A Survey of Measurement-based Spectrum Occupancy Modelling for Cognitive Radios

Yunfei Chen, *Senior Member, IEEE,* Hee-Seok Oh

Corresponding Address:

Yunfei Chen

School of Engineering

University of Warwick, Coventry, U.K. CV4 7AL

Tel: +44 (0)24 765 23105

e-mail: Yunfei.Chen@warwick.ac.uk

Yunfei Chen is with the School of Engineering, University of Warwick, Coventry, U.K. CV4 7AL (e-mail: Yunfei.Chen@warwick.ac.uk)

Hee-Seok Oh is with the Department of Statistics, Seoul National University, Seoul, Korea 151-747 (e-mail: heeseok.oh@gmail.com)

Abstract

Spectrum occupancy models are very useful in cognitive radio designs. They can be used to increase spectrum sensing accuracy for more reliable operation, to remove spectrum sensing for higher resource usage efficiency or to select channels for better opportunistic access, among other applications. In this survey, various spectrum occupancy models from measurement campaigns taken around the world are investigated. These models extract different statistical properties of the spectrum occupancy from the measured data. In addition to these models, spectrum occupancy prediction is also discussed, where the auto-regressive and/or moving-average models are used to predict the channel status at future time instants. After comparing these different methods and models, several challenges are also summarized based on this survey.

Index Terms

Cognitive radio, measurement, modelling, prediction, spectrum occupancy.

I. Introduction

Cognitive radio (CR) equips a radio device with cognition by learning from and adapting to the radio environment during the operation [1]. More specifically, CR finds the parts of the radio spectrum that are not being occupied at some specific times in some specific locations and move its operation to these parts called ”spectrum holes” for opportunistic access. Thus, CR has two main functions: spectrum sensing and data transmission. Among them, spectrum sensing is probably more important than data transmission in many cases, as it determines the amount of interference to other systems, including possible licensed systems, which is exactly the current fixed spectrum access policy tries to avoid and is the main concern of the regulators. Consequently, it is of paramount importance to obtain spectrum sensing results as accurately as possible. The accuracy of spectrum sensing depends on several factors but ultimately it depends on the occupancy status of the spectrum, as the radio spectrum is a dynamic environment determined by the usage of the spectrum. Therefore, knowledge of statistics or models of the spectrum occupancy will greatly benefit and improve the designs of CR, and indeed such knowledge has already been used to improve CR performances.

To name a few, in [2], the authors used the Markov chain model of the spectrum occupancy in the Bayesian cost factors of missed detection and false alarm to enhance the accuracy of spectrum sensing. The proposed weighted sequence detection algorithm is optimal in minimizing the overall detection error. In [3], the authors proposed proactive opportunistic access scheme to remove spectrum sensing so that cognitive data transmission is not frequently interrupted by spectrum sensing. In this case, the spectrum availability is predicted by using the spectrum occupancy model based on all historical channel information, not by spectrum sensing that is based on detection using only a snapshot of channel measurements. In [4], the authors used the spectrum occupancy information in cognitive radios to select the best channels for control and data transmission purposes. It was shown there that the overall time required to switch cognitive radio from one channel to the other channel due to collision can be reduced by up to 55% such that the throughput of the system has been greatly improved. In [5] and [6], the statistics of the spectrum occupancy were used to control the transmission power of the cognitive radio to maximize the bandwidth efficiency. Specifically, the average transmission rate is maximized subject to a total average power constrain and the optimal transmission power becomes a function of the statistics of the spectrum occupancy. Similar works have also been conducted in [7], where more practical spectrum occupancy models were used. In [8], the authors used the spectrum occupancy information to predict the channel status in the following time slots so that optimal spectrum sensing order can be achieved by comparing them. The proposed selective scheme can improve the throughput of the system while meeting quality of service requirements. In [9] and [10], this spectrum occupancy information was used to achieve trade-off between data buffering and channel switching that can save up to 50% of the energy consumption, and to reduce the required number of spectrum handoffs considerably, respectively. These works and other works use statistics or models of the spectrum occupancy to improve either the physical layer spectrum sensing or the upper layer spectrum management for cognitive radios. Thus, spectrum occupancy models are very important for cognitive radio designs.

This paper conducts a survey of the state-of-the-art spectrum occupancy models that are obtained from measurement campaigns up to 2013 used for cognitive radio designs. To start with, various measurement campaigns around the world that extract important sample statistics, mainly the duty cycle, are discussed. Then, more in-depth works that extract complete statistical models out of the measurement data, including cumulative distribution function, probability density function, Markov chain and linear regression models, are investigated. These models can also be classified as time-dimension models, frequency-dimension models, location-dimension models or their mixtures in terms of the dimension of spectrum occupancy model. As the last part of this survey, various models for spectrum occupancy prediction are presented that predict the value of the spectrum occupancy in the future. Below is an outline of this paper.

- Spectrum modelling
  - Statistics
    - Power: [11] − [17], [20], [25], [27]
    - Occupancy: [12] − [15], [17] − [19], [21] − [24], [26], [28], [29]
    - Duty cycle: [12] − [19], [28]
  - CDF, PDF
    - Power: [32] − [37]
    - Duty cycle: [34], [38]
    - Holding time: [35], [39]
    - Available channels: [40] − [42]
  - Markov chain:
    - CTMC: [43] − [46]
    - CTSMC: [47] − [50]
    - DTMC and heuristic: [51] − [53]
  - Linear regression
    - Laycock-Gott: [54] − [57]
    - Space: [58] − [61]
- Spectrum prediction
  - ARMA: [62]
  - AR: [63] − [65]

II. Spectrum Occupancy Statistics

There have been a large number of measurement campaigns around the world to study the spectrum utilization. Some of these measurement campaigns give detailed data analysis in terms of complete statistical models, such as probability functions and random processes. Other measurement campaigns only give simple but important sample statistics. We start with the measurement campaigns that give the sample statistics. Before proceeding to discuss some of the important works, it is useful to make a few summary statements regarding these campaigns.

First, most of these campaigns have been conducted by using an antenna that collects data, a spectrum analyser that displays and processes data and a computer that analyses data. Other minor equipment for calibration and pre-selection are also used in some campaigns for better results. Different choices of equipment will of course affect the accuracy of the measurement. However, as long as the equipment works reasonably well, the difference in equipment has marginal effect on the statistical behaviours of spectrum occupancy and therefore, the equipment difference is not discussed in this paper.

Second, all these campaigns consider either outdoor or indoor locations. Overwhelmingly, the outdoor locations are chosen on the roof of a high building that overlooks a certain area in order to reduce the effect of radio propagation loss on the measurements. The indoor locations are often chosen in an office building, a typical application environment for wireless communications. Some of the works tried to analyse the location variation of spectrum occupancy by taking measurements from different locations but this is generally expensive and difficult. As a result, most works focus on time- and frequency-variations by fixing locations. The time span of the measurement varies from a few days in some campaigns to a few years in other campaigns. Intuitively, the longer the time span is, the more useful but also the more expensive the measurement campaign will be. For short-term effects they are of similar value. Also, most of these campaigns focus on the frequency range between 30 MHz and 3 GHz. This range covers some of the very important applications of wireless communications, such as FM radio, TV broadcasting, cellular communications. Thus, measurements of frequency bands in this range will provide useful insights on the current status of spectrum utilization. Another reason for this choice is that higher frequencies are more vulnerable to propagation loss and thus, the measurements will be less accurate as an indicator of the channel occupation status when the frequency is higher.

Third, these works only consider some simple statistics of the spectrum occupancy, such as the maximum, the average and the minimum of the power level, the spectrum occupancy and the duty cycle. More complicated statistical models, such as cumulative distribution function and Markov model, will be discussed in the next section.

One of the earliest measurement campaigns is perhaps done by Sanders for the Institute of Telecommunication Science in the USA [11]. In this work, using a radio spectrum measurement system with custom-made hardware and software, Sanders measured the frequency bands from 108 MHz to 19.3 GHz in three different US cities, Denver, San Diego and Los Angeles, for two weeks. All the measurements were performed outdoors. Using these measurements, it was found that San Diego has considerably more radio activities than Denver. Thus, Sanders concluded that coastal cities have higher spectrum occupancy than midwestern cities, due to the presence of maritime radars. By analysing the data from Denver, Sanders also found that radar bands have severe adjacent interferences and thus, the necessity for guard bands. The spectrum occupancy from microwave ovens is also evident in the 2400 - 2500 MHz ISM band.

More extensive measurement campaign was done by McHenry and his colleagues in [12]. In fact, some of these measurements are publicly available in [13]. In [12], measurements were taken from 30 MHz to 3 GHz for a few hours. The location was fixed to the roof of a high-rise building in the centre of Chicago. Pre-selector was used to improve measurement sensitivity and dynamic range and the data were calibrated to measure the power level at the antenna input, in contrast to other campaigns that measure the spectrum analyser input. The results were presented in terms of the maximum power level, the instantaneous spectrum occupancy, and the duty cycle. From the results, some TV bands have the heaviest occupancy with an average duty cycle of 70.9%. The cellular band also has a heavy occupancy with an average duty cycle of 55%. On the other hand, some frequency bands allocated by FCC, including satellite bands, were completely unused. An interesting result from this study is that by comparing Chicago and New York, the two cities have sizable differences in terms of spectrum occupancy, although both are large cities in the USA. A related measurement campaign in the same city Chicago has drawn similar conclusions [14] and [15].

In [16], a similar measurement campaign was done by Wellens and his colleagues in Germany. The measurements were taken on the roof of a building that overlooks the area and a room in an office building. This is also one of the few works that consider indoor environments. The frequency range is from 20 MHz to 6 GHz, measured for seven days. The resolution bandwidth is 200 kHz that evenly divides the whole frequency range. The results were presented in terms of the power levels, and the duty cycle for different locations and different frequencies. For the outdoor location, it was found that the spectrum occupancy is almost 100% from 20 MHz to 3 GHz, and very low from 3 GHz to 6 GHz. For the indoor location, the spectrum occupancy is about 32% from 20 MHz to 3 GHz. Thus, the spectrum occupancy is highly related to the application scenario, an intuition that is confirmed by [16].

In [17], another measurement campaign was performed in Spain. This was also done for an outdoor application but in the frequency range between 75 MHz and 3 GHz for two days. The equipment set-up is similar to [16] without any pre-selection or calibration. The resolution is 10 kHz that evenly divides different blocks of frequencies. The results were presented in terms of the power levels, spectrum occupancy and duty cycle. The statistics focus on the cellular bands, showing significant spectrum opportunities or low spectrum occupancy in frequencies above 1 GHz. An interesting result from this work is that the spectrum occupancy is also related to the frequency bin, that is, the frequency separation between two frequencies measured. This shows the complexity and difficulty of the measurement campaign.

In [18], a measurement campaign was done in Singapore for the frequency range from 80

MHz to 5.85 GHz with a resolution bandwidth of 10 kHz. The measurements were taken on the roof of a building for 12 weekdays. Duty cycle and spectrum occupancy results show that the spectrum occupancy in Singapore is as low as 4.54% in terms of used bandwidth. The busiest band in Singapore is the GSM900 band, although there are plenty of spectrum opportunities or low spectrum occupancy in the radar bands, ISM band, and above 1 GHz. This measurement campaign also calculates the received signal at the antenna input but without pre-selection. The uniqueness about this campaign is that Singapore is a relatively small country whose radio activities might be mixed with those from neighbouring countries.

There are other works that extract important sample statistics out of the measurement campaigns. For example, in [19], spectrum occupancy in three different locations in two different countries was studied, based on which spectrum occupancy and duty cycle were calculated. Although this is a good attempt in understanding the spectrum usage difference between different countries, it is in general difficult to obtain such understanding. In [20], another campaign in Atlanta, USA, was performed where the power spectral density results were given. In [21], [22], [23] and [24], some specialized systems, such as Wi-Fi, GSM and public safety systems, were studied, where measurements were taken to show that the Wi-Fi band was still relatively vacant while the public safety band was relatively full, at the time of study. A more important conclusion from [23] is that there are seasonal, weekly and daily trends in the spectrum occupancy statistics. Also, reference [24] has used the measurement results to study the design of spectrum sensing schemes for better performances. In [25], the measurement campaign conducted in Qatar was presented. The frequency bands from 700 MHz to 3 GHz were measured for seven days with a resolution of 300 kHz in four outdoor locations. Different from most previous works, this work measured the four locations at the same time and thus, location analysis in this work is more convincing, as it has been revealed in the previous works that spectrum occupancy changes with time so sequential measuring in different locations would lead to a mixed effect of time and location. It was found in this work that the spectrum utilization is highly related to the natural environment where the device is deployed. In [26], two outdoor locations and one indoor location in Japan were studied, focussing on the TV bands, for 24 hours. The results in terms of duty cycle, spectrum occupancy rate and amplitude probability distribution revealed that there is higher spectrum occupancy in outdoor locations than indoor location. The work has also explicitly compared the spectrum occupancy difference between night time and day time and has determined a safe distance from the primary user for the operation of secondary user as 0.68 km in urban areas and 1.7 km in suburban areas. In [27], a measurement campaign in Amsterdam was performed, where, unlike the other works that used measurement equipment at a fixed location, mobile equipment was also used. Based on this campaign, the effect of location variation was studied. The effect of different areas was also studied. This is an interesting work that extends previous works on a ”point” to an area. Such a mobile monitoring system also takes communications that only happen locally into account. This may be important, as wireless relaying also happens locally between peers that may be difficult to be captured by settings used in other measurement campaigns. In [28] and [29], long-term measurement campaigns were conducted. While most previous works take measurements for a few days, [28] took measurements for six months while [29] took measurements for three years. This will allow the extraction of long-term statistics from the measurements. Indeed, reference [28] revealed the possibility of using the difference between weekdays and weekends to explore spectrum opportunities and determined a set of ”suitable” channels above 1 GHz, while reference [29] examined the seasonal, weekly and daily trends in different frequency bands.

An important method that is used in almost all these works is energy detection, where the measurement is compared with a predetermined threshold. The channel is considered vacant if the measurement is below this threshold or occupied if the measurement is above this threshold. Thus, assuming that the i-th measurement at frequency f is $P_i(f)$ and the threshold is T, one has the instantaneous spectrum occupancy rate for the i-th measurement at frequency f as

$$B_i(f) = 1, \quad if \quad P_i(f) > T$$
$$B_i(f) = 0, \quad if \quad P_i(f) < T. \tag{1}$$

where $i = 1, 2, \cdots, N(f)$ and $N(f)$ is the total number of measurements taken at frequency f. Using (1), the average duty cycle used in the measurement campaigns is calculated as

$$r(f) = \frac{\sum_{i=1}^{N(f)} B_i(f)}{N(f)} \tag{2}$$

which is a function of frequency. From (2), the duty cycle is always smaller than 1. The larger the value of the duty cycle is, the higher the spectrum occupancy will be.

The challenging part of the above calculation is the setting of the threshold T. In fact, the spectrum occupancy could be significantly changed when the threshold varies [29]. This is expected, as if the threshold is too low, there will be more false alarms while if the threshold is too high, there will be more missed-detections. A natural choice of the threshold is the noise floor. As mentioned before, each measurement campaign has a slightly different system setting. Thus, for a specific system, the noise floor can be obtained by replacing the antenna in the measuring system with a 50 ohms load and taking measurements for this load. These noise samples are then averaged to find the noise power or the noise floor. They are also used to find the distribution of the noise. It turns out that the noise floor increases with the frequency [26]. Also, in some cases, the noise is not Gaussian [28]. Finally, instead of using the calculated noise floor, the detection threshold in (1) is often set a few dBs above the noise floor such that the probability of false alarm satisfies a certain value to take into account strong noise samples that are comparable to signals. Consequently, the detection threshold in (1) equals

$$T = W(f) + M(f) \tag{3}$$

where $W(f)$ is the calculated noise floor at frequency f and $M(f)$ is a fixed margin to satisfy certain probability of false alarm criterion. From (3), the detection threshold should be a function of frequency too.

In [16] and [17], $M(f) = 3\ dB$ was chosen to satisfy a probability of false alarm of 0.01 such that the detection threshold varies with frequency. In [26], a probability of false alarm of 0.015 was chosen such that both $M(f)$ and the detection threshold vary with frequency. In [12], a fixed threshold of -90 dBm or -110 dBm was chosen for different bands. In [25], a fixed threshold of -78 dBm was used. In [28], a fixed margin of 5 dB was used while the detection threshold varies with frequency between -77.4 dBm and 69.5 dBm. In [19], a fixed margin of 7 dB was used while the detection threshold varies with frequency. In [18], the detection threshold was set 6 dB above the minimum power level, not the noise floor as in (3). In [22], the noise samples were assumed Gaussian such that the margin $M(f)$ was chosen as a function of the standard deviation to achieve a probability of false alarm of 0.003. There are also other ways of setting the detection threshold, borrowed from other research areas, such as Otsu's algorithm, recursive and adaptive thresholding [30]. In fact, reference [30] has a very detailed discussion of thresholding as well as other data processing techniques used for the calculation of spectrum occupancy. In addition to the widely used energy detection, other detection methods are also available [31].

Tables I and II give a summary of the campaigns. All the afore-mentioned works only obtain statistics, mainly the average duty cycle, to show the spectrum occupancy. The main conclusions from these works are that there are more spectrum opportunities above 1 GHz than below 1 GHz and there are more spectrum opportunities indoor than outdoor. The actual spectrum occupancy varies with frequency, time, location, detection threshold and system setting. These initial results provide important guidance for further studies. The duty cycle statistic is effective in giving a general idea of the availability of different frequency bands. It can also be used in cognitive radio designs as *a priori* knowledge of the channel status. However, for more sophisticated applications, more details about the spectrum occupancy are required. In the next section, complete statistical models for the spectrum occupancy are surveyed. These models are more useful for dynamic access and control of opportunistic spectrum.

TABLE I

SUMMARY OF THE MEASUREMENT CAMPAIGN RESULTS. (POWER (P), OCCUPANCY (O), DUTY CYCLE (D))

Campaign	Frequency	Time	Location	Statistics
[11]	108 MHz - 19.3 GHz	two weeks	outdoors	P
[12] - [15]	30 MHz - 3 GHz	hours	outdoors	P, O, D
[16]	20 MHz - 6 GHz	seven days	outdoors, indoors	P, D
[17]	75 MHz - 3 GHz	two days	outdoors	P, O, D
[18]	80 MHz - 5.85 GHz	twelve days	outdoors	O, D
[19]	400 MHz - 3 GHz	six days	outdoors	O, D
[20]	400 MHz - 7.2 GHz	several months	outdoors	P
[21]	Wi-Fi band	seven days	outdoors	O
[22]	GSM band	N/A	outdoors	O
[23]	public safety band	several months	outdoors	O
[24]	public safety band	two days	outdoors	O
[25]	700 MHz - 3 GHz	three days	outdoors	P
[26]	90 MHz - 3 GHz	one day	outdoors, indoors	O, D
[27]	100 MHz - 500 Mhz	one day	outdoors	P
[28]	300 MHz - 4.9 GHz	six months	outdoors	O, D
[29]	30 MHz - 6 GHz	three years	outdoors	O

III. SPECTRUM OCCUPANCY MODELS

In this section, complete statistical models for spectrum occupancy are surveyed. Note that the measurement system settings in these works are similar to those discussed in the previous section. In fact, some of the works discussed in this section are from the same measurement campaigns as those in the previous section, such as the Aachen measurement campaign, only with more comprehensive analysis. Thus, unless necessary, the measurement settings will not be

TABLE II

SUMMARY OF THE MEASUREMENT CAMPAIGNS' MAIN FINDINGS.

Campaign	Findings
[11]	Coastal cities have higher occupancy. Radar bands have severe adjacent interferences.
[12] - [15]	TV and celluar bands are heavily occupied. Satellite bands are free.
[16]	Occupancy is lower for above 3 GHz and indoor locations.
[17]	Occupancy is lower for above 1 GHz and depends on frequency bin.
[18]	Occupancy is low in Singapore in most bands. GSM band is the busiest.
[19]	Occupancy depends on culture and economical development level.
[20]	5.6 GHz bands in urban area are vacant and 6.6 GHz band in rural area are vacant.
[21]	Wi-Fi band is relatively vacant.
[22]	Occupancy depends on load scenarios and traffic channels but is generally low.
[23]	Public safety band is relatively full with seasonal, weekly and daily trends.
[24]	Public safety band is relatively full.
[25]	Occupancy is highly related to the natural environment.
[26]	Occupancy is lower for indoor. Night time and day time are different.
[27]	Mobile spectrum monitoring is necessary to reveal location-specific information.
[28]	Difference between weekdays and weekends can be used to enhance utilization.
[29]	Seasonal, weekly and daily trends exist in different frequency bands.

discussed in the following but bear in mind that these works may also have to select the times, frequencies, locations, and detection thresholds measured and these selections may affect the results and therefore, the conclusions. We start with the probability function models, including the probability density function (PDF) and cumulative distribution function (CDF).

A. CDF and PDF

The motivation of CDF and PDF modelling is two-fold. First, these models describe the range of possible values for the primary user signal and how often these values occur. This is useful for the choices of cognitive transceiver parameters, such as dynamic range and transmission period. Second, these models can be used to improve cognitive performances. For example, the PDF of the power level may be used as prior knowledge to improve primary user signal detection and estimation. These works can be categorized into two types: some works model the CDF and/or PDF of the power level that are obtained directly from the measurement systems before energy detection, while others model the CDF and/or PDF of the duty cycle and associated random variables that are calculated from the instantaneous spectrum occupancy rate after energy detection.

In [32] - [37], the probability models for the power level have been studied. Specifically, in [32], another measurement campaign was performed in New Zealand that covers the frequencies from 806 MHz to 2.75 GHz for 12 weeks. One outdoor location on the roof of a tall building and one indoor location were studied. From the measurements, the amplitude probability distribution, where the probability that a certain power level occurs throughout the campaign is defined as a function of frequency and power level, was calculated and plotted with x axis being the frequency, y axis being the power level and z axis being the probability of occurrence. Since the measured range covers several wireless systems with possibly different occupancy statuses, it is necessary to differentiate the amplitude probabilities over different frequencies. In the calculation, the probability is averaged over all sampling times at a fixed frequency. Mathematically, denote P_{ij} as the measurement taken at the i-th sampling time t_i in the j-th frequency f_j, where $i = 1, 2, \cdots, N_I$ and $j = 1, 2, \cdots, N_J$, and P_T as a certain threshold with $\min\{P_{ij}\} < P_T < \max\{P_{ij}\}$. Then, the amplitude probability distribution is calculated as

$$APD(P_T, f_j) = Pr[P_{ij} > P_T] = \frac{1}{N_I}\sum_{i=1}^{N_I} S_{ij} \tag{4}$$

where

$$S_{ij} = 1, \quad if \quad P_{ij} > P_T$$
$$S_{ij} = 0, \quad if \quad P_{ij} < P_T. \tag{5}$$

In this case, the amplitude probability distribution is a function of threshold P_T and frequency f_j. Using the amplitude probability distribution, the spectrum was divided into three types: white space, grey space and black space. The black space has more than 70% chance that the power level is above certain threshold, not suitable for any exploration, the grey space has a chance between 10% and 30% that the power level is above certain threshold, possible for opportunistic access, while the white space has a chance between 5% and 7% that the power level is above certain threshold, ideal for exploration. Also, indoor location and outdoor location do not have much difference in this study. Similar amplitude probability distributions were also obtained in [16] and [26] from different campaigns. To have a better view of the overall distribution of the power level, in [33], the amplitude probability distribution in [32] was further averaged over different frequencies and defined as spectrum opportunity, that is, the probability that a certain power level occurs is averaged over both all frequencies and all times. This gives

$$APD(P_T) = Pr[P_{ij} > P_T] = \frac{1}{N_I N_J} \sum_{i=1}^{N_I} \sum_{j=1}^{N_J} S_{ij}. \tag{6}$$

The effects of different system settings on the CDF of the power level were examined. The Beta distribution was then used to fit the empirical CDFs. In [34], the PDF of the power level was calculated by calculating the probability over the measurements in all frequencies and in a 12-hour period to account for the fact that day time and night time have different patterns. The PDF is asymmetric and it always rises quickly and then drops slowly when the power level increases, that is, there are more weak noises than strong signals.

On the other hand, for narrow-band measurements, the probability of the power level only needs to be averaged over all sampling times. In [35], the PDF of the power level over several selected channels of GSM system was obtained by calculating the probability over all measurements obtained in one day in that channel. The PDF is in general asymmetric. The calculated PDF for the GSM900 channel rises slowly then drops quickly when the power level increases, while the calculated PDF for the GSM1800 channel rises quickly and then drops slowly when the power level increases. In [22], another measurement campaign for the GSM system was performed for different traffic channels. The obtained PDF showed similar asymmetry to [34], but interestingly, some traffic channels in some scenarios show double-peaked PDFs. In [36], the VHF band was measured and studied. The obtained CDF has confirmed the spectrum sparsity in the VHF band. There is little persistent activity above -87 dBm in the urban area and above -110 dBm in the rural area. The best candidate channels for CR operations were also identified based on the CDF. In [37], another measurement campaign for the GSM system in China was conducted. The empirical PDF was first obtained from the measurements. Again, these PDFs show asymmetry near zero, implying that there are more weak noises in the measurements than signals. The characteristic function was then fitted using Nakagami-m distribution for both real and imaginary parts. The fitting works better for the real part than for the imaginary part. Unlike the other works that merely give the empirical PDF or CDF, this is perhaps the only work that tries to fit the power level to some known random variable, although the fitting still needs improvement.

To summarize, the current works on the probability models for the power level are mainly empirical, with the exception in [37]. Most of them focus on a specific band, such as the GSM band and the VHF band, while only a few, such as [34], [26] and [33], consider all the frequencies studied in the previous section. The PDF and CDF reflect the spectrum occupancy status by using the signal strength directly. Thus, unlike the duty cycle measure, it does not suffer from the noise uncertainty caused by the setting of the threshold in energy detection. Thus, the PDF and CDF measures are less affected by the measuring system. On the other hand, as can be seen, the PDF and CDF still heavily rely on the times, frequencies and locations used in the calculation. For the same measurements, calculations over one day or one week may lead to totally different conclusions. Thus, it is advisable to try different methods of calculation in order to find the most reliable results. Next, we discuss probability models for the duty cycle and its associated metrics for spectrum occupancy.

In the seminal paper [34], the empirical CDF of the duty cycle was obtained. These CDFs are significantly different for different locations and frequency bands. Also, the slopes for very low duty cycle near 0 and very high duty cycle near 1 are very large. Based on this observation, a modified Beta distribution for the duty cycle was proposed, which leads to an ordinary Beta function combined with two delta functions at 0 and 1, respectively, as

$$f(x) = P_0\delta(x) + P_1\delta(x-1) + \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \tag{7}$$

where $B(\cdot,\cdot)$ is the Beta function. The proposed modified Beta distribution fits the measurements quite well for different locations and frequency bands. The effect of detection threshold was also examined. It was found that the detection threshold will change the modified Beta distribution parameters but has insignificant impact on the distribution. Thus, the modified Beta distribution is a very useful model for duty cycles in different locations, frequencies or using different detection thresholds. The frequency correlation of the duty cycle was studied as well, which shows that adjacent channels have high correlation in duty cycle within 1.7 MHz for DECT systems and 5 MHz for UMTS systems. In another work [38], based on the measurements taken from the UHF and GSM bands in a fixed location for two days, the empirical CDF of the duty cycle was fitted with several widely used distributions and it was found that the lognormal distribution and the Beta distribution are the best candidates in this case in terms of Kolmogorov-Smirnov distance.

The duty cycles used in [34] and [38] are calculated as the average fraction of time at each frequency when the channel is occupied. In some applications, it is also important to know how long the channel will stay occupied or vacant, in addition to the average fraction. Essentially, they are the actual channel holding times. For example, in CR transmission, the channel holding time can be used to determine spectrum sensing interval and data transmission interval for sensing-throughput trade-off. In [39], these intervals, defined as run length and burst length, were studied. The complementary CDFs of the run length and the burst length were fitted with different distributions, in addition to other results in [39]. It was found that long-tailed distributions, in this case log-normal distribution, can fit the run length and burst length quite well in many cases. On the other hand, it was also pointed out that such conclusions heavily rely on the frequencies and times chosen. In [35], the time interval between two opportunities was studied for the GSM system. This is actually the burst length considered in [39]. In this case, it was found that this time interval can be well approximated as an exponential random variable, with the exception of the first sample point. This verifies the observation from [39] that the PDF of the run length and burst length depends on the frequency and time considered.

In [40] - [42], another important metric related to the duty cycle was studied as the number of free channels or spectrum availability. Due to the analytical difficulty, the spectrum availability was derived by proposing several approximations and calculating the approximate distribution parameters from real measurements. It was found that the Poisson-Normal approximation is accurate in terms of Chi-square test and the Camp-Paulson approximation is accurate in terms of maximum absolute error, to model the distribution of the number of free channels. There is no comparison between these two approximations using the same criterion though. The CDF of the Camp-Paulson approximation is given by [42, eq. (11)]

$$F_K(k) = \Phi\left(\frac{\Omega - \mu}{\sigma}\right) \tag{8}$$

where $\Omega = (1-\beta)\rho^{\frac{1}{3}}$, $\mu = 1-\alpha$, $\sigma = \sqrt{\beta\rho^{\frac{2}{3}} + \alpha}$, $\alpha = \frac{1}{9N-9k}$, $\beta = \frac{1}{9k+9}$, $\rho = \frac{N(k+1)(1-E\{K\}/N)}{(N-k)E\{K\}}$ and N is the total number of channels. The Poisson-Normal approximation is obtained by finding the CDF of the sum of a normal random variable and two Poisson random variables.

In the above, the probability models for the duty cycle and the related channel holding time and spectrum availability have been proposed. For the duty cycle, the modified Beta distribution is a good model, for the channel holding time, the log-normal distribution fits the purpose well, while for the spectrum availability, the Poisson-Normal distribution could be a good choice.

Similar to the statistics in the previous section, since this modelling is based on the spectrum occupancy rate after energy detection, the accuracies of these models are related to the detection threshold. The relationship between the statistics studied in Section II and the probability models studied in this subsection is that the statistics in Section II can be considered as statistical averages of the random variables for which the probability models in this subsection are established. In other words, results in Section II are the first-order statistics of the random variables studied in this subsection. From these results, it has been clearly shown that the spectrum occupancy is time-varying. Thus, a random variable may not be enough to model it. In the following, a more complete statistical model, random process, is used to model the spectrum occupancy. More specifically, the overwhelmingly used random process in this case is the Markov chain.

B. Markov Chain

The Markov chain is a very natural choice for statistical modelling of the spectrum occupancy, as the spectrum occupancy rate is either 0 or 1 after energy detection and the occupancy status changes between these two cases. Early works found that the continuous-time Markov chain (CTMC) models the spectrum occupancy well. With more measurement campaigns performed for larger numbers of frequency bands, recent works also showed that the continuous-time semi-Markov chain (CTSMC) models the spectrum occupancy better. In a Markov chain, the random process switches between different states and these are characterized by the transition probabilities. In each state, the random process is characterized by the sojourn time or channel holding time. In a CTMC, this holding time follows an exponential distribution, while in CTSMC, this holding time follows an arbitrary distribution.

The CTMC model is widely used in the modelling of HF bands that dates back to the 1970's. Reference [43] is perhaps one of the first works that proposed the use of a first-order CTMC to model the spectrum occupancy. By defining the spectrum occupancy as the fraction of time that the measured power level exceeds a certain threshold, it was found that this value is asymptotically Gaussian distributed whose mean and variance only depend on the means of the sojourn times, not their actual distributions. Then the modelling boils down to the estimation of the probability that the measured power level exceeds a certain threshold from measurements. In [44] and [45], this model was extended to the two-dimensional case by taking the channel dependence or frequency dependence into account. In particular, in the transition probabilities, in addition to the dependence on the previous state in time, the dependence on the previous state in an adjacent channel was also added, giving eight transition probabilities instead of four. As a result, one more parameter was added and the three parameters were calculated from the measurements. The developed model is more accurate than the original first-order CTMC in [43], at the cost of higher computational complexity. In [46], the model in [44] and [45] was extended to cyclostationary chain by taking the diurnal variation into account. In particular, the eight transition probabilities in [44] and [45] are functions of the hour of operation during the day now. Again, the three parameters were then calculated from the measurements.

The CTMC model requires that the sojourn time or the channel holding time follow exponential distributions. However, measurements in several campaigns revealed that this is not the case. Thus, the CTSMC model is used where another distribution is used to fit the channel holding time. References [47] and [48] studied the 2.4 GHz WLAN channel. The study first showed that the initial four states in the Markov chain can be well simplified to two states that denote either a transmit status or an idle status. However, unlike CTMC, the channel holding time of the idle status does not follow exponential distribution. Instead, it discovered that the generalized Pareto distribution fits the measurements better in [47] and the hyper-Erlang distribution fits the measurements better in [48], from the Kolmogorov-Smirnov test. The parameters of these models were then calculated from the measurements. These results were obtained based on the assumption that the channel does not suffer from any interferences. However, in reality, the 2.4 GHz WLAN channel has considerable interferences from other applications, such as microwave ovens and cordless phones. In [49], this more realistic scenario was studied. This study revealed that the hyper-exponential distribution provides better fits to the empirical curve, while simpler distributions, such as the generalized Pareto distribution, also provides good accuracy. The time-variance of this distribution was also discussed by calculating the sample average of the occupancy. Such results were also found in [34] and [38] for other frequency bands that favoured the long-tailed distributions, such as log-normal distribution, over the exponential distribution. In [50], more simple distributions and more frequency bands were studied for the distribution of the channel holding time in the CTSMC model. This study confirmed the invalidity of the exponential distribution and also suggested the generalized Pareto distribution for different frequency bands, when the sampling rate is relatively low.

There are a few other works that do not use the standard Markov chain. An empirical discrete-time Markov chain was used to model the spectrum occupancy in [51] and [52]. This work proposed the use of discrete-time Markov chain such that the channel does not stay in any of the states. Instead, it keeps switching between states. Thus, this model cannot be used to describe the channel holding time that is quite common in practice. To accommodate this in the discrete-time, the transition probabilities were made functions of time. Both deterministic method and stochastic method were used to determine the transition probabilities as functions of time from the measurements and the empirical curves seem to match very well with the fitted curves using this model. In [53], a heuristic model was proposed by assuming exponential holding times, Gaussian transmission powers and uniform centre frequencies. A detailed description of the proposed model was given. This is the first time that centre frequency and transmission power are considered. However, compared with the Markov chain, this model is quite heuristic.

C. Linear Regression

The Markov chain mainly describes the variation of spectrum occupancy with time. Indeed, it provides an accurate description of the time dimension of the spectrum occupancy. On the other hand, in addition to time, the spectrum occupancy is also largely dependent on the location and the frequency considered, as shown by almost all the afore-mentioned works. In this case, the Markov chain is not very useful to describe the frequency dimension and the space dimension of the spectrum occupancy. To describe the frequency and space dimensions, or in some cases a mixture of different dimensions, the linear regression model is often used.

This method was first proposed and termed as the Laycock-Gott model in [54] for HF spectrum occupancy modelling. The Laycock-Gott model uses the logit transformation to the spectrum occupancy rate, which is then modelled as a linear function of all the parameters that affect the spectrum occupancy, including time, frequency, location, and threshold. Denote Q as the spectrum occupancy rate, one has

$$y = \ln\left(\frac{Q}{1-Q}\right) = \sum_{i}^{I} a_i x_i \tag{9}$$

where a_i is a coefficient to be determined, x_i is a parameter that affects the spectrum occupancy and I is the total number of parameters that affect the spectrum occupancy. The measured spectrum occupancy rates with the corresponding parameters, such as threshold and frequency, were then used in this model to calculate the coefficients. If a binomial distribution for the occupancy rate in each HF band is assumed, further works in [54] showed that the model can be simplified to [54, eq. (3)]

$$y = A_k + B \times detectionthreshold(dBm) + (C_0 + C_1 f_k + C_2 f_k^2) \times sunspotnumber \tag{10}$$

where A_k, B, C_0, C_1 and C_2 are the coefficients to be determined, and f_k is the centre frequency of the k-th HF band. These coefficients can be determined using the maximum likelihood method. This model was revisited in [55], where the Laycock-Gott model was applied to a specific measurement system to model the HF spectrum occupancy.

In [56], the Laycock-Gott model was extended to account for the seasonal variation by adding the week of the year in the linear function, again, for HF spectrum occupancy modelling. Specifically, using similar assumptions, the model in this case becomes

$$
\begin{aligned}
y &= A_k + (B_0 + B_1 f_k + B_2 f_k^2) \times detectionthreshold(dBm) \\
&\quad + (C_0 + C_1 f_k) \times sunspotnumber \\
&\quad + (D_0 + D_1 f_k + D_2 f_k^2 + D_3 f_k^3)\cos(\omega \cdot week) \\
&\quad + (E_0 + E_1 f_k + E_2 f_k^2 + E_3 f_k^3)\cos(\omega \cdot week) \times sunspotnumber \\
&\quad + F_0 \cos(2\omega \cdot week)
\end{aligned} \tag{11}
$$

where $\omega = \frac{2\pi}{52}$. The cosine function is used to recognize the fact that the seasonal variation is a cycle. In this case, there are more coefficients to be determined and thus the model is more complicated. One sees that the Laycock-Gott model is very flexible and has great potential to be extended in many related cases.

In [57], linear regression was used to describe the mixed effects of the measuring times, frequency bands and locations. Three frequency bands in the range of 88 MHz - 3 GHz, five different locations in the USA and different times were chosen as the parameters, whose coefficients were fitted from the measurements. More specifically,

$$
\begin{aligned}
Q &= a_0 + a_1 \times band1 + a_2 \times band2 + a_3 \times band3 \\
&\quad + b_1 \times location1 + b_2 \times location2 + b_3 \times location3 + b_4 \times location4 \\
&\quad + c_1 \times weekend + c_2 \times afternoon + \epsilon
\end{aligned} \tag{12}
$$

where a_0, a_1, a_2, a_3, b_1, b_2, b_3, b_4, c_1 and c_2 are the coefficients to be determined from the measurements, ϵ is the error term and the rest of the variables are indicators of either 1 or 0. Results showed that a good fit could be achieved using their measurements. Comparing (12) with the Laycock-Gott model, apparently (12) does not consider the fact that the occupancy rate can only be between 0 and 1 and thus, transformations would be necessary before regression. Also, unlike the Laycock-Gott model, the variables have been treated equally in (12).

Most of the above works focus on the time and/or frequency dimensions of the spectrum occupancy. However, the space dimension is as important as the time and frequency dimensions. In fact, many of these works show that the spectrum occupancies in different locations are considerably different. In a realistic system, different locations in the same network will have different coverage and thus, different spectrum occupancies. This motivates the investigation of the space dimension of the spectrum occupancy. Ideally, this could be done by running several identical measuring systems simultaneously in a grid of locations with reasonable grid sizes and analysing the obtained measurements for spatial distributions. This would incur a very high cost of the measurement campaign, one of the reasons why there have been so few works on the space dimension of the spectrum occupancy. Nevertheless, some works have modelled the space dimension of the spectrum occupancy.

In [58] and [59], the space dimension model of the spectrum occupancy was proposed using a deterministic model for the duty cycle as a function of various parameters in the system, such as the activity factor, the probability of false alarm, the mean and variance of the primary user power. This model is built by combining the assumed number and locations of the primary users and the assumed primary user power and propagation patterns. Using the number of primary users, the transmission power and the power loss during the propagation of each primary user, and the locations of the primary users, one can calculate the received power at the CR for the spectrum occupancy as a linear combination of powers from all transmitting primary users. The advantage of this model is that it is deterministic. Thus, given the primary user propagation parameters, the primary user transmission powers and the distances to the primary users, the duty cycle can be calculated from the formula. The disadvantage is its inflexibility. Any mismatch between the assumed patterns and the real patterns may lead to a poor modelling.

In [60], a similar deterministic model was also proposed by using the assumed primary user transmission power patterns and propagation models. However, unlike [58] and [59], this model does not assume known number and locations of the primary users. Instead, it uses the random field theory and the spatial point process to model the number and locations of the primary users.

TABLE III

MAIN PROS AND CONS OF DIFFERENT MODELLING METHODS

Method	Pros	Cons
Statistics	simple, reliable	incomplete model
CDF, PDF	basic statistical model	no time-, frequency-, location-, or threshold-variance
Markov Chain	comprehensive model with time-variance	complicated, no frequency-, location-, or threshold-variance
Linear regression	comprehensive model with time-, frequency-, location- and threshold-variance	complicated and heuristic

This model is then fitted to the measurements obtained in a grid of area. It was shown that this model works relatively well in terms of the semivariograms.

In another relevant work [61], the spatial distribution of the spectrum occupancy om the TV band was simply obtained by locating all the primary TV transmitters in the UK, calculating the ”safe” distances from the transmitters using the propagation model, and counting the number of TV channels available in this distance. This is a reasonable method for modelling, as the TV transmitters are of fixed locations. The study found that on average a total of about 150 MHz channels are available for low-power CR operations but the available channels are scattered around. Thus, orthogonal frequency division multiplex may be required to collect all the channels for wideband services.

Note that these spectrum occupancy models are statistical models that are used to improve the average accuracy of the occupancy detection, not the instantaneous accuracy of the occupancy detection. The instantaneous accuracy depends on the specific realization of the spectrum occupancy as a random variable or random process and varies from realization to realization, but its performance can be improved on average by using its statistical behaviours, as in [2] - [10]. This is also the purpose of many other statistical detection and estimation applications that aim for average performance improvement rather than instantaneous performance improvement, such as average bit error rate improvement in fading channels. Another advantage of the spectrum occupancy model is its use in proactive spectrum access. The work in [3] used the Markov chain model of the spectrum occupancy to detect the primary user based on past observations. Traditional spectrum sensing detects the primary user based on current observations and switches channels after detecting the primary user, causing unavoidable disruption, while proactive spectrum access based on the spectrum occupancy model predicts the primary user status using the expected channel idle times and switches to channels with longer idle times preemptively. Computer simulation has shown that proactive access based on spectrum occupancy model can increase the channel utilization by 3% and reduce interference to the primary user by 30%. Using a testbed it was also shown that the average throughput increases by 10% with less fluctuations. This confirms the advantages of using spectrum occupancy model. It is not possible to fairly compare spectrum prediction with spectrum sensing in terms of receiver operating characteristics (ROC) in this case. First, it is difficult to compare prediction with detection as two different statistical methods. Second, prediction and sensing are determined by different parameters. Although the definitions of probabilities of detection and false alarm are the same for both prediction and sensing, the probabilities of detection and false alarm as functions of system parameters are unknown for prediction. Thus, one does not know how to adjust these parameters for a fair comparison. For example, prediction often depends on the number of predictors. It is not clear how to choose this parameter for sensing, as sensing does not depend on it. Finally, the purpose of ROC comparison has already been served by comparisons in terms of channel utilization and disruption rate. The false alarm probability is related to the channel utilization, while the detection probability is related to the disruption rate. It is not necessary to provide a ROC comparison in this paper either, as the main purpose of this paper is to provide a survey on spectrum modeling, where [3] is only used as a motivation for existing works. It is more appropriate and easier for [3] to provide such a comparison using their existing testbed and codes.

IV. Spectrum occupancy prediction

The measurement-based modelling methods discussed so far are related. For example, the average duty cycle studied in Section II can be considered as a distribution parameter for the CDF or PDF of the duty cycle in Section III-A. The CDF or PDF of the duty cycle in Section III-A can be used to describe the channel holding time of the Markov chain in Section III-B. They do describe the statistical behaviours of the spectrum occupancy as a whole, but they do not give the actual value of the spectrum occupancy. In some applications, it is also of interest to obtain such a value so that one is able to predict what will happen in the future. In this case, spectrum occupancy prediction may be used.

A commonly used method for spectrum occupancy prediction is the auto-regressive (AR), moving-average (MA) or auto-regressive moving-average (ARMA) model. Define X_t as a time series. The AR model of order p is of the form

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t,$$

where $\phi_1, \phi_2, \ldots, \phi_p$ ($\phi_p \neq 0$) denote the coefficients and are constants, and ϵ_t is a Gaussian white noise series with mean zero and variance σ^2. Note that the mean of X_t is zero. If the mean of X_t is not zero, say μ, then replace X_t by $X_t - \mu$, that is,

$$X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \epsilon_t,$$

where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$. The MA model of order q is defined as

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q},$$

where the coefficients $\theta_1, \theta_2, \ldots, \theta_q$ $(\theta_q \neq 0)$ are constant, and the noise ϵ_t is assumed to be Gaussian white noise. The ARMA model of orders p and q is given by

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}.$$

If X_t has a nonzero mean μ, the model is expressed as

$$X_t = \alpha + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q},$$

where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$.

Different methods may be applied to different data. One way of choosing these models is to calculate the auto-correlation function (ACF) and partial auto-correlation function (PACF). If the ACF decreases gradually with the lag and the PACF is zero when the lag is larger than p, a suitable model is likely to be the AR model. If the ACF is zero when the lag is larger than q and the PACF decreases gradually with the lag, the MA model may be more suitable. If both the ACF and PACF decrease gradually with the lag, the ARMA model is a good choice.

In [62], the duty cycle was predicted by using the measurements for the frequency range of 100 MHz - 2.4 GHz in a fixed outdoor location during one week. The GSM band data were used as an example to show that it can be described by an ARMA model, as it has cyclic trends. The fitted data were compared with the measurements to show good agreement in terms of the Akaike information correction criterion. It was pointed out that for stationary data, such as the TV band, a simpler AR model can be used.

In [63], the radio resource availability of the WLAN channel, defined as the sum of the packet occupation time in the channel in seconds, was predicted using the AR model. The order of the AR model was chosen between 10 and 18, depending on the data considered and the mean squared error requirement. This work also compared the prediction that is based on n-step-ahead for a time series with intervals of 1 second and the prediction based on 1-step-ahead for a time series with intervals of n seconds. They almost have identical prediction performances.

In [64] and [65], the duty cycle was predicted by using the AR model with logit transformations of the linearly combined binary spectrum occupancy rates to account for the fact that the duty cycle is between 0 and 1. Four locations for the GSM band were examined and a model order of $p = 3$ was shown to have acceptable performances. The prediction error was calculated and was shown to be dependent on the location.

Note that the above model applies to the general case when the spectrum statuses are correlated. In the special case when the spectrum statuses are independent, this model can still be applied but may not produce a meaningful prediction. From the expression of the ARMA model, the correlation of the time series comes from the terms $X_{t-1}, \cdots, X_{t-p}$ and $\epsilon_{t-1}, \cdots, \epsilon_{t-q}$ and thus, the coefficients of ψ and θ should be zero when the spectrum statuses are independent. Then, $X_t = \mu + \epsilon_t$, giving a constant trend plus noise. Since the white noise ϵ_t has a flat spectrum and does not contain any useful information, the prediction becomes the mean of X_t in the mean squared error sense. One does not have a better prediction in this model. However, since the actual value is the sum of the mean and the noise and the predicted value is the mean, this is equivalent to tossing a coin randomly. Thus, the produced prediction as the mean is not meaningful, which agrees with the intuition that one cannot use past statuses to predict the occupancy accurately when the spectrum statuses are independent, although one could still produce such a prediction. In reality, the spectrum statuses are correlated, as shown in [66] and [34]. Thus, the prediction done in [62] - [65] is useful.

There are other works on auto-regressive prediction, such as [67]. Also, there exist other works that use machine learning for spectrum sensing, such as neural networks [68] and hidden Markov model [69]. A survey of artificial intelligence for cognitive radios can be found in [70]. Since [70], more recent works on the use of machine learning for spectrum sensing include [71] that used artificial neural network technique to detect primary user in low signal-to-noise ratio scenarios, [72] that used unsupervised learning to evolve the classifier in sensing with security countermeasures, [73] that used support vector machine to outperform energy detection, [74] that used unsupervised K-means clustering and Gaussian mixture model as well as supervised support vector machine and K-nearest neighbour for cooperative spectrum sensing, and [75] that also used support vector machine to detect weak primary user signals, to name a few. These works do not use measurement data for verification. Since this paper focuses on measurement-based models and [70] already provides a good survey for non-measurement-based works, they are not discussed here. Table III shows the main contributions of the works discussed in this survey.

In general, the spectrum occupancy models can be applied to standard cognitive radio systems

TABLE IV

MAIN CONTRIBUTIONS OF THE WORKS DISCUSSED IN THE SURVEY

Model	Contributions
Statistics	more opportunities beyond 1 GHz and indoor
PDF of power	generally asymmetric and sometimes double-peaked
PDF of duty cycle	modified Beta
PDF of holding time	hyper-Erlang, hyper-exponential, generalized Pareto
PDF of available channels	Poisson-normal, Camp-Paulson
Markov chain	continuous-time semi-Markov chain
Linear regression	Laycock-Gott model
Prediction	AR or ARMA

in several ways. First, the simple statistics can be used as prior knowledge to improve current sensing accuracies. For example, they can be used as the *a priori* probabilities for the null and alternative hypotheses in energy detection in the physical layer. Second, the PDF and CDF may be used in signal detection. Current detectors are based on the maximum likelihood principle, while these PDFs and CDFs can be used for maximum *a posteriori* detection. Third, the Markov chain, the linear regression and the ARMA models can be used in the network layer for resource management. All of them will improve the system performance, as shown in [2] - [10].

It is worth mentioning that [76] also provided a brief review of some of the measurement campaigns discussed in this survey up to 2009. Reference [77] compiled the results in [39], [51] and [59]. In addition, they proposed a new model considering time-, frequency- and space-variance by focusing on the Markov process and the correlation. Reference [78] provides a complementary review of spectrum models, focusing more on theoretical models instead of measurement-based models. It has a brief discussion of the Markov models and time series as well.

V. CHALLENGES

It can be seen from the above that spectrum occupancy modelling for obtained measurements is a very complicated and challenging task to perform and it depends on a lot of known and unknown factors. Although much has been done on measurement-based spectrum occupancy modelling, especially in the recent years due to the great interest in opportunistic access to under-utilized frequency bands, a lot of open problems remain.

First, future communications systems are more and more likely to be a heterogeneous network that consists of long-range communications, such as cellular communications and satellite communications, as well as short-range local communications, such as Bluetooth and peer-based relaying. These local communications have largely been ignored in the measurement campaigns performed, as most outdoor measurements have been taken on the roofs of high buildings so far. If the measurements are taken at the street level, the radio activities might be significantly different, and in this case the local radio hot spots may have a huge impact. The challenge here is that there are so many possible hot spots that it may be difficult to find a typical scenario. A mobile monitoring system may be a solution to this problem.

Second, most models are established based on the observations from the measurements and are extracted by fitting them to the well-known theoretical models. This leads to a large number of different models for similar situations. They are very inconvenient to use, as one has to determine which model to use first before making any use of the model. The question arises whether a general, if not universal, model exists that can unify most existing models. This problem is likely to be solved by using a complicated model with more parameters but if such a model exists, it does provide great convenience.

Third, it has been discussed in most of these works that time, frequency, location and detection threshold are the four most important factors that affect the spectrum occupancy. So far, most modelling works have studied the time-dependence of the spectrum occupancy. A few of them have also studied the frequency-dependence of the spectrum occupancy. And a very limited number of them have studied the effects of location and detection threshold on the spectrum occupancy. One would ask if a more comprehensive model that takes all these important factors into account can be derived. Linear regression may be a solution to this problem but certainly more sophisticated regression models would be required than those in the literature.

Finally, although most frequency bands in the range between 30 MHz and 3 GHz have been measured, current modelling works are limited to only a few bands, such as the GSM band, the ISM band and the HF bands, especially for the Markov chain modelling and the linear regression modelling. It is yet unknown if other bands have similar statistical behaviours or completely different. Thus, it is worthwhile to extend existing modelling methods to all frequency bands to see if they have any common behaviours.

REFERENCES

[1] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE Journal of Selected Areas in Communications*, vol. 23, pp. 201 - 220, Feb. 2005.

[2] Z. Sun, G.J. Bradford, J.N. Laneman, ”Sequency detection algorithms for PHY-layer sensing in dynamic spectrum access networks,” *IEEE J. Selected Topics in Signal Processing*, vol. 5, pp. 97 - 109, Sept. 2010.

[3] L. Yang, L. Cao and H. Zheng, ”Proactive channel access in dynamic spectrum networks,” *Physical Commun.*, vol. 1, pp. 103 - 111, June 2008.

[4] M. Hoyhtya, S. Pollin, A. Mammela, ”Classification-based predictive channel selection for cognitive radios,” *Proc. ICC’10*, pp. 1 - 5, Cape Town, South Africa, May 2010.

[5] X. Zhou, J. Ma, Y. Li, Y.H. Kwon, A.C.K. Soong and G. Zhao, ”Probability-based transmit power control for dynamic spectrum access,” *Proc. DySPAN’08*, pp. 1 - 5, Chicago, USA, Oct. 2008.

[6] X. Zhou, J. Ma, G.Y. Li, Y.H. Kwon, A.C.K. Soong, ”Probability-based optimization of inter-sensing duration and power control in cognitive radio,” *IEEE Trans. Wireless Commun.*, vol. 8, pp. 4922 - 4927, Oct. 2009.

[7] K.W. Sung, S.-L. Kim, J. Zander, ”Temporal spectrum sharing based on primary user activity prediction,” *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3848 - 3855, Dec. 2010.

[8] G. Yuan, R.C. Grammenos, Y. Yang, W. Wang, ”Performance analysis of selective opportunistic spectrum access with traffic prediction,” *IEEE Trans. Vehicular Technology*, vol. 59, pp. 1949 - 1959, May 2010.

[9] M. Timmers, S. Pollin and F. Catthoor, ”On the value of prediction in opportunistic radio systems,” *Proc. ICC’11*, pp. 1 - 5, Kyoto, Japan, June 2011.

[10] F. Bouali, O. Sallent, J. Perez-Romero and R. Agusti, ”Strengthening radio environment maps with primary-user statistical patterns for enhancing cognitive radio operation,” *Proc. CrownCom’11*, pp. 256 - 260, Osaka, Japan, June 2011.

[11] F.H. Sanders, ”Broadband spectrum surveys in Denver, CO, San Diego, CA, and Los Angeles, CA: methodology, analysis and comparative results,” *Proc. 1998 IEEE Int. Symp. on Electromagnetic Compatibility*, vol. 2, pp. 988 - 993, Denver, USA, Aug. 1998.

[12] M.A. McHenry, P.A. Tenhula and D. McCloskey, ”Chicago spectrum occupancy measurements & analysis and a long-term studies proposal,” *Proc. TAPAS’06*, 2006.

[13] M. A. McHenry et al., Spectrum occupancy measurements, Shared Spectrum Company, Vienna, VA, USA. (available at http://www.sharedspectrum.com)

[14] D.A. Roberson, C.S. Hood, J.L. LoCicero, and J.T. MacDonald, ”Spectral occupancy and interference studies in support of cognitive radio technology deployment,” *Proc. SDR’06*, pp. 26 - 35, Reston, USA, Sept. 2006.

[15] R. B. Bacchus, A. J. Fertner, C. S. Hood, and D. A. Roberson, ”Long-term, wide-band spectral monitoring in support of dynamic spectrum access networks at the IIT spectrum observatory,” *Proc. DySPAN’08*, pp. 1 - 10, Oct. 2008.

[16] M. Wellens, J. Wu and P. Mahonen, ”Evaluation of spectrum occupancy in indoor and outdoor scenario in the context of cognitive radio,” *Proc. CrownCom’07*, pp. 420 - 427, Orlando, USA, Aug. 2007.

[17] M. Lopez-Benitez, A. Umbert, and F. Casadevall, ”Evaluation of spectrum occupancy in Spain for cognitive radio applications,” *Proc. VTC-Spring’09*, pp. 1 - 5, Apr. 2009.

[18] M. H. Islam, C. L. Koh, S. W. Oh, X. Qing, Y. Y. Lai, C. Wang, Y.-C. Liang, B. E. Toh, F. Chin, G. L. Tan, and W. Toh, ”Spectrum survey in Singapore: Occupancy measurements and analyses,” *Proc. CrownCom’08*, pp. 1 - 7, May 2008.

[19] V. Valenta, R. Marselek, G. Baudoin, M. Villegas, M. Suarez, and F. Robert, ”Survey on spectrum utilization in Europe: measurements analyses and observations,” *Proc. CrownCom’10*, pp. 1 - 5, June 2010.

[20] A. Petrin and P. G. Steffes, ”Analysis and comparison of spectrum measurements performed in urban and rural areas to determine the total amount of spectrum usage,” *Proc. ISART’05*, pp. 9 - 12, Mar. 2005.

[21] M. Biggs, A. Henley and T. Clarkson, ”Occupancy analysis of the 2.4 GHz ISM band,” *IEE Proceedings - Commun.*, vol. 151, pp. 481 - 488, Oct. 2004.

[22] V. Blaschke, H. Jaekel, T. Renk, C. Kloeck, F.K. Jondral, ”Occupation measurements supporting dynamic spectrum allocation for cognitive radio design,” *Proc. CrownCom’07*, pp. 50 - 57, Orlando, USA, Aug. 2007.

[23] R. Bacchus, T. Taher, K. Zdunek, D.A. Roberson, ”Spectrum utilization study in support of dynamic spectrum access for public safety,” *Proc. DySPAN’10*, pp. 1 - 11, Singapore, Apr. 2010.

[24] S. D. Jones, E. Jung, X. Liu, N. Merheb, and I.-J.Wang, ”Characterization of spectrum activities in the U.S. public safety band for opportunistic spectrum access,” *Proc. DySPAN’07*, pp. 137 - 146, Dublin, Ireland, Apr. 2007.

[25] K. A. Qaraqe, H. Celebi, A. Gorcin, A. El-Saigh, H. Arslan, and M. S. Alouini, ”Empirical results for wideband multidimensional spectrum usage,” *Proc. PIMRC’09*, pp. 1262 - 1266, Sept. 2009.

[26] S. Contreras, G. Villardi, R. Funada, H. Harada, ”An investigation into the spectrum occupancy in Japan in the context of TV white space systems,” *Proc. CrownCom’11*, pp. 341 - 345, Osaka, Japan, June 2011.

[27] R. Schiphorst and C. H. Slump, ”Evaluation of spectrum occupancy in Amsterdam using mobile monitoring vehicles,” *Proc. VTC-Spring’10*, pp. 1 - 5, May 2010.

[28] T. Harrold, R. Cepeda and M. Beach, ”Long-term measurements of spectrum occupancy characteristics,” *Proc. DySPAN’11*, pp. 83 - 89, Aachen, Germany, May 2011.

[29] T.M. Taher, R.B. Bacchus, K.J. Zdunek and D.A. Roberson, ”Long-term spectral occupancy findings in Chicago,” *Proc. DySPAN’11*, pp. 100 - 107, Aachen, Germany, May 2011.

[30] D. Datla, A.M. Wyglinski, G.J. Minden, ”A spectrum surveying framework for dynamic spectrum access networks,” *IEEE Trans. Vehicular Technol.*, vol. 58, pp. 4158 - 4169, Oct. 2009.

[31] A. Petrin, ”Maximizing the utility of radio spectrum: broadband spectrum measurements and occupancy model for use by cognitive radio,” Ph.D. thesis, Georgia Institute of Technology, 2005.

[32] R. I. C. Chiang, G. B. Rowe, and K. W. Sowerby, ”A quantitative analysis of spectral occupancy measurements for cognitive radio,” *Proc. VTC-Spring’07*, pp. 3016 - 3020, Dublin, Ireland, Apr. 2007.

[33] P.F. Marshall, ”Closed-form analysis of spectrum characteristics for cognitive radio performance analysis,” *Proc. DySPAN’08*, pp. 1 - 12, Chicago, USA, Oct. 2008.

[34] M. Wellens and P. Mahonen, ”Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model,” *Mobile Netw. Appl.*, vol. 15, pp. 461 - 474, 2010.

[35] L. Mendes, L. Goncalves, A. Gameiro, ”GSM downlink spectrum occupancy modeling,” *Proc. PIMRC’11*, pp. 546 - 550, Toronto, Canada, Sept. 2011.

[36] S.W. Ellingson, ”Spectral occupancy at VHF: Implications for frequency agile cognitive radios,” *Proc. VTC-Fall’05*, pp. 1379 - 1382, Sept. 2005.

[37] S. Yin, Q. Zhang, E. Zhang, L. Yin, S. Li, ”Statistical modeling for spectrum usage characterizing wireless fading channels and mobile service dynamics,” *IEEE Trans. Vehicular Technol.*, vol. 62, pp. 3800 - 3812, Oct. 2013.

[38] , M. Pavloski, V. Rakovic, V. Atanasovski, L. Gavrilovska, ”Multi-level modeling of spectrum use,” *Proc. ISWCS’12*, pp. 81 - 85, Paris, France, Aug. 2012.

[39] M.Wellens, J. Riihijrvi, and P. Mohonen, ”Empirical time and frequency domain models of spectrum use,” *Phys. Commun.*, vol. 2, no. 1/2, pp. 10 - 32, Mar. 2009.

[40] C. Ghosh, S. Roy and M.B. Rao, ”Modeling and validation of channle idleness and spectrum availability for cognitive networks,” *IEEE J. Sele. Areas in Commun.*, vol. 30, pp. 2029 - 2040, Nov. 2012.

[41] C. Ghosh, S. Roy, M.B. Rao, D.P. Agrawal, ”Spectrum occupancy validation and modeling using real-time measurements,” *Proc. CoRoNet’10*, pp. 25-30, New York, USA, Sept. 2010.

[42] K. Arshad and K. Moessner, ”Statistical models of spectrum opportunities for cognitive radio,” *Proc. PIMRC’11*, pp. 2310 - 2314, Toronto, Canada, Sept. 2011.

[43] A.D. Spaulding and G.H. Hagn, ”On the definition and estimation of spectrum occupancy,” *IEEE Trans. Electromagnetic Compatibility*, vol. EMC-19, pp. 269 - 280, Aug. 1977.

[44] A.J. Gibson and L. Arnett, ”Statistical modeling of spectrum occupancy,” *Electronics Letters*, vol. 29, pp. 2175 - 2176, Dec. 1993.

[45] A.J. Gibson and L. Arnett, ”Measurements and statistical modeling of spectrum occupancy,” *Proc. HF Radio Systems and Techniques’94*, pp. 150 - 154, York, UK, July 1994.

[46] D.J. Percival, M. Kraetzl, and M.S. Britton, ”A Markov model for HF spectral occupancy in central Australia,” *Proc. HF Radio Systems and Techniques’97*, pp. 14 - 18, Nottingham, UK, July 1997.

[47] S. Geirhofer, L. Tong, and B.M. Sadler, ”A measurement-based model for dynamic spectrum access in WLAN channels,” *Proc. MILCOM’06*, pp. 1 - 7, Oct. 2006.

[48] S. Geirhofer, L. Tong, and B. M. Sadler, ”Dynamic spectrum access in the time domain: modeling and exploiting white space,” *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 66 - 72, May 2007.

[49] L. Stabellini, ”Quantifying and modeling spectrum opportunities in a real wireless environment,” *Proc. WCNC’10*, pp. 1 - 6, Apr. 2010.

[50] M. Lopez-Benitez and F. Casadevall, ”Time-dimension models of spectrum usage for the analysis, design and simulation of cognitive radio networks,” *IEEE Trans. Vehicular Technology*, vol. 62, pp. 2091 - 2104, June 2013.

[51] M. Lopez-Benitez and F. Casadevall, ”Empirical time-dimension model of spectrum use based on a discrete-time Markov chain with deterministic and stochastic duty cycle models,” *IEEE Trans. Vehicular Technology*, vol. 60, pp. 2519 - 2533, July 2011.

[52] M. Lopez-Benitez and F. Casadevall, ”Discrete-time spectrum occupancy model based on Markov chain and duty cycle models,” *Proc. DySPAN’11*, pp. 90 - 99, May 2011.

[53] C. Ghosh, S. Pagadarai, D. Agrawal, A. Wyglinski, ”A framework for statistical wireless spectrum occupancy modeling,” *IEEE Trans. on Wireless Communications*, vol. 9, pp. 38-44, Jan. 2010.

[54] P.J. Laycock, M. Morrell, G.F. Gott and A.R. Ray, ”A model for HF spectral occupancy,” *Proc. HF Radio Systems and Techniques*, pp. 165 - 171, London, UK, Apr. 1988.

[55] C.A. Pantjiaros, P.J. Laycock, G.F. Gott and S.K. Chan, ”Development of the Laycock-Gott occupancy model,” *IEE Proceedings-Communications*, vol. 144, pp. 33 - 39, Aug. 2002.

[56] C.A. Pantjiaros, J.A. Wylie, J. Brown, G.F. Gott and P.J. Laycock, ”Extended UK models for high frequency spectral occupancy,” *IEE Proceedings - Communications*, vol. 145, pp. 168 - 174, June 1998.

[57] S. Pagadarai and A.M. Wyglinski, ”A linear mixed-effects model of wireless spectrum occupancy,” *EURASIP J. Wireless Commun. and Networking*, vol. 2010, pp. 1 - 7, 2010.

[58] M. Lopez-Benitez and F. Casadevall,”Statistical prediction of spectrum occupancy perception in dynamic spectrum access networks,” *Proc. ICC’11*, pp. 1 - 6, June 2011.

[59] M. Lopez-Benitez and F. Casadevall, ”Spatial duty cycle model for cognitive radio,” *Proc. PIMRC’10*, pp. 1629 - 1634, Sept. 2010.

[60] M. Wellens, J. Riihijarvi, and P. Mahonen, ”Spatial statistics and models of spectrum use,” *Elsevier Computer Commun.*, vol. 32, pp. 1998 - 2011, Aug. 2009.

[61] M. Nekovee, ”Quantifying the availability of TV white spaces for cognitive radio operation in the UK,” *Proc. ICC’09*, pp. 1 - 5, Dresden, Germany, June 2009.

[62] Z. Wang and S. Salous, ”Spectrum occupancy statistics and time series models for cognitive radio,” *J. Signal Process. Syst.*, vol. 62, no. 2, pp. 145 - 155, Feb. 2011.

[63] S. Kaneko, S. Nomoto, T. Ueda, S. Nomura, K. Takeuchi, ”Predicting radio resource availability in cognitive radio - an experimental examination,” *Proc. CrownCom’08*, pp. 1 - 8, Singapore, May 2008.

[64] A. Gorcin, H. Celebi, K.A. Qaraqe, and H. Arslan, ”An auto-regressive approach for spectrum occupancy modeling and prediction based on synchronous measurements,” *Proc. PIMRC’11*, pp. 705 - 709, Toronto, Canada, Sept. 2011.

[65] S. Yarkan, H. Arslan, ”Binary time series approach to spectrum prediction for cognitive radio,” *Proc. VTC-Fall’07*, pp. 1563 - 1567, Baltimore, USA, Sept. 2007.

[66] M. Wellens, J. Riihijarvi, and P. Mahonen, ”Empirical time and frequency domain models of spectrum use,” *Elsevier Physical Commun.*, pp. 10 - 32, 2009.

[67] Z. Wen, T. Luo, W. Xiang, S. Majhi, Y. Ma, ”Autoregressive spectrum hole prediction model for cognitive radio systems,”’ *Proc. ICC’08*, pp. 154 - 157, Beijing, China, May 2008.

[68] V.K. Tumuluru, P. Wang, D. Niyato, ”A neural network based spectrum prediction scheme for cognitive radio,” *Proc. ICC’10*, pp. 1 - 5, Cape Town, South Africa, May 2010.

[69] I.A. Akbar and W.H. Tranter, ”Dynamic spectrum allocation in cognitive radio using hidden markov models: Poisson distributed case,” in *Proc. of IEEE SoutheastCon*, pp. 196 - 201, March 2007.

[70] A. He, K.Y. Bae, T.R. Newman, J. Gaeddert, K. Kim, R. Menon, L. Morales-Tirado, J. Neel, Y. Zhao, J.H. Reed and W.H. Tranter, ”A survey of artificial intelligence for cognitive radios,” *IEEE Trans. Vehicular Technol.*, vol. 59, pp. 1578 - 1592, May 2010.

[71] Y.-J. Tang, Q.-Y. Zhang, W. Lin, ”Artificial neural network based spectrum sensing method for cognitive radio,” *Proc. WiCOM 2010*, pp. 1 - 4, Chengdu, China, Sept. 2010.

[72] T.C. Clancy, A. Khawar, T.R. Newman, ”Robust signal classification using unsupervised learning,” *IEEE Trans. Wireless Commun.*, vol. 10, pp. 1289 - 1299, Apr. 2011.

[73] D. Zhang, X. Zhai, ”SVM-based spectrum sensing in cognitive radio,” *Proc. WiCOM 2011*, pp. 1 - 4, Wuhan, China, September 2011.

[74] K.M. Thilina, K.W. Choi, N. Saquib, E. Hossain, ”Machine learning techniques for cooperative spectrum sensing in cognitive radio networks,” *IEEE J. Select. Areas Commun.,*, vol. 31, pp. 2209 - 2221, Nov. 2013.

[75] O.P. Awe, Z. Zhu, S. Lambotharan, ”Eigenvalue and support vector machine techniques for spectrum sensing in cognitive radio networks,” *Proc. TAAI 2013*, pp. 223 - 227, Taipei, China, Dec. 2013.

[76] M. Wellens, Empirical Modelling of Spectrum Use and Evaluation of Adaptive Spectrum Sensing in Dynamic Spectrum Access Networks, Ph.D. Thesis, 2010.

[77] M. Lopez-Benitez and F. Casadevall, ”Spectrum usage models for the analysis, design and simulation of cognitive radio networks,” *Cognitive Radio and its Application for Next Generation Cellular and Wireless Networks*, Lecture Notes in Electrical Engineering (ed. by H. Venkataraman and G.-M. Muntean), Springer, 2012.

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

chen_photo.eps

Yunfei Chen (S’02-M’06-SM’10) received his B.E. and M.E. degrees in electronics engineering from Shanghai Jiaotong University, Shanghai, P.R.China, in 1998 and 2001, respectively. He received his Ph.D. degree from the University of Alberta in 2006. He is currently working as an Associate Professor at the University of Warwick, U.K. His research interests include wireless communications, cognitive radios, channel modeling, SNR estimation, diversity, modulation and UWB systems.

heeseok.eps

Hee-Seok Oh received the Ph.D. degree in statistics from Texas A&M University, College Station, U.S.A. in May 1999. He was with Department of Mathematics, University of Bristol, Bristol, U.K. (1999-2000), National Center for Atmospheric Research, Boulder, CO, U.S.A. (2000-2002), and Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada (2002-2004). Since September 2004, he has been with Department of Statistics, Seoul National University, Seoul, Korea, as a Professor. His current research interests include multiscale methods in statistics, function estimation, statistical signal processing, spatio-temporal data analysis, and statistical climatology.