Abstract—This work addresses the problem of accurately modeling the spectrum occupancy patterns of real radio communication systems, an essential aspect in the study of cognitive radio networks. The main drawbacks and limitations of previous works are identified and the methodological procedures on which they rely are improved and extended. Two sophisticated measurement platforms, providing low and high time resolutions, are used to obtain extensive real-world data from a multi-band spectrum measurement campaign, embracing a wide variety of spectrum bands of practical interest for cognitive radio applications. A comprehensive, systematical and rigorous analysis of the statistical properties observed in the measurement data is then performed in order to find accurate models capable to capture and reproduce, within reasonable complexity limits, the statistical properties of temporal patterns, at both short and long timescales, in real wireless systems. Innovative modeling approaches capable to simultaneously describe statistical properties at both timescales are developed as well. In summary, this work contributes realistic and accurate time-dimension spectrum usage models for their application to the study and development of cognitive radio.

Index Terms—Cognitive radio, dynamic spectrum access, spectrum usage models, time dimension.

I. INTRODUCTION

Cognitive Radio (CR) has become one of the most intensively studied paradigms in wireless communications [1]–[4]. A CR is a context-aware intelligent radio capable of autonomous reconfiguration by learning from and adapting to the communication environment. An important specific application of CR is Dynamic Spectrum Access (DSA) [2], [5]. Despite being a broader concept [6]–[8], DSA is commonly understood as an opportunistic spectrum access method whereby unlicensed (secondary) systems are allowed to access, in a non-interfering manner, licensed bands not occupied by the licensed (primary) systems for a certain period of time. Secondary users take profit of opportunities arise when the primary system remains inactive (idle most of the time). This work focuses on the time dimension of DSA. Temporal opportunities arise when the primary system remains inactive for a certain period of time. Secondary users take profit of these inactivity periods (white spaces or spectrum holes [9]) to opportunistically access the spectrum. The DSA/CR concept has been motivated by the results of spectrum measurement campaigns performed all over the world over both wide frequency ranges [10]–[22] and specific bands [23]–[29], which demonstrated that spectrum remains idle most of the time. This suggests that new communication systems based on DSA/CR can coexist with legacy systems in the same spectrum, thus leading to a more efficient exploitation of the spectrum.

Owing to the opportunistic nature of the DSA/CR principle, the behavior and performance of a secondary network depends on the spectrum occupancy patterns of the primary system. A realistic and accurate modeling of such patterns becomes therefore essential and extremely useful in the domain of DSA/CR research [30]. Spectrum usage models can be employed in analytical studies, the design and dimensioning of DSA/CR networks, the implementation of new simulation tools and the development of more efficient DSA/CR techniques. However, the practical utility of such models depends on their degree of realism and accuracy. Unfortunately, the models of spectrum usage widely accepted and commonly used to date by the research community are limited in scope, and based on oversimplifications or assumptions that have not been validated with empirical measurement data. In this context, this work presents models that accurately capture and reproduce the statistical properties of temporal spectrum opportunities in real radio communication systems.

The modeling of spectrum occupancy in the time dimension from a discrete-time viewpoint was addressed in [31]. In contrast, this work addressed the problem from a continuous-time perspective. From the viewpoint of DSA/CR, the occupancy pattern of a Radio Frequency (RF) channel can be modeled by means of a Continuous-Time Markov Chain (CTMC) with two states, namely busy (i.e., channel occupied by a primary user and therefore not available for opportunistic access) and idle (i.e., available for secondary use). According to this model, the channel is constantly changing to the alternative state after having remained in the previous state for a random time interval, referred to as state holding time or sojourn time, which is modeled as an exponentially distributed random variable. The CTMC model has been widely employed in the study of various aspects of DSA/CR networks such as Medium Access...
Control (MAC) protocols for spectrum sharing [32], [33], MAC-layer sensing schemes [34]–[36], adaptive spectrum sensing solutions [37], the sensing-throughput tradeoff [38], [39], and the performance of DSA/CR sensor networks [40]. However, the analysis of field measurements demonstrates that the lengths of busy and idle periods in real systems are not exponentially distributed and the CTMC model is therefore unrealistic. An alternative channel model is the Continuous-Time Semi-Markov Chain (CTSMC) model, where the state holding times can follow any arbitrary distribution. Based on this modeling approach, previous works have attempted to characterize spectrum occupancy patterns [41]–[45], but unfortunately, they lack of a sufficiently comprehensive treatment of the problem. In this context, this work aims to cover the deficiencies of previous studies. The main drawbacks and limitations of previous works reported in the literature are identified and the methodological approaches and procedures on which they rely are improved and extended. A comprehensive, systematical and rigorous analysis of the statistical properties observed in field measurements of real wireless systems is performed, and innovative modeling approaches are developed as well.

The remainder of this work is organized as follows. First, Section II reviews previous related studies. Section III then summarizes the main novelties of this work, highlighting the differences with respect to related previous studies. Two sophisticated measurement platforms, providing low and high time resolutions, are used to obtain real-world empirical data, which are described in Section IV. The considered probability distribution models and the goodness-of-fit metrics employed to assess their suitability in fitting the empirical data are presented in Sections V and VI, respectively. The most suitable distribution models are discussed in Sections VII and VIII, based respectively on low and high time-resolution measurements. Additionally, two innovative modeling approaches to simultaneously describe spectrum occupancy statistics at both short and long timescales are proposed in Section IX. Finally, Section X summarizes the work.

II. PREVIOUS WORK

Based on a CTSMC, a model that statistically describes the busy and idle periods of an IEEE 802.11b Wireless Local Area Network (WLAN) is proposed in [41]–[43]. The model is based on data obtained from measurements performed with a vector signal analyzer in the 2.4 GHz ISM band, considering a controlled laboratory setup and under high Signal-to-Noise Ratio (SNR) conditions. The experimental setup considers a traffic source of UDP packets with a constant packet length (512 bytes) and Poisson-distributed inter-departure times at different rates [41], as well as more realistic traffic sources such as FTP [42], VoIP [42] and HTTP [43] streams generated by real applications. The high sampling rate provided by vector signal analyzers enables time accuracies down to the symbol level and thereby the identification of the IEEE 802.11b MAC protocol behavior in the captured traces. The sequence of states corresponding to data transmission and acknowledgment is found to be essentially deterministic, which results in a deterministic sojourn time in the busy state. The idle sojourn time is fitted to a generalized Pareto distribution [41], a mixture of uniform distribution (associated to the effects of the contention window) and generalized Pareto distribution (associated to truly unused channels) [42], [43], and a hyper-Erlang distribution [42], [43], which represents a good tradeoff between accuracy and tractability of the model.

While [41]–[43] considers an interference-controlled environment with a single packet flow artificially generated, [44] analyzes the distribution of busy and idle periods in a real environment with heterogeneous wireless devices operating in the 2.4 GHz ISM band. The study reported in [44] concludes that complex models such as the hyper-exponential distribution provide excellent fits, but simpler models such as the generalized Pareto distribution still lead to good matches with empirical data, thus providing a reasonable tradeoff between complexity and accuracy.

The work reported in [45] performs a similar study over a wider set of spectrum bands and based on a spectrum analyzer. Spectrum analyzers are characterized by significantly lower sampling rates, which may result in under-sampling of the measured signals, but enable high dynamic ranges, high sensitivity levels and broadband measurements. The work performed in [45] concludes that state holding times can appropriately be described by means of geometric distributions. For channels with low (high) loads, the duration of idle (busy) periods increases notably, leading to heavy-tailed distributions for which a log-normal model is found to provide more accurate fits.

III. NOVELTIES OF THIS STUDY

This section identifies the main drawbacks and limitations of previous modeling studies reported in the literature and explains how they are overcome in this work.

- High time-resolution measurement equipments have been employed in [41]–[43] (vector signal analyzer) and [44] (wireless transceiver in a laptop) to obtain spectrum occupancy data in the 2.4 GHz ISM band. Although high time resolutions enable more accurate models, the studies performed in [41]–[44] focus on the 2.4 GHz ISM band exclusively. A wider set of spectrum bands have been embraced by the study performed in [45], but making use of a low time-resolution device (spectrum analyzer). This work jointly employs both low and high time-resolution devices to measure the spectral activity in a wide range of allocated spectrum bands and discusses the consequences of different time resolutions on the resulting models.
- As opposed to some previous works [41]–[43] where a single traffic flow is generated and measured, this chapter exhaustively measures the occupancy patterns of a significantly high number of channels for each analyzed spectrum band, which ensures that the resulting models are representative of the true spectrum occupancy of channels in real wireless systems.
- Previous works have tried to fit a reduced number of probability distributions to empirical data, in some cases considering complex models such as phase-type distributions (e.g., hyper-Erlang or hyper-exponential) that are
obtained as a linear (weighted) combination of a number of simpler distributions of the same class. Such complex models have been shown to provide good accuracy levels and can be implemented in simulation tools. However, their applicability in analytical studies appears to be difficult due to their complex mathematical expressions and the high number of model parameters involved. By contrast, this work analyzes the suitability of a wider set of simpler distributions, some of which have not been considered before, showing that satisfactory fits can be achieved as well.

- Model parameters have usually been derived from empirical data based on Maximum Likelihood Estimation (MLE) techniques [41]–[44]. This work also considers the Method Of Moments (MOM) and evaluates the resulting fits under both inference methods.
- The fits for various distribution models have solely been evaluated based on distance-type metrics. This type of metrics provide a single indication of the goodness of the fit for a certain probability model over the whole range of values of the parameter under study (i.e., the duration of busy and idle periods in this case). Although this type of metrics is also considered in this work, the fit of the considered models in particular regions of the parameter under study (i.e., short and long periods) is individually evaluated as well.

In summary, this work provides an adequate treatment of the problem by performing a comprehensive, systematical and rigorous study on the probability distributions that can be employed to accurately describe the statistical properties of spectrum usage in real wireless systems.

IV. MEASUREMENT SETUP AND METHODOLOGY

Two measurement platforms, providing low and high time resolutions, are employed in this study. The first measurement platform (see Figure 1) relies on a spectrum analyzer setup where different external devices have been added to improve the detection capabilities. The design is composed of two broadband discone-type antennas (75-7075 MHz), a Single-Pole Double-Throw (SPDT) switch to select the desired antenna, several filters to remove undesired out-of-band and overloading FM signals, a low-noise pre-amplifier to enhance the sensitivity (the overall noise figure of the whole platform is 4 dB), and a high performance spectrum analyzer to record the spectral activity. The spectrum analyzer is connected to a laptop via Ethernet and controlled by a tailor-made software based on the Matlab’s Instrument Control Toolbox. A more detailed description of this platform can be found in [46], [47].

The second measurement platform (see Figure 2) relies on a Universal Software Radio Peripheral (USRP). Target signals are captured with a broadband discone-type antenna (75-3000 MHz) and down-converted by TVRX (50-860 MHz, 8 dB typical noise figure) and DBS2RX (800-2400 MHz, 3-5 dB typical noise figure) RF front-end boards to the Intermediate Frequency (IF) at which the main USRP board performs sampling and filtering. Digital signal samples are down-converted (decimated) to Base Band (BB) and sent via USB to a laptop running the GNU Radio software, where signal samples are saved to files for off-line processing and analysis. A more detailed description of this platform can be found in [48]. Both measurement platforms provide complimentary characteristics. On the one hand, USRP can handle up to 8 MHz bandwidth, meaning that only one or a few RF channels can be measured at a time. Moreover, due to its high sampling rate (1 µs between samples) and the resulting huge volume of generated data, a channel can be monitored continuously for a relatively short period (20 minutes in our experiments1). However, high time-resolution measurements are useful to accurately extract the true occupancy pattern of RF channels (see Figure 3). On the other hand, spectrum analyzers can handle entire spectrum bands, and due to their lower sampling rates (in our experiments, 2.58-5.70 seconds between samples, depending on the considered band), can be used for much longer measurements (7 days in our experiments). The low effective sampling rates of spectrum analyzers, however, result in a significant under-sampling of the measured signals (the channel state may change between two consecutive observations as illustrated in Figure 3). The occupancy pattern observed in such a case, although inaccurate, is interesting as it can be thought of as the perception of a DSA/CR user that periodically senses the channel and observes its state at discrete time instants. Thus, while USRP measurements are useful to accurately describe the true channel occupancy pattern at short timescales, spectrum analyzer measurements are useful to model the occupancy pattern perceived by DSA/CR users at longer timescales.

Both measurement setups were employed to monitor spectrum bands allocated to amateur systems (144–146 MHz), paging systems (157–174 MHz), Private/Public-Access Mobile Radio (PMR/PAMR) systems such as TETRA UL (410–420 MHz) and TETRA DL (420–430 MHz), cellular mobile

1For most radio technologies, a measurement period of 20 minutes is enough to obtain a sufficiently large number of samples of the period durations and derive statistically reliable estimations of the empirical distributions at short timescales. In a few particular cases, however, several measurement sessions of 20 minutes were required to obtain a sufficiently large sample.
communication systems such as E-GSM 900 UL (880–915 MHz), E-GSM 900 DL (925–960 MHz), DCS 1800 UL (1710–1785 MHz) and DCS 1800 DL (1805–1880 MHz), cordless telephone systems such as DECT (1880–1900 MHz), and open bands such as ISM (2400–2500 MHz). Most of these bands were measured from a strategically selected building rooftop with direct line-of-sight to several transmitters a few tens or hundreds of meters apart. For the DECT and ISM bands, however, measurements were performed in indoor environments, where short-range devices using these bands are commonly deployed. Measurement locations were carefully selected to maximize the receiving SNR and ensure a reliable and accurate estimation of the true busy/idle states for the measured channels. Although this work does not present detailed results for all the considered bands, extensive and detailed analyses were performed for all of them.

Binary busy/idle channel occupancy patterns were extracted from spectrum data based on energy detection [49]. Power samples provided by the spectrum analyzer were processed individually, leading to the effective time resolutions mentioned above (2.58–5.70 seconds between samples). USRP individually, leading to the effective time resolutions mentioned above (2.58–5.70 seconds between samples). USRP spectrum analyzer data were processed based on the classical energy detection method [49] (as in [31]), while the Improved Energy Detection (IED) scheme presented in [50] (with parameter \( L = 5 \)) was used to extract binary occupancy information from USRP data. The IED method makes use of past channel observations to determine the current channel state, which is not sensible when the time period between consecutive observations is in the order of several seconds. This prevents the application of the IED scheme to the spectrum analyzer data. However, the high time resolution of the USRP platform enables the application of the IED method, which was experimentally observed to result in an improved detection performance and a more accurate estimation of the true channel occupancy.

V. CONSIDERED PROBABILITY DISTRIBUTIONS

Based on the binary occupancy patterns extracted from empirical data, the length of busy and idle periods was computed

\[ \tau \]

Fig. 3. Low versus high time-resolution measurements.

For each RF channel and the corresponding empirical Cumulative Distribution Function (CDF) was derived and compared to the probability distribution models shown in Table I. Some complex distributions studied in previous works (phase-type distributions such as hyper-Erlang or hyper-exponential) are not considered. Such distributions have been proven to be accurate but involve complex mathematical expressions and a high number of parameters, which hinders their application in analytical studies. By contrast, this work considers a wider set of simpler and more tractable models, some of which have not been considered before. The considered CDF models, as it will be shown, can provide satisfactory fits to empirical data.

The exponential distribution is analyzed to determine the validity of the widely employed CTMC model. The interest of the generalized exponential distribution [51] relies on its ability to reproduce other distributions with a single analytical expression: for \( \alpha = 1 \) it becomes the exponential distribution, while for certain ranges of the shape \( \alpha \) and scale \( \lambda \) parameters it closely resembles the log-normal, gamma, and Weibull distributions, which are explicitly considered as well. The log-normal distribution has been suggested as an adequate model for heavy-tailed trends [45], while the suitability of the gamma and Weibull distributions has not been studied before. The Pareto and generalized Pareto distributions, considered in previous studies, are also analyzed.

It is worth noting that the exponential and Pareto distributions are particular cases of their generalized counterparts when \( \alpha = 1 \) and \( \mu = \lambda/\alpha \), respectively. In such cases, the relations \( \mu_{GE} = \mu_{E} \) and \( \lambda_{GE} = \lambda_{E} \) hold for the former, while \( \lambda_{GP} = \lambda_{P}/\alpha_{P} \) and \( \alpha_{GP} = 1/\alpha_{P} \) hold for the latter (see Table I). There is a reason, however, to explicitly consider the particular cases, instead of solely considering the generalized distributions. When the particular cases are sufficient to provide accurate fits, the numerical methods employed to estimate the parameters of the generalized distributions do not necessarily lead to numerical values satisfying \( \alpha = 1 \) and \( \mu = \lambda/\alpha \), and hence the need to explicitly consider the particular cases in order to identify those situations where simpler distributions with fewer parameters suffice. In some cases, a better accuracy may be obtained with the particular cases than with the generalized distributions, which is merely an artifact of the employed numerical methods and indicates the suitability of the particular cases. This circumstance also highlights the need to explicitly consider the particular cases of the general distributions in order to guarantee optimum fits, an aspect that seems to have been neglected in previous works, where only generalized distributions have been considered.

As mentioned in Section III, the distributions parameters have frequently been derived from the empirical data based on MLE techniques [41]–[44]. MLE is also employed in this work to estimate the best fitting parameters for the exponential [53]–[55], generalized exponential [51], Pareto [56], generalized Pareto [57]–[59], log-normal [53], [54], [60], [61], gamma [53], [54], [60], [61], and Weibull [53]–[55] distributions. Additionally, MOM inference techniques are also considered, which consist in equating statistical moments with sample moments and then solving those equations for the estimated parameters [62]. MOM parameter estimates are computed.
TABLE I
CONSIDERED PROBABILITY DISTRIBUTION MODELS. DISTRIBUTION NAMES: E (EXPONENTIAL), GE (GENERALIZED EXPONENTIAL), P (PARETO), GP (GENERALIZED PARETO), LN (LOG-NORMAL), G (GAMMA), AND W (WEIBULL). DISTRIBUTION PARAMETERS: μ (LOCATION), λ (SCALE), AND α (SHAPE). T_i REPRESENTS THE PERIOD LENGTH. E{ } AND V{ } REPRESENT THE MEAN AND THE VARIANCE OF THE DISTRIBUTION, RESPECTIVELY.

<table>
<thead>
<tr>
<th>Distribution Function</th>
<th>Parameters</th>
<th>Moments</th>
</tr>
</thead>
<tbody>
<tr>
<td>F_E(T_i; μ, λ) = 1 − e^{−λ(T_i−μ)}</td>
<td>T_i ≥ μ &gt; 0, λ &gt; 0</td>
<td>E{T_i} = μ + \frac{1}{λ}</td>
</tr>
<tr>
<td>F_{GE}(T_i; μ, λ, α) = 1 − e^{−{(T_i−μ)/α}^α}</td>
<td>T_i ≥ μ &gt; 0, λ &gt; 0, α &gt; 0</td>
<td>E{T_i} = μ + \frac{2(α+1)/2}{λ}</td>
</tr>
<tr>
<td>F_P(T_i; λ, α) = 1 − \left(\frac{λ}{α}\right)^α</td>
<td>T_i ≥ λ, λ &gt; 0, α &gt; 2</td>
<td>E{T_i} = \frac{αλ}{(α−1)^2(α−2)}</td>
</tr>
<tr>
<td>F_{GP}(T_i; μ, λ, α) = 1 − \left[1 + \frac{e(T_i−μ)}{λ}\right]^{−1/α}</td>
<td>T_i ∈ [μ, μ − \frac{λ}{α} (α &lt; 0), μ − \frac{λ}{α} (α &gt; 0)]</td>
<td>E{T_i} = μ + \frac{α}{λ}</td>
</tr>
<tr>
<td>F_{LN}(T_i; μ, λ) = \frac{1}{2} \left[1 + \text{erf} \left(\frac{ln(T_i−μ/λ)}{\sqrt{2}\sqrt{λ}}\right)\right]</td>
<td>T_i ≥ μ &gt; 0, λ &gt; 0, α &lt; 1/2</td>
<td>E{T_i} = e^{μ+λ^{1/2}}</td>
</tr>
<tr>
<td>F_G(T_i; μ, λ, α) = \frac{γ(\frac{T_i−μ}{λ}, α)}{Γ(α)}</td>
<td>T_i ≥ μ &gt; 0, λ &gt; 0, α &gt; 0</td>
<td>E{T_i} = μ + λα</td>
</tr>
<tr>
<td>F_W(T_i; μ, λ, α) = 1 − \exp \left[−\left(\frac{T_i−μ}{λ}\right)^α\right]</td>
<td>T_i ≥ μ &gt; 0, λ &gt; 0, α &gt; 0</td>
<td>E{T_i} = μ + λΓ(1 + \frac{1}{α})</td>
</tr>
</tbody>
</table>

Based on the sample mean and sample variance of the length of busy/idle periods (see Table I). While MLE is widely accepted as a preferred inference method, MOM can sometimes provide better fits to empirical data.

VI. GOODNESS-OF-FIT METRICS
In order to assess the suitability of the considered probability distributions in fitting the empirical data sets, several tests and their underlying Goodness-Of-Fit (GOF) metrics are employed. Based on the Kolmogorov-Smirnov (KS) test, the KS distance between the empirical CDF \( F_{emp}(T_i) \) of period lengths \( T_i \) (\( i \in \{0, 1\} \)) denotes the period type, with \( T_0 \) and \( T_1 \) being idle and busy periods, respectively) and the CDF model \( F_{fit}(T_i) \) is computed as [63]:

\[
D_{KS} = \max\{|F_{emp}(T_i) − F_{fit}(T_i)|\}
\]

A symmetric version of the Kullback-Leibler (KL) divergence [63] is also employed:

\[
D_{KL}^{sym} = D_{KL}(F_{emp}(T_i) || F_{fit}(T_i)) = D_{KL}(F_{fit}(T_i) || F_{emp}(T_i))
\]

where \( f_{emp}(T_k) \) and \( f_{fit}(T_k) \) represent the empirical Probability Density Function (PDF) and the evaluated PDF model, respectively, assumed to be computed for a discrete number of \( K \) values of the period length \( T_i, T_k^k (k = 1, 2, \ldots, K) \) represents the \( k \)-th value of \( T_i \), and \( D_{KL}(f_A || f_B) \) denotes the KL divergence between PDFs \( f_A \) and \( f_B \).

Finally, the Bhattacharyya distance is also employed, which is defined as [64]:

\[
D_B = − \ln \left( \sum_{k=1}^{K} \sqrt{f_{emp}(T_k^i) \cdot f_{fit}(T_k^i)} \right)
\]

These metrics provide a single numerical value indicating the GOF of a certain probability model over the whole range of values of \( T_i \). Additionally, the GOF of the considered models in particular regions of the CDF corresponding to low and high values of \( T_i \) will be evaluated as well.

VII. LONG TIMESCALE MODELS
This section analyzes the suitability of the considered CDF models in describing the length of busy and idle periods based on spectrum analyzer measurements. The high number of factors involved in the analysis were handled as follows. First, the CDF models of Table I were fitted to the empirical CDFs of busy/idle periods based on the mean and variance of the distribution. As it can be appreciated, the generalized
Pareto distribution provides the best fit in all cases, which can also be concluded based on the average KS distance. Similar trends were observed for other GOF metrics and inference methods, for both busy and idle periods, thus indicating that the analysis can be undertaken in terms of average values since such values are sufficiently representative of the results obtained for individual bands. Based on this appreciation, the GOF metrics were averaged over all bands in order to obtain the average GOF metrics shown in Table III.

In order to compare the fitting accuracy of the considered distributions, the minimum value of the GOF metrics resulting from the MLE and MOM approaches was selected as the representative result. Notice that the best possible fit (either MLE or MOM) is the one that really indicates the ability of a distribution function to describe a set of empirical data. Thus, the comparison among the considered distributions is performed based on the minimum values obtained for each GOF metric, which are shown in bold in Table III. As it can be appreciated, all the considered GOF metrics indicate that the best fit, thus indicating that the GOF of the generalized Pareto distribution is the most accurate model according to the KS distance, but not from the point of view of the symmetric KL divergence and the Bhattacharyya distance. However, in these two cases it is interesting to note that the fitting accuracy of the generalized Pareto distribution is only marginally worse than the best fit, thus indicating that the GOF of the generalized Pareto distribution in these cases can be considered to be acceptable. Based on these observations, the generalized Pareto distribution can be selected as an appropriate model for the lengths of both busy and idle periods in real systems. Other alternative models such as the generalized exponential, log-normal, gamma and Weibull distributions are able to achieve approximately 75% of the channels were best fitted with the MLE method. Therefore, MLE outperforms MOM in general, but not always provides the best fit (see for example the Weibull distribution in Table III).

The previous analysis has been based solely on distance metrics, which provide a single numerical value as an indication of the GOF of a certain probability model over the whole range of values of $T_i$. The fitting accuracy in particular regions of the CDF is analyzed in the following. Figures 4 and 5 show the empirical and fitted distributions based on the MLE and MOM methods, respectively. These figures correspond to a single selected channel from a particular spectrum band but are representative of the behavior observed for other channels and bands. Each figure is composed of four graphs. Upper graphs show the statistics of busy periods, while lower graphs show the statistics of idle periods. The graphs on the left-hand side show the empirical and fitted CDFs with the abscissa axis in logarithmic magnitude, which enables a finer detail of appreciation of the fitting accuracy for low values of $T_i$ (i.e., short periods). The graphs on the right-hand side show the same information in terms of the Complementary Cumulative Distribution Function (CCDF) with both axes in logarithmic magnitude, which allows for a finer detail of appreciation of the fitting accuracy for high values of $T_i$ (i.e., long periods).

The first interesting observation from Figures 4 and 5 is that the exponential distribution is not able to describe with sufficient accuracy the length of busy and idle periods observed in real channels, meaning that the CTMC channel model widely employed in the literature is unrealistic. It is important to note that the exponential distribution considered in this study (see Table I) includes a location parameter $\mu$ that is not usually considered in the conventional exponential distribution widely employed in the literature. Without such location parameter (i.e., $\mu = 0$), the resulting accuracy was observed to be notably worse, thus confirming that a CTMC does not constitute an adequate model. In terms of accuracy, similar comments can be made for the Pareto distribution, which suggests that simple one/two-parameter distributions do not seem to be adequate models.

For the rest of distributions, it is interesting to note that each inference method improves the fitting accuracy in one particular region at the expense of degrading the fit in the opposite region. With some particular exceptions, the best

<table>
<thead>
<tr>
<th>Band</th>
<th>E</th>
<th>GE</th>
<th>P</th>
<th>GP</th>
<th>LN</th>
<th>G</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amateur</td>
<td>0.22</td>
<td>0.14</td>
<td>0.17</td>
<td>0.08</td>
<td>0.12</td>
<td>0.14</td>
<td>0.23</td>
</tr>
<tr>
<td>Paging</td>
<td>0.26</td>
<td>0.17</td>
<td>0.23</td>
<td>0.13</td>
<td>0.15</td>
<td>0.16</td>
<td>0.25</td>
</tr>
<tr>
<td>TETRA UL</td>
<td>0.25</td>
<td>0.13</td>
<td>0.17</td>
<td>0.06</td>
<td>0.10</td>
<td>0.13</td>
<td>0.17</td>
</tr>
<tr>
<td>DCS 1800 DL</td>
<td>0.20</td>
<td>0.13</td>
<td>0.24</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.26</td>
</tr>
<tr>
<td>DECT</td>
<td>0.24</td>
<td>0.16</td>
<td>0.16</td>
<td>0.12</td>
<td>0.16</td>
<td>0.15</td>
<td>0.30</td>
</tr>
<tr>
<td>ISM</td>
<td>0.18</td>
<td>0.18</td>
<td>0.22</td>
<td>0.16</td>
<td>0.21</td>
<td>0.18</td>
<td>0.43</td>
</tr>
<tr>
<td>Average</td>
<td>0.23</td>
<td>0.15</td>
<td>0.11</td>
<td>0.14</td>
<td>0.14</td>
<td>0.15</td>
<td>0.26</td>
</tr>
</tbody>
</table>

3An exhaustive analysis for all channels and bands indicated that approximately 75% of the channels were best fitted with the MLE method. Therefore, MLE outperforms MOM in general, but not always provides the best fit (see for example the Weibull distribution in Table III).

4Previous work [65] has shown that the PDF of channel vacancy durations can be fitted with an exponential-like function of the form $f(T_i) \approx a + b \cdot \exp(-c \cdot T_i)$, with $a, b, c \in \mathbb{R}^+$, which is not an exponential distribution.
fitting accuracies in the region of short periods is normally achieved by the MLE method, while the best fit for long periods is in general attained with the MOM approach. However, the generalized Pareto distribution is the only model that simultaneously provides acceptable fits for both short and long periods, regardless of the inference method applied, which is not observed for any of the other considered distributions. This observation confirms the suitability of the generalized Pareto distribution for the state holding times of the CTSMC channel model. The obtained results demonstrated that the generalized Pareto distribution provides a remarkably good fitting accuracy for short/long busy/idle periods in channels with low/high loads over a wide range of spectrum bands and technologies.

In order to facilitate to researchers the application of the CTSMC model, Table IV provides numerical values for the distribution’s parameters extracted from field measurements. Various channel loads are shown in terms of the channel Duty Cycle (DC). To reproduce an arbitrary DC, $\Psi$, the parameters of the distribution need to be chosen such that the following equality holds [66]:

$$\Psi = \frac{\mathbb{E}\{T_1\}}{\mathbb{E}\{T_0\} + \mathbb{E}\{T_1\}}$$

(5)

where $\mathbb{E}\{T_0\}$ and $\mathbb{E}\{T_1\}$ represent the mean idle and busy periods, respectively, which are related with the parameters $\mu$, $\lambda$ and $\alpha$ as shown in Table I. The values shown in Table IV for the location parameter $\mu$ are determined by the time resolution of the spectrum analyzer. This parameter could be tailored to the particular scenario under study in order to match the considered spectrum sensing period or the minimum period of activity/inactivity of a primary RF channel.

VIII. SHORT TIMESCALE MODELS

This section analyzes the suitability of the considered CDF models in describing the length of busy and idle periods based on measurements performed with the USRP platform. Given the limited bandwidth capabilities of the USRP hardware, selected channels were analyzed instead of entire bands as in Section VII. An individual analysis for each channel was feasible in this case and, as a result, the GOF metrics were not averaged over channels belonging to the same band. As opposed to Section VII, no single distribution was observed to provide acceptable fits for all bands, thus indicating the need of an individual analysis for each considered band. Moreover, there was no clear predominance of an inference method over the other. The results shown for each distribution correspond to the inference method providing the best fitting accuracy in each case. In general, the three considered GOF metrics showed agreement on the inference method attaining the best fit. In those cases where some disagreement was observed, the inference method indicated by any two of the three GOF metrics was selected (i.e., a majority vote criterion).

An individual analysis for each band is provided in the following. The results shown for each band are representative examples of the typical channel loads observed in each case. Other load levels can be reproduced by selecting the parameters of the distribution based on (5).

A. Amateur bands

Figure 6 shows the fitted distributions for a channel measured in amateur bands. Two of the three GOF metrics indicated that the generalized Pareto distribution provides the best fit for busy periods. The KS distance, however, indicated that the best fit corresponds to the log-normal distribution. As appreciated in Figure 6, the latter provides the best fit for long busy periods, but it cannot accurately model the minimum period duration. Therefore, it can be concluded that the generalized Pareto distribution provides a better fit over the whole range of busy period durations. For idle periods, the most precise models are the generalized Pareto and Weibull distributions, but similar accuracies are attained with the generalized exponential and gamma models. To facilitate the application of these models, Table V provides the values of

3The activity patterns of IEEE 802.11 systems operating in the 2.4 GHz ISM band have already extensively been studied in [41]-[44] using high time-resolution devices and will not be considered in this section.
TABLE IV
PARAMETERS OF THE GENERALIZED PARETO DISTRIBUTION FOR BUSY AND IDLE PERIODS EXTRACTED FROM EMPIRICAL MEASUREMENT RESULTS BY MEANS OF MLE (\(T_i\) IN TABLE I EXPRESSED IN SECONDS).

<table>
<thead>
<tr>
<th>Load</th>
<th>Duty cycle</th>
<th>(\mu) (secs)</th>
<th>(\lambda)</th>
<th>(\alpha)</th>
<th>(\mu) (secs)</th>
<th>(\lambda)</th>
<th>(\alpha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low</td>
<td>0.09</td>
<td>3.5150</td>
<td>1.6960</td>
<td>0.0284</td>
<td>3.6100</td>
<td>38.3633</td>
<td>0.2125</td>
</tr>
<tr>
<td>Low</td>
<td>0.29</td>
<td>3.5150</td>
<td>2.6240</td>
<td>0.1884</td>
<td>3.5780</td>
<td>10.9356</td>
<td>0.1784</td>
</tr>
<tr>
<td>Medium</td>
<td>0.51</td>
<td>3.5150</td>
<td>5.1483</td>
<td>0.1978</td>
<td>3.5160</td>
<td>4.6583</td>
<td>0.2156</td>
</tr>
<tr>
<td>High</td>
<td>0.71</td>
<td>3.5470</td>
<td>10.7968</td>
<td>0.1929</td>
<td>3.5310</td>
<td>2.6272</td>
<td>0.2119</td>
</tr>
<tr>
<td>Very high</td>
<td>0.93</td>
<td>3.5940</td>
<td>52.8611</td>
<td>0.2377</td>
<td>3.5160</td>
<td>1.6609</td>
<td>0.0068</td>
</tr>
</tbody>
</table>

Fig. 6. Empirical CDF and fitted distributions for amateur bands.

Fig. 7. Empirical CDF and fitted distributions for paging bands.

Fig. 8. Empirical CDF and fitted distributions for PMR/PAMR bands.

the fitted parameters, along with the corresponding values of the GOF metrics and a typical DC value observed in channels of the amateur band.

B. Paging bands

Figure 7 shows the fitted distributions for a channel measured in paging bands. The numerical results of GOF metrics indicated that the best fit for busy periods is attained with the Weibull distribution. Nevertheless, Figure 7 shows that the generalized exponential and gamma distributions are also able to provide a comparable fitting accuracy, even slightly better, over the whole range of busy period lengths, which indicates that these distributions are adequate models as well. For idle periods, the best fit corresponds to the Pareto distribution. Table VI provides the values of the fitted distribution parameters, GOF metrics and typical DC value.

C. Private/public access mobile radio bands

Figure 8 shows the fitted distributions for a channel measured in the TETRA DL band, which represents an example of a PMR/PAMR system. The best fit for busy periods is attained with the generalized Pareto distribution. Nevertheless, the simpler Pareto distribution is able to attain a similar level of accuracy as well. For idle periods, the Weibull distribution provides the best fit to empirical data. It is worth noting that other channels within the TETRA DL band showed a slightly better fit with Weibull distributions for busy periods and (generalized) Pareto distributions for idle periods. This circumstance indicates that both busy and idle periods in the TETRA DL band can indistinctly be modeled with the Weibull and (generalized) Pareto distributions (in some cases, however, the Pareto distribution was observed not be accurate enough and only the generalized Pareto distribution was acceptable). Table VII provides the values of the fitted parameters, GOF metrics and typical DC value. The results in this case are shown for two different channels.
TABLE V
PARAMETERS OF BUSY AND IDLE PERIOD DISTRIBUTIONS FOR AMATEUR BANDS ($T_i$ IN TABLE I EXPRESSED IN MILLISECONDS).

<table>
<thead>
<tr>
<th>Period type</th>
<th>Fitted distribution</th>
<th>GOF metrics</th>
<th>$\alpha$</th>
<th>$\mu$ (msecs)</th>
<th>$\lambda$</th>
<th>$\alpha$</th>
<th>Duty cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Busy</td>
<td>GP</td>
<td>$D_{KS}$</td>
<td>0.15</td>
<td>2.80</td>
<td>0.35</td>
<td>6.8659 - 10^4</td>
<td>6.0417 - 10^4</td>
</tr>
<tr>
<td>Idle</td>
<td>GP</td>
<td>$D_{KS}$</td>
<td>0.07</td>
<td>1.13</td>
<td>0.14</td>
<td>2.2208 - 10^4</td>
<td>1.0230 - 10^4</td>
</tr>
<tr>
<td></td>
<td>GP</td>
<td>$D_{KL}$</td>
<td>0.10</td>
<td>0.07</td>
<td>0.05</td>
<td>9.7262 - 10^3</td>
<td>1.1801</td>
</tr>
<tr>
<td></td>
<td>GP</td>
<td>$D_{KL}$</td>
<td>0.07</td>
<td>1.15</td>
<td>0.14</td>
<td>2.2208 - 10^4</td>
<td>1.3271 - 10^4</td>
</tr>
<tr>
<td></td>
<td>GP</td>
<td>$D_{KL}$</td>
<td>0.07</td>
<td>1.15</td>
<td>0.14</td>
<td>2.2208 - 10^4</td>
<td>7.0195 - 10^4</td>
</tr>
</tbody>
</table>

TABLE VI
PARAMETERS OF BUSY AND IDLE PERIOD DISTRIBUTIONS FOR PAGING BANDS ($T_i$ IN TABLE I EXPRESSED IN MILLISECONDS).

<table>
<thead>
<tr>
<th>Period type</th>
<th>Fitted distribution</th>
<th>GOF metrics</th>
<th>$\alpha$</th>
<th>$\mu$ (msecs)</th>
<th>$\lambda$</th>
<th>$\alpha$</th>
<th>Duty cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Busy</td>
<td>W</td>
<td>$D_{KP}$</td>
<td>0.13</td>
<td>0.25</td>
<td>0.05</td>
<td>0.2560</td>
<td>0.5922</td>
</tr>
<tr>
<td>Idle</td>
<td>G</td>
<td>$D_{KP}$</td>
<td>0.14</td>
<td>0.26</td>
<td>0.09</td>
<td>0.2560</td>
<td>3.0029</td>
</tr>
</tbody>
</table>

TABLE VII
PARAMETERS OF BUSY AND IDLE PERIOD DISTRIBUTIONS FOR PMR/PAMR BANDS ($T_i$ IN TABLE I EXPRESSED IN MILLISECONDS).

<table>
<thead>
<tr>
<th>Period type</th>
<th>Fitted distribution</th>
<th>GOF metrics</th>
<th>$\alpha$</th>
<th>$\mu$ (msecs)</th>
<th>$\lambda$</th>
<th>$\alpha$</th>
<th>Duty cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Busy</td>
<td>GP</td>
<td>$D_{KS}$</td>
<td>0.08</td>
<td>0.21</td>
<td>0.03</td>
<td>0.2560</td>
<td>0.0621</td>
</tr>
<tr>
<td>Idle</td>
<td>W</td>
<td>$D_{KL}$</td>
<td>0.12</td>
<td>0.24</td>
<td>0.11</td>
<td>0.2560</td>
<td>3.1084</td>
</tr>
<tr>
<td>Busy</td>
<td>W</td>
<td>$D_{KL}$</td>
<td>0.04</td>
<td>0.23</td>
<td>0.05</td>
<td>0.2560</td>
<td>3.1084</td>
</tr>
<tr>
<td>Idle</td>
<td>P</td>
<td>$D_{KL}$</td>
<td>0.11</td>
<td>0.35</td>
<td>0.08</td>
<td>0.2560</td>
<td>1.0186</td>
</tr>
</tbody>
</table>

D. Cellular mobile communication bands

Figure 9 shows the fitted distributions for a channel measured in the E-GSM 900 DL band, which represents an example of a cellular mobile communication system (the results are also representative of the DCS 1800 system). The generalized Pareto distribution provides the best fit for busy periods over the whole range of period durations. For idle periods, the best fit is attained with the generalized exponential distribution, although the gamma distribution provides a similar accuracy. Table VIII provides the values of the fitted distribution parameters, GOF metrics and typical DC value. The location parameter (i.e., minimum period duration) should be equal to the GSM/DCS time-slot duration ($\mu = 0.577$ ms). The value shown in Table VIII is affected by the time resolution of 0.128 ms of the measurement device (see Section IV).

It is worth noting that the stair-shaped empirical CDF in Figure 9 is a natural consequence of the time-slotted structure employed by the multiple access mechanism of the GSM/DCS system, which defines a frame structure with 8 time slots of 577 $\mu$s each. This circumstance suggests the possibility of modeling busy/idle periods in time-slotted systems from a discrete perspective, where period lengths are described in terms of the number of time-slots. Based on this observation, and considering a time-slot duration of 577 $\mu$s, the empirical CDF of Figure 9 was discretized and compared to several discrete-time CDF models. In particular, geometric, Poisson and negative binomial distributions were fitted to the empirical data (see Table IX). The binomial distribution was also considered, but it was observed to be unable to fit empirical data and it is therefore not reported. Figure 10 and Table X show the fitted distributions and obtained GOF metrics. As it can be appreciated, the best fit for both busy and idle periods is attained with the negative binomial distribution. For busy periods, the Poisson and negative binomial distributions provide the same fit and are indistinguishable from each other in Figure 10. For busy periods, the optimum fit with the Poisson distribution is obtained with parameter $\lambda = 2.9239$. With the negative binomial distribution, the optimum fit is obtained with parameters $r = 3.0294 \cdot 10^3$ and $p = 0.9990$ for busy periods and parameters $r = 0.7430$ and $p = 0.0734$ for idle periods.$^7$

E. Cordless telephone bands

The continuous-time distributions presented in Section V were fitted to the channels measured in the DECT band, which represents an example of a cordless telephone system. However, no satisfactory fits were observed for this particular

$^6$The negative binomial distribution converges to the Poisson distribution as its parameter $r$ tends towards infinity. Hence, they may closely resemble each other for $r$ sufficiently large.

$^7$In its original definition, the negative binomial distribution is a discrete probability distribution of the number of successes in a sequence of Bernoulli trials before a specified (non-random) number of failures occurs. The distribution is characterized by two parameters: $r \in \mathbb{N}^*$ (the number of failures until the experiment is stopped) and $p \in [0, 1]$ (the success probability in each experiment). The original definition can be extended to real values of the parameter $r$, which is sometimes referred to as the Pólya distribution. For the purposes of this study, the distribution is regarded as a discrete probability distribution with two parameters that are fitted to empirical data.
TABLE VIII
PARAMETERS OF BUSY AND IDLE PERIOD DISTRIBUTIONS FOR CELLULAR MOBILE COMMUNICATION BANDS (T̄ in TABLE I EXPRESSED IN MILLISECONDS).

<table>
<thead>
<tr>
<th>Period</th>
<th>Fitted distribution</th>
<th>GOF metrics</th>
<th>Parameters</th>
<th>Duty cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>D_{KS}</td>
<td>D_{KS}^{α = 3}</td>
<td>D_{KL}</td>
</tr>
<tr>
<td>Busy</td>
<td>GP</td>
<td>0.13</td>
<td>3.09</td>
<td>0.37</td>
</tr>
<tr>
<td>Idle</td>
<td>GE</td>
<td>0.05</td>
<td>2.12</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0.07</td>
<td>2.13</td>
<td>0.26</td>
</tr>
</tbody>
</table>

TABLE IX

<table>
<thead>
<tr>
<th>Distribution function</th>
<th>Parameters</th>
<th>Moments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F_{Geom}(k; p) = 1 - (1 - p)^{k+1}</td>
<td>k ∈ N_0 = {0, 1, 2, 3,...}</td>
<td>E{k} = \frac{1 - p}{p}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V{k} = \frac{1 - p}{p^2}</td>
</tr>
<tr>
<td>F_{Pois}(k; λ) = e^{-λ} \sum_{i=0}^{k} \frac{λ^i}{i!}</td>
<td>k ∈ N_0 = {0, 1, 2, 3,...}</td>
<td>E{k} = λ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V{k} = λ</td>
</tr>
<tr>
<td>F_{NBin}(k; r, p) = I_p(r, k + 1)</td>
<td>k ∈ N_0 = {0, 1, 2, 3,...}</td>
<td>E{k} = \frac{(1 - p)r}{p}</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V{k} = \frac{(1 - p)r}{p^2}</td>
</tr>
</tbody>
</table>

Fig. 9. Empirical CDF and fitted distributions for cellular mobile communication bands.

Fig. 10. Empirical CDF and fitted distributions for cellular mobile communication bands (discrete-time models).

system. Since DECT employs a time-slotted frame structure, the discrete-time distributions considered in Section VIII-D were then fitted to the empirical data following a similar procedure for the discretization of the continuous-time empirical CDF. As an example of the results obtained in this case, Figure 11 shows the fitted distributions for a channel measured in the DECT band. As it can be appreciated, busy periods appear to be perfectly fitted with a geometric distribution. This result, however, should be interpreted carefully. The empirical CDF of Figure 11 indicates that the busy periods observed in the field measurements were always one time-slot long, and in

8The DECT radio interface [67] is based on a Multi Carrier, Time Division Multiple Access, Time Division Duplex (MC/TDMA/TDD) radio access methodology. The basic DECT frequency allocation defines 10 carrier frequencies. In the time domain, each carrier frequency is divided in 10-ms frames, which are composed of 24 time-slots of 417 µs each. The first (last) 12 time-slots of a frame are used for downlink (uplink) transmissions.
this particular case the geometric distribution perfectly fits the resulting simple CDF. Nevertheless, the suitability of the geometric distribution for longer busy periods of two or more time-slots cannot be concluded from the results of Figure 11. Unfortunately, this could not be verified with empirical data since busy periods were always one time-slot long for all the measured DECT channels. However, this circumstance suggests that this may be the most common case in practice. Concerning idle periods, the results of Figure 11 indicate that the channel inactivity periods may either be 11 or 23 time-slots long. None of the considered distributions can be fitted to the resulting empirical CDF with a reasonable level of accuracy.

Although the previous discussion indicates that a probabilistic modeling of busy and idle period lengths does not seem to be valid for DECT channels, the results of Figure 11 suggest an alternative, simpler modeling approach. In particular, two well-defined cases can be inferred from Figure 11. The first case corresponds to a base station transmitting broadcast in-coming call set-up. This information is transmitted in one (busy) time-slot. If there are no active communications, the rest of downlink and uplink time-slots in the frame will be empty until the next broadcast message (time-slot). Therefore, in this case the channel occupancy pattern is composed of one busy time-slot followed by 23 idle time-slots. According to Figure 11, this occupancy pattern was observed for about 90% of the time in the considered channel. The second inferrable case corresponds to a single communication link between the base station and one portable device. In this case, two time-slots are used for communication, one in the downlink part of the frame and the other one in the uplink part. In this other case, the channel occupancy pattern is composed of one busy time-slot followed by 11 idle time-slots. According to Figure 11, this occupancy pattern was observed for about 10% of the time in the considered channel.

Note that the two possible occupancy patterns inferred from the results of Figure 11 are characterized by a completely deterministic sequence of busy and idle periods, for which a probabilistic modeling may not be well suited. In the eventual case of busy periods of two or more time-slots (i.e., two or more communication links with the base station), the length of busy and idle periods would depend on the particular position of busy time-slots within the DECT frame structure. In this other case, a probabilistic modeling would be more appropriate. However, as mentioned above, busy periods of one time-slot appear to be the most common situation in real DECT systems and in such a case the deterministic modeling approach discussed above results more convenient.

Before concluding this section, it is worth noting that DECT makes use of a continuous dynamic channel selection and allocation mechanism. All DECT equipment is obliged to regularly scan its local radio environment at least once every 30 seconds. After that period of time, the system may decide to switch to a different operating carrier frequency. As a result, a particular carrier frequency may exhibit in practice long inactivity periods (while the system is operating over other carrier frequencies) followed by periods of activity (while the system is operating over that carrier frequency). During these activity periods, the carrier frequency exhibits busy/idle intervals at a much shorter timescale (i.e., at the slot level), which is due to the the two aforementioned occupancy patterns. This observation suggests that the real activity pattern of a channel may be more appropriately described by means of a two-layer approach: one modeling level describes channel usage at a high timescale, while another modeling level describes in detail the true occupancy pattern at a low timescale. This modeling approach is discussed in Section IX.

### IX. Two-Layer Modeling Approaches

This section explores two different approaches integrating the models of Sections VII and VIII to simultaneously reproduce the statistical properties of spectrum usage at long and short timescales.

The first proposed method comprises four distribution functions. As illustrated in Figure 12(a), two distribution functions are used to characterize the lengths of inactivity and activity periods at long timescales, $F^L(T_0)$ and $F^L(T_1)$ respectively. During the periods of primary activity, a sequence of idle and busy periods are present at a shorter timescale as described by the distribution functions $F^S(T_0)$ and $F^S(T_1)$. The functions $F^L(T_0)$ and $F^L(T_1)$ can be generalized Pareto distributions as concluded in Section VII, while the distribution functions

<table>
<thead>
<tr>
<th>Busy</th>
<th>Geometric</th>
<th>Poisson</th>
<th>Neg. binominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{KL}$</td>
<td>0.13</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>$D_{sym}$</td>
<td>0.64</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>$D_{KL}$</td>
<td>0.09</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Idle</th>
<th>Geometric</th>
<th>Poisson</th>
<th>Neg. binominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{KL}$</td>
<td>0.14</td>
<td>0.28</td>
<td>0.14</td>
</tr>
<tr>
<td>$D_{sym}$</td>
<td>0.36</td>
<td>1.49</td>
<td>0.29</td>
</tr>
<tr>
<td>$D_{KL}$</td>
<td>0.13</td>
<td>0.32</td>
<td>0.10</td>
</tr>
</tbody>
</table>

![Fig. 11. Empirical CDF and fitted distributions for cordless telephone bands (discrete-time models).](image)

<table>
<thead>
<tr>
<th>METRICS FOR CELLULAR MOBILE COMMUNICATION BANDS (DISCRETE-TIME MODELS).</th>
<th>Geometric</th>
<th>Poisson</th>
<th>Neg. binominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>$D_{KL}$</td>
<td>$D_{sym}$</td>
<td>$D_{KL}$</td>
</tr>
<tr>
<td>Busy</td>
<td>0.13</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Idle</td>
<td>0.14</td>
<td>0.28</td>
<td>0.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE X</th>
<th>11</th>
<th>42</th>
<th>54</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.13</td>
<td>0.27</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>0.09</td>
<td>0.29</td>
<td>0.14</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>$D_{KL}$</th>
<th>$D_{sym}$</th>
<th>$D_{KL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Busy</td>
<td>0.13</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Idle</td>
<td>0.14</td>
<td>0.28</td>
<td>0.14</td>
</tr>
</tbody>
</table>

![Fig. 11. Empirical CDF and fitted distributions for cordless telephone bands (discrete-time models).](image)
is appropriate for channels of amateur, paging, PMR/PAMR and cordless telephone bands. For channels of cellular mobile communication systems such as E-GSM 900 and DCS 1800, the existence of inactivity periods lasting for several seconds is unlikely. For this particular case, a second modeling approach is proposed. This alternative considers two distribution functions to describe the length of idle and busy periods at short timescales, \( F^S(T_0) \) and \( F^S(T_1) \) respectively, which can be negative binomial distributions as concluded from Section VIII-D. The behavior at long timescales is included by means of a DC model that describes the channel load variation over time as illustrated in Figure 12(b). The deterministic DC models proposed in [31] for cellular mobile communication systems can be employed to this end. Based on this approach, the parameters of the distribution functions \( F^S(T_0) \) and \( F^S(T_1) \) are regularly adjusted according to (5) so as to meet the corresponding DC at any time. Field measurements indicated that this alternative modeling approach results more appropriate in the case of GSM/DCS systems.

In summary, the models developed in Sections VII and VIII provide a simple yet realistic and accurate means to characterize the statistical properties of channel occupancy patterns at long and short timescales, and can be combined into a two-layer modeling approach in order to provide a holistic characterization of the spectrum usage patterns observed in real wireless systems.

X. Conclusion

This work has analyzed the spectrum occupancy patterns of various radio technologies in the time domain from the point of view of the DSA/CR technology. The deficiencies and limitations of previous works have been overcome by performing a comprehensive, systematical and rigorous study on the set of probability distributions that can be employed to accurately describe the lengths of busy and idle periods in real radio communication systems. The study has relied on field measurements performed with two sophisticated measurement platforms providing various levels of time resolution, which guarantees the realism and accuracy of the models. Numerical values for the models’ parameters, extracted from empirical data, have been provided in order to facilitate the practical application of the models.

The obtained results indicate that the assumption of exponentially distributed busy and idle periods is invalid, meaning that the CTMC model widely employed in the literature is unrealistic. In real systems, other distributions are observed. At long timescales, a single distribution function (generalized Pareto) has been proven to adequately describe the state holding times for all the considered bands. At short timescales, however, the obtained results indicate that the most convenient distribution depends on the considered radio technology. For time-slotted systems, channel occupancy patterns can also be described from a discrete-time viewpoint where state holding times are expressed as an integer number of time-slots, following a negative binomial distribution. While a probabilistic approach has been proven to be adequate for most radio technologies, it may not be appropriate in some particular

---

\( F^S(T_0) \) and \( F^S(T_1) \) depend on the particular radio technology considered, as observed in the results of Section VIII.

The previous modeling approach is motivated by the following observation. Low time resolution measurement devices, such as spectrum analyzers, constitute a reasonable choice to determine the length of long inactivity periods. Although the real state of a channel might change between consecutive channel observations without being detected by the spectrum analyzer, this situation can be considered to be rather unlikely in most cases as the effective sampling period is not excessively long, and activity periods are not short enough to go unnoticed between two consecutive channel observations. The simultaneous monitoring of RF channels with both measurement platforms confirmed this statement. Therefore, the length of the idle periods reported by the spectrum analyzer can be considered as an acceptable estimation of the channel inactivity periods \( F^S(T_0) \). Spectrum analyzer measurements can also be employed to determine when a primary transmitter is active and therefore the length of its activity periods \( F^S(T_1) \), although in this case the length of the real busy and idle periods at short timescales cannot be determined due to the limited time resolution.\(^9\) High time resolution measurements can then be employed to extract the real channel occupancy pattern in terms of busy and idle periods when a primary transmitter is active, i.e. \( F^S(T_0) \) and \( F^S(T_1) \). Based on this discussion, the models derived from low and high time-resolution measurements can be combined as mentioned above in order to jointly describe the channel occupancy patterns at long and short timescales.

The spectrum usage patterns observed for various radio technologies indicated that the previous modeling approach

\(^9\)The employed spectrum analyzer, with the selected configuration, sweeps at an approximate rate of 25 ms per megahertz of scanned bandwidth. This means, for instance, that a 200-kHz GSM/DCS channel is averaged for a time period of 5 ms (i.e., more than one GSM/DCS frame) and a 1.728-MHz DECT channel is averaged for several DECT frames. If a single or a few time-slots are busy within the frame with a sufficiently high power level, the spectrum analyzer will report the carrier frequency as busy. However, the exact time-slot(s) that are active cannot be determined.
cases where the channel occupancy is characterized by strong deterministic patterns as it has been observed in this study for the DECT system. In such cases, other alternative modeling approaches taking into account technology-specific features at the physical and higher layers may result more convenient. Finally, a two-layer modeling approach combining models at long and short timescales has been proposed as a holistic means to describe the spectrum occupancy patterns observed in real radio communication systems.

REFERENCES


Miguel López-Benítez (S’08, M’12) received the B.Sc. (2003) and M.Sc. (2006) degrees in Communications Engineering (First-Class Honors) from Miguel Hernández University (UMH), Elche, Spain, and a Ph.D. degree (2011) in Communications Engineering from the Department of Signal Theory and Communications (TSC) of the Technical University of Catalonia (UPC), Barcelona, Spain.

He has been a Research Assistant in the Ubiquitous Wireless Communications Research (Uwicore) laboratory at UMH, and a Research and Teaching Assistant in the Mobile Communication Research Group (GRCM) at UPC. He currently is a Research Fellow in the Centre for Communication Systems Research (CCSR) at the University of Surrey, Guildford, Surrey, UK. His research interests include the field of mobile radio communication systems, with a special emphasis on radio resource management, heterogeneous wireless systems, quality of service provisioning, spectrum modeling and opportunistic/dynamic spectrum access in cognitive radio networks. He has been or is actively involved in the European-funded projects AROMA, NEWCOM++, FARAMIR, QoS4MOS and CoRadat along with Spanish projects COGNOS and ARCO. He has co-authored 2 book chapters and more than 40 papers in refereed journals and recognized conferences, and has been TPC for several IEEE conferences as well as reviewer for IEEE journals and conferences.

Dr. López-Benítez was the recipient of the 2003 and 2006 University Education National Awards, a distinction from the Spanish Ministry of Education and Science to the best national academic records, as well as some other distinctions from the Spanish professional association of telecommunications engineers. His M.Sc. thesis was awarded a national research prize from the France Telecom foundation (today Orange foundation) in the context of the 5th Archimedes University Competition, the main science contest for young researchers in Spain organized by the Spanish Ministry of Education and Science (+100 applications). He was shortlisted as a finalist for the IET Innovation Awards 2012 in the category of Telecommunications (+400 applications). For more details, please visit http://www.lopezbenitez.es.

Fernando Casadevall (M’87) received the Engineer (1977) and Doctor Engineer (1983) degrees in telecommunications engineering from Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. In 1978, he joined UPC, where he was an Associate Professor from 1983 to 1991. He is currently a Full Professor with the Department of Signal Theory and Communications, UPC. After graduating, he was concerned with equalization techniques for digital fiber-optic systems. He has also been working in the field of digital communications, with particular emphasis on digital radio and its performance under multipath propagation conditions. Over the last 15 years, he has been mainly concerned with the performance analysis and development of digital mobile radio systems. He has published around 150 technical papers in both international conferences and magazines, most of them corresponding to IEEE publications. His particular research interests include cellular and personal communication systems, multipath transceiver design, radio resource management, and end-to-end quality-of-service issues. During the last 15 years, he has participated in more than 30 research projects funded by both public and private organizations. In particular, he has actively participated in 15 research projects funded by the European Commission, being the Project Manager for three of them: Advanced Radio Resource Management for Wireless Systems (ARROWS), Evolutionary Strategies for Radio Resource Management in Cellular Heterogeneous Networks (EVEREST), and Advanced Resource Management Solutions for Future All IP Heterogeneous Mobile Radio Environments (AROMA) (see http://www.gct.tsc.upc.edu for details).

F. Casadevall has been a Technical Program Committee Member for different international IEEE supported conferences and a Reviewer for several IEEE magazines. From October 1992 to January 1996, he was in charge of the Information Technology Area, National Agency for Evaluation and Forecasting (Spanish National Research Council).