [11] L. Yin, S.-X. Yin, S. Wang, E.-Q. Zhang, W.-J. Ho, and S. fang LI, "Quantitative spectrum occupancy evaluation in china: based on a large scale concurrent spectrum measurement," The Journal of China Universities of Posts and Telecommunications, vol. 19, no. 3, pp. 122–128, 2012.
 [12] A. Pintor, M. To, J. Salenga, G. Geslani, D. Agpawa, and M. Cabatuan, "Spectrum survey of vhf and uhf bands in the philippines," in TENCON2012 - 2012 IEEE Region 10 Conference, Nov 2012, pp. 1–6.
 [13] C. Ghosh, S. Pagadarai, D. Agrawal, and A. M. Wyglinski, "A framework for statistical wireless spectrum occupancy modeling," Wireless Communications, IEEE Transactions on, vol. 9, no. 1, pp. 38–44, January 2010.
 [14] K. K. Toshiyasu Kurasugi and Y. Takahashi, "A markovian model of coded video traffic which exhibits long-range dependence in statistical analysis," Journal of the Operations Research, vol. 42, no. 1, 1999.

- [15] M. Ehsan and D. Dahlhaus, "Statistical modeling of ism data traffic in indoor environments for cognitive radio systems," in *Digital Information, Networking, and Wireless Communications (DINWC)*, 2015 Third International Conference on, Feb 2015, pp. 88–93..
- [16] M. F. Tommaso Costa, Giuseppe Boccignone, "Gaussian mixture model of heart rate variability," *PLOS ONE*, vol. 7, no. 5, May 2012.
- [17] G. Mclachlan, *Finite Mixture Models*. Cambridge, MA: John Wiley & Sons, Inc., 2000.
- [18] D. Choudhary and A. Robinson, "A new approach to wireless channel modeling using finite mixture models," *International Journal of Digital Information and Wireless Communications (IJDIWC)*, vol. 4, no. 2, pp.44 –51, 2014.
- [19] Seber and G.A.F, *Multivariate Observations*. Hoboken, NJ: John Wiley and Sons, 1984.