REVIEW   **Open Access**

CrossMark

Statistical spectrum occupancy prediction for dynamic spectrum access: a classification

Hamid Eltom[1*] ⓘ, Sithamparanathan Kandeepan[1], Robin J. Evans[2], Ying Chang Liang[3] and Branko Ristic[1]

Abstract

Spectrum scarcity due to inefficient utilisation has ignited a plethora of dynamic spectrum access solutions to accommodate the expanding demand for future wireless networks. Dynamic spectrum access systems allow secondary users to utilise spectrum bands owned by primary users if the resulting interference is kept below a pre-designated threshold. Primary and secondary user spectrum occupancy patterns determine if minimum interference and seamless communications can be guaranteed. Thus, spectrum occupancy prediction is a key component of an optimised dynamic spectrum access system. Spectrum occupancy prediction recently received significant attention in the wireless communications literature. Nevertheless, a single consolidated literature source on statistical spectrum occupancy prediction is not yet available in the open literature. Our main contribution in this paper is to provide a statistical prediction classification framework to categorise and assess current spectrum occupancy models. An overview of statistical sequential prediction is presented first. This statistical background is used to analyse current techniques for spectrum occupancy prediction. This review also extends spectrum occupancy prediction to include cooperative prediction. Finally, theoretical and implementation challenges are discussed.

Keywords: Dynamic spectrum access, Spectrum occupancy, Cognitive radio, Spectrum prediction, Sequential prediction, Markov models, Universal prediction, Cooperative prediction, Mixture models, Bayesian prediction

1 Introduction

Spectrum scarcity has been a major research topic for the past few decades [1, 2]. Fixed spectrum allocation inefficiency has generated a proliferation of dynamic spectrum access solutions to accommodate the growing demand for wireless and mobile applications. Dynamic spectrum access (DSA) systems typically consist of licensed *primary* users and opportunistic *secondary* users. Primary users are the incumbent owners of the spectrum, while the secondary users opportunistically access the spectrum, and are required to inflict limited interference on the primary users (Fig. 1). To fulfil such requirements, secondary users must be equipped with a *cognitive ability*, and *reconfigurability*, to identify and exploit instantaneous availability of spectrum opportunities (holes) [1, 3]. Spectrum management framework classifies such cognitive ability into few generic functions, referred to as cognitive radio cycle
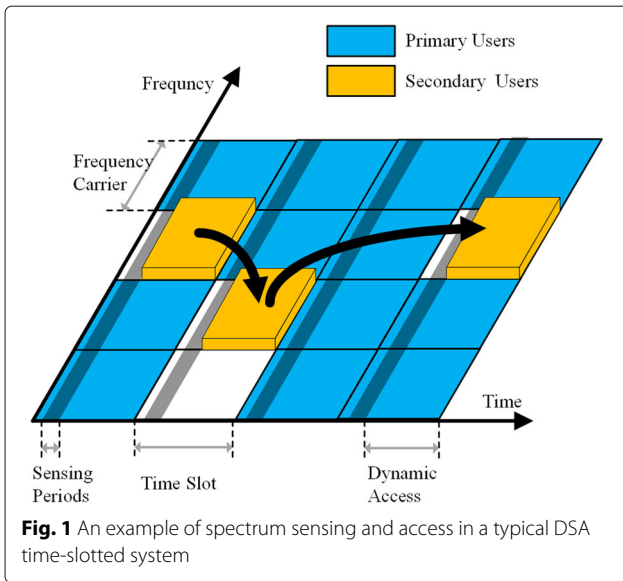
functions. These functions are represented by the secondary user's ability to perform spectrum *sensing*, *decision*, *sharing*, and *mobility* [3, 4]. Spectrum occupancy prediction (SOP) models were proposed in DSA literature to optimise cognitive cycle functions [5]. SOP models add agility, and adaptability to cognitive radio functions to optimise periodic spectrum sensing scheduling, and channel selection in spectrum decision (Fig. 2) [3]. Similarly, SOP models allow the implementation of a proactive spectrum mobility strategy based on predicted occupancy patterns which avoids collisions with incumbent primary users [5, 6].

SOP models for DSA systems broadly target occupancy parameters such as *channel availability*, i.e. prediction of channel status as idle or busy, as well as, *duty cycle*, i.e. prediction of the average fraction of time the primary user is occupying the channel [7, 8]. Measurements on spectrum occupancy show as in Fig. 3 that spectrum prediction is much required to improve spectrum utilisation efficiency. The common motivation for SOP techniques is to minimise the accumulated *time delay* due to cognitive cycle processing. By predicting the channel status in

*Correspondence: hamid.o.eltom@ieee.org
[1]School of Engineering, RMIT University, 124 La Trobe St, 3000 Melbourne, Australia
Full list of author information is available at the end of the article

Eltom *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:29

Page 2 of 17

Fig. 1 An example of spectrum sensing and access in a typical DSA time-slotted system

advance, more processing time is available for spectrum sensing, decisions, and mobility [5]. SOP models address *prediction* either *explicitly* [9–11] or *implicitly*. Implicit approaches present SOP models as primary/secondary user activity models. In this review, we address both implicit and explicit formulations as *statistical* SOP models. Statistical SOP models proposed for spectrum occupancy analysis include Poisson processes [12, 13], Bayesian prediction [9, 14], and linear regression [15, 16]. Machine learning-based techniques have also been proposed for model *learning* including neural networks, time regression, and space vector machines [5, 17, 18]. The surveys in [5, 6] provide a good taxonomy of primary user's

Fig. 2 Prediction spectrum prediction in dynamic spectrum access framework
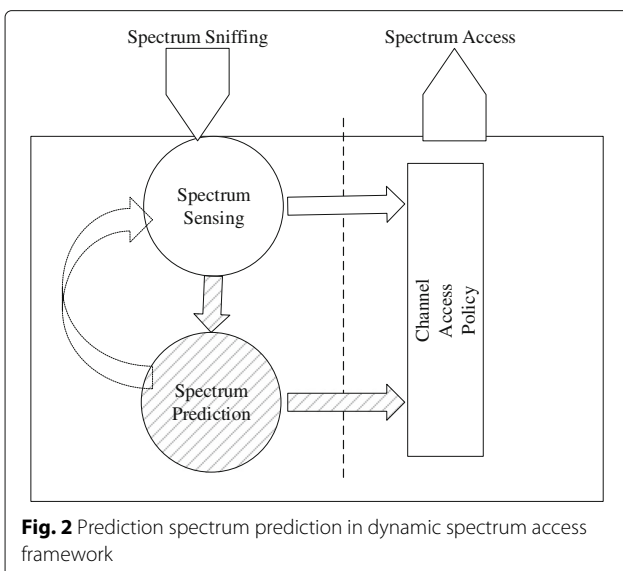
activity model collection. This review abstracts and consolidate SOP models in DSA systems, and extends the aforementioned works.

Our contribution in this review paper is a consolidated top-down classification of spectrum occupancy prediction. We present SOP taxonomy in a sequential prediction-based framework. This allows the authors to dissociate the spectrum prediction model from the application assumptions. In other words, this review paper addresses spectrum prediction model selection based on the theoretical sequential prediction stochastic class. The review places techniques adopted in literature into categories based on their theoretical predictor classes. This classification approach highlights candidate prediction techniques suitable for SOP scenarios not extensively covered in current literature. Firstly, we review the fundamentals of statistical prediction. Then, based on the stochastic mixture model framework, we review parametric and non-parametric approaches for underlying stochastic source assignment. Secondly, we describe spectrum occupancy prediction in terms of the stochastic class assignment. We extend mixture model formulation to cooperative spectrum occupancy prediction using decision (hard) and data (soft) fusion techniques. Finally, we elaborate on additional theoretical and practical challenges of sequential spectrum occupancy prediction implementation.

In Section 2, we outline the fundamentals of statistical sequential prediction and detail relevant aspects in section Section 3. A brief review of empirical and statistical-based approaches for SOP modelling is presented in Section 4. Then, we provide a review of current spectrum occupancy techniques in Sections 5, 6, and 7, respectively. This is followed by a review on cooperative spectrum prediction and fusion rules in Section 8. Lastly, we list the challenges in spectrum occupancy prediction in Section 9, and concluding remarks in Section 10.

2 Background
Is it possible to forecast the short-term evolution of an event? And if possible, how can we quantify the performance of this forecast? [19, 20]. Prediction theory asks such questions and attempts to formulate the problem, and quantify the prediction accuracy. *Sequential prediction* is deeply embedded in statistics [21], information theory [22, 23], machine learning [22–25], source coding theory [25], and gambling [26] among many other disciplines. The term prediction in the literature generally refers to sequential prediction, with an implicit notion of time dependency. However, unlike the estimation problem, the sequential prediction does not seek an interpretation of information, but rather an exploitation of the information to forecast future events [23]. A well-known
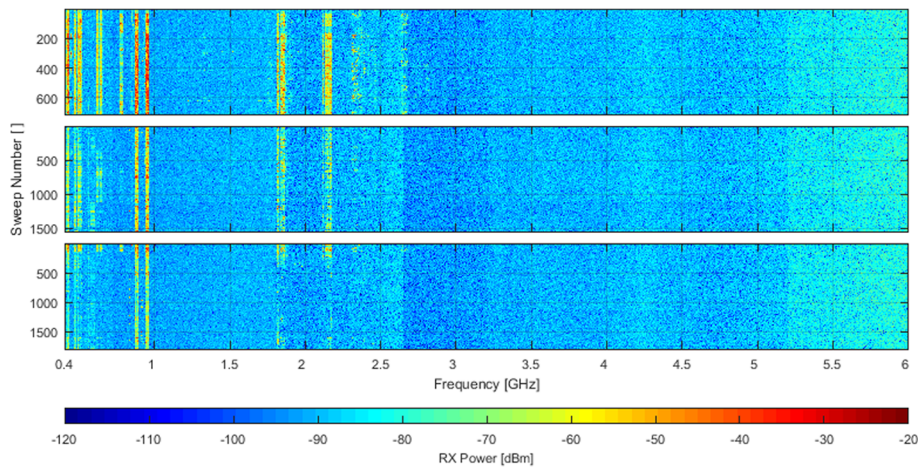
Eltom *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:29

Page 3 of 17

Fig. 3 Power measurement campaign sample for Melbourne LTE system measurements [8]

definition of the sequential prediction problem is [19, 20, 23, 27]:

Let a predictor receive a series of sequential observations $x^{t-1} = \{x_1, x_2 ..., x_{t-1}\}$ drawn from a sample space $\mathcal{X}$. At time instant t, the predictor performs an action a_t based on the previous observations x^{t-1} before the observation x_t is available. Once x_t is available, the predictor then updates the loss function $l(a_t, x_t)$.

The loss function $l(a_t, x_t)$ is a distance measure, e.g. a squared error $l(a_t, x_t) = (x_t - a_t)^2$. The action a_t is generally assigned $a_t = \hat{x}_t$ (where $\hat{x}_t$ is the predictor's guess of x_t) for "next event prediction". Alternatively, a_t can represent the *confidence* in next event prediction, i.e. the conditional probability $a_t = p_t\left(x_t|x^{t-1}\right)$ of one-step ahead prediction, given a series of observations up to $t - 1$. General loss function assignments transform sequential prediction problem into a decision problem [20, 27].

There are two main formulations of the sequential prediction problem. The first is *classical prediction* where the underlying source is assumed *known*, and the observations are assumed identically distributed (not necessarily independent). The second formulation is *universal prediction*, where no specific assumptions are made about how the observed series is generated[1]. Conceptually, universal prediction compares the designed predictor to an indexed set $\mathbb{M}$ of stochastic sources (e.g. distributions, codes, or polynomials). The true observation generating mechanism is generally assumed to be a member of the *predictor* stochastic source set $\mathbb{M}$ [20, 28]. The universal prediction algorithm is expected to perform at least as well as the *best* member of set $\mathbb{M}$ in terms of prediction loss [19, 29, 30]. The universal predictor is not necessarily a member of $\mathbb{M}$ [30], but can be created as a mixture

of predictor set $\mathbb{M}$ [31]. Universal prediction formulation can be summarised as:

Let $\mathbb{M}$ be an indexed set of arbitrary predictors. There exist prediction strategies for each sequence x^{t-1} that can possibly be realised, which can predict essentially as well as the predictor in $\mathbb{M}$ that turns out to be best for that sequence "with hindsight" [19, 30].

For example, a universal predictor may be compared to (or constructed from) a parametrised stochastic set $\{P_\theta, \theta \in \mathbb{M}\}$ such as a set of memoryless Poisson sources, a finite set of kth-order Markov models, or a set of autoregressive models of order p [19, 20, 23, 29]. However, the sequential predictor performance generally depends on the predictor set $\mathbb{M}$ class "complexity" or richness, which quantifies the class type, size, and statistical regression between observations [19, 20, 23, 29]. Thus, a set of finite k*th*-order Markov models is more practical for the predictor design than the set of all arbitrary order Markov models due to the set size (see [20] for universality guarantee and indexed class size). If the predictor utilises Bayesian methods, a well-known *Bayesian mixture model* is constructed as a weighted *linear* sum of the parametrised sources. Bayesian mixture models are the most common algorithms for predictor design (see Bayesian mixture models and redundancy-capacity theorem for optimality analysis [20, 23, 28, 31]). However, they are by no means the only available methods, nor perform well for all arbitrary loss functions [19, 29, 30][2].

3 Statistical prediction

In broad terms, a sequential predictor is either fitted to the observations series, i.e. *curve fitting* or the observation generating stochastic distribution, i.e. *density fitting* to estimate future observations. Thus, statistical prediction

is categorised based on the assumptions about the existence or non-existence of an underlying stochastic source, see [19, 20, 23, 29]. Statistical prediction is commonly presented under either probabilistic or deterministic settings. Prediction loss function, regret, and redundancy are discussed in Subsection 3.3, while Subsection 3.4 provides an overview of Bayesian-based techniques.

3.1 Probabilistic settings
The classical definition of the sequential prediction problem assumes an arbitrary *known* stochastic process $\{\mathbb{P}_\theta, \theta \in \mathbb{M}\}$ is responsible for generating the observations x^t [21, 32, 33]. Accordingly, optimal prediction is formulated as the minimisation of the expected value of the predictor loss function [20, 23, 27, 28, 34]. For example, if $\{X_t\}$ is an arbitrary parametrised random source, the action $a_t = \hat{x}_t$ is set as next observation prediction, and the loss function is the squared distance $l(a_t, x_t) = E(a_t - x_t)^2$ then the optimal predictor will always choose the conditional mean as it is predicted value. One of the most well-known techniques that utilises this approach is the Kalman filter [24, 35, 36] (see Section 6). Practically, the underlying stochastic process are unknown, so a *replacement stochastic assignment* $\mathbb{Q}$ is created based on the predictor set $\mathbb{M}$ of stochastic predictors. The performance of the designed sequential predictor $\mathbb{Q}$ is compared to the best predictor $\mathbb{P}$ in the class $\mathbb{M}$. The designed predictor $\mathbb{Q}$ has asymptotically small prediction regret compared to $\mathbb{P}$ [29, 30].

3.2 Deterministic settings
There are two sequential prediction approaches when the underlying source is assumed deterministic. The first is *curve fitting*, where a deterministic function $f(x)$ is assumed responsible for generating the observations. Curve fitting generally exploits statistical regression in the observation series. *Moving average* and *autoregressive* linear models (see Section 7) are commonly used for deterministic settings prediction. The second approach seeks a universal deterministic predictor. The predictor avoids over-fitting the predictor to a specific sequence, i.e in deterministic settings, a predictor that is applicable to different sets of sequences [20, 37]. The predictor class set $\mathbb{M}$ is a set of polynomials or code sequences. This construction avoids probabilistic assumptions about the observation source. However, when the designed predictor $\mathbb{Q}$ is constructed from the predictor set class $\mathbb{M}$, a prior probability distribution is often assumed [19, 30].

3.3 Loss function and regret
One step-ahead prediction commonly seeks the estimated state value at the next prediction slot $a = \hat{x}$. Alternatively, the action is set $a_t = p_t (x_t | x^{t-1})$ as a conditional probability assignment to measure the confidence in next

step prediction. Probabilistic prediction assignment provides more information about the state of the system compared to next event prediction. The loss in prediction is measured between the designed predictor's guess and the true value of x_t. Absolute, squared distance measures are common for next event prediction, while log distance is commonly used for probabilistic settings prediction. However, 0/1 loss function poses a challenge to several universal prediction algorithms including Bayesian mixture models [29, 30].

The predictor regret expresses the instantaneous loss due to choice of probability assignment $\mathbb{Q}$ rather than the true source $\mathbb{P}$. Subsequently, redundancy loss refers to the statistical expectation of regret for an observation sequence of length n [20, 27]. For example, if a source $\mathbb{Q}$ is used in place of $\mathbb{P}$, and a self information loss function is assumed $a_t = p_t (x_t | x^{t-1})$, $l(a_t, x_t) = -\log (p_t (x_t | x^{t-1}))$ then the redundancy loss limit to be achieved by an optimal predictor is the entropy rate of the source $\mathbf{H}(\mathbb{P})$ [20, 27]. In other words, no additional loss due to the use of $\mathbb{Q}$ [29, 30]. KL-divergence is commonly used to measure performance distance and can be defined by the *cross entropy* between $\mathbb{P}$ and $\mathbb{Q}$ as

$$d_t \left(x^{t-1}\right) := \log \frac{\mathbb{P}\left(x_t | x^{t-1}\right)}{\mathbb{Q}\left(x_t | x^{t-1}\right)}$$

$$D_n := \sum_{t=1}^{n} \mathbb{E}\{d_t\}$$

d_t is the instantaneous Kullback-Leibler (KL) divergence, and D_n is the total distance counterpart [20, 38]. Other possible choices for distance between $\mathbb{P}$ and $\mathbb{Q}$ are absolute, squared, Hellinger, and absolute divergence distances [28].

3.4 Bayesian methods for source assignment
Bayesian *mixture* models with self-information (entropy) loss were extensively studied in information and coding theory [28, 31, 34]. Bayesian algorithms are minimax optimal and are universal under self information loss functions [20, 23, 27]. They perform well under both probabilistic and deterministic non-stochastic settings [20, 27, 28, 38, 39]. Probability source assignment for $\mathbb{Q}$ is either *parametric* or *non-parametric*. The former assumes a single parametrised source $\{\mathbb{Q} = P_{\hat{\theta}}\}$ in the predictor set $\mathbb{M}$, while the later assumes $\mathbb{Q}_w$ as a mixture of sources with prior $\{w(\theta), \theta \in \mathbb{M}\}$ [27]. Mixture source assignment utilises a weighted linear sum of distributions $\{P_\theta, \theta \in \mathbb{M}\}$ with a prior distribution on the predictor index set $\mathbb{M}$ [20, 23]. Using a non-negative normalised weighting function $w(\theta)$. The mixture model density function is defined as

$$\mathbb{Q}_w \left(x^t\right) = \int_{\mathbb{M}} w(\theta) P_\theta \left(x^t\right) d\theta$$

Eltom *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:29

Page 5 of 17

The challenge in such models is the appropriate choice of the weights $w(\theta)$, i.e the prior distribution of the parameter $\theta \in \mathbb{M}$. Upper and lower loss bounds for Bayesian mixtures are defined using minimax and maximin approaches [20, 27]. Mixture models differ in terms of the size of the predictor index class C, stochastic class type P_θ, and mixture prior $w(\theta)$. Different mixture models can be grouped into the four approaches:

3.4.1 Plug-in approach

This approach can be considered as a mixture model with the number of mixtures $C = 1$. The underlying source is assumed to be a single parametrised by θ. The chosen predictor $\{P_{\hat{\theta}}\}$ probability function is created by estimating the value of $\hat{\theta}$ based on the series x_{t-1}. The parameter $\hat{\theta}_t = \hat{\theta}_t (x^{t-1})$ can be estimated using a maximum likelihood estimator [20, 24]. However, plug-in approaches are heuristic and lack theoretical justification [20, 23].

3.4.2 Finite mixture models

In finite mixture models, the replacement source $\mathbb{Q}_w$ is a sum of finite number of stochastic sources. The number of mixtures $C < \infty$ is generally decided beforehand based on the application objectives or through trial and error with different values of C. Prior distribution often set in advance (uninformative uniform distribution is common choice).

$$\mathbb{Q}_w\left(x^t\right) = \sum_{i=1}^{C} P_{\theta_i}(x^t)w(\theta_i), \quad \theta_i \in \{\theta_1, .., \theta_C\}$$

Expectation-maximisation (EM) algorithm is used to estimate the parameter set $\theta \in \mathbb{M}$ [24, 40, 41].

3.4.3 Kernel density estimation

Kernel density estimation places a *kernel*, i.e a function that satisfies probability density axioms on each observation sample. The samples are assumed independent and identically distributed. The stochastic source $\mathbb{Q}_w$ is defined as

$$\mathbb{Q}_w\left(x^n\right) = \frac{1}{nh} \sum_{t=1}^{n} K\left(x - x_t\right)$$

$h > 0$ is the smoothing parameter, and the kernel $K(., .)$ is a non-negative density function. Uniform, triangular, Epanechnikov, and normal kernels are some of common choices. [24, 40, 41].

3.4.4 Infinite mixture models

When the class $\mathbb{M}$ size is infinite, the prior distribution on θ is a smooth continuous function. The prior distribution is generally assumed drawn from a hyper-parametrised distribution, i.e. a probability distribution over probability distributions. A common non-parametric Bayesian method is the Dirichlet process $D(\alpha, G)$, where α is concentration parameter, and G is the distribution over $\theta \in \mathbb{M}$. Samples of θ_t at each time instant t are calculated iteratively from G using Monte-Carlo Markov chain methods. Infinite mixture model allows dynamic classification of data into clusters without having to specify the number of clusters in advance [40–42].

4 Review

The flow chart in Fig. 4 highlights the temporal sequence spectrum occupancy prediction process presented in this section. This section focuses on model selection, while
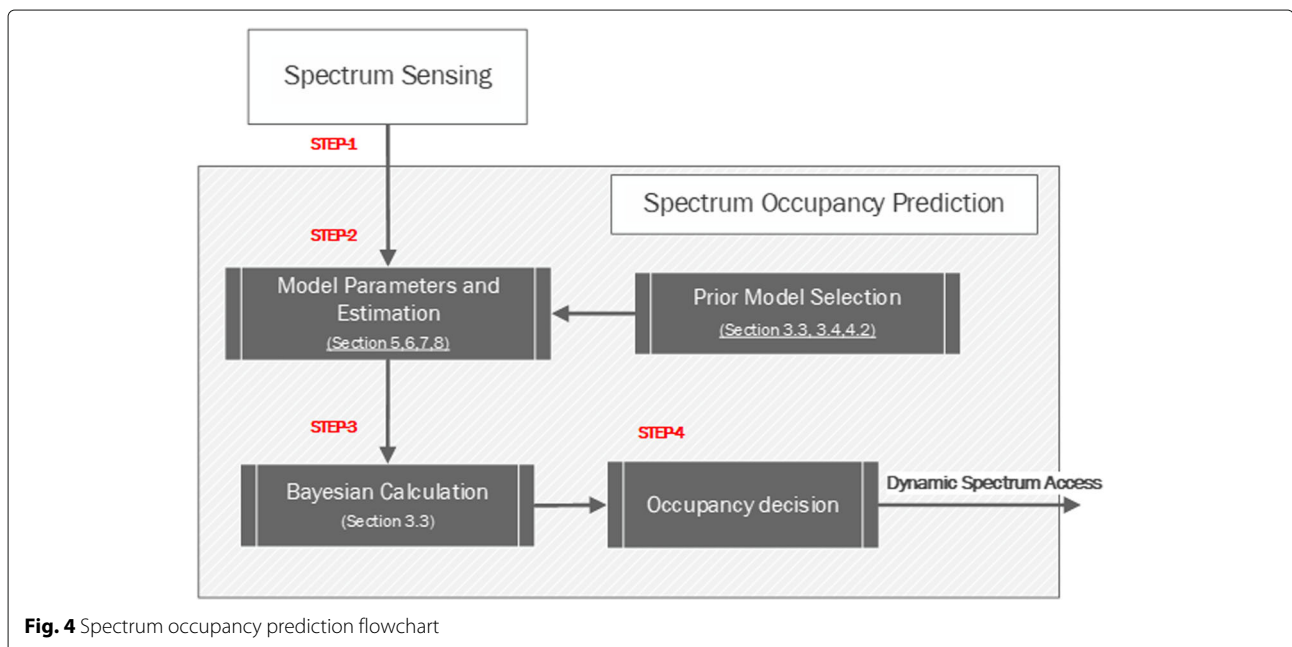

Fig. 4 Spectrum occupancy prediction flowchart

the next three sections address selected model classes. We present current spectrum occupancy prediction techniques using the statistical sequential prediction definition. Current spectrum occupancy research can be broadly divided into measurement campaigns and statistical occupancy modelling. Notably, spectrum measurements are often used to estimate the selected SOP model parameters. For the spectrum prediction either the measurements or the models can be used.

4.1 Spectrum measurement campaigns
A spectrum measurement campaign is an empirical data collection conducted for specific scenarios, e.g. indoor/outdoor, to collect spectrum occupancy samples on pre-selected frequency bands, e.g television white bands/cellular phones. Statistical analysis and estimation are conducted to generate an approximate statistical description of average power or channel occupancy. Though such modelling captures real-life spectrum occupancy scenarios, it is riddled with sampling inaccuracy, as well as spectral, spatial, and temporal dependency. However, the data collected in these measurement campaigns are utilised to infer a suitable class set $\mathbb{M}$ for the predictor design [7, 8, 43, 44]. Campaigns in Hong Kong in [44] and Melbourne [8] assessed spectrum occupancy patterns for a large section of radio spectrum. The survey by Chen and Oh [7] provides an intensive review of several measurement campaigns for selected wireless communication technologies.

Figure 3 presents raw spectrogram results of spectrum monitoring experiment conducted in three different urban environments in Melbourne metropolitan [8]. The spectrum campaign addressed spectral allocation for cognitive radio device-to-device communications and small cell networks. The spectrum occupancy is quantised by comparing the received signal level to an adaptive detection threshold based on the noise power. Raw samples collected over all frequency sweeps are shown for three urban environment class. The work results indicated that frequency range 402−460 MHz and 520−820 MHz (vacated analogue TV band) are suitable candidate for DSA applications [8].

4.2 Statistical occupancy modelling
Alternatively, statistical occupancy modelling estimates the observation generating mechanism often based on empirical samples. The scheme utilises a prior belief about the occupancy state and updates such belief as new observations are available. Given the estimated statistical model, spectrum occupancy prediction at future instances is achievable. Such models examine several statistical techniques with a major literature focus on Markov processes [10, 45], Poisson processes [12, 13], Bayesian models [9, 14], neural networks [5, 11, 46], linear regression

[15, 16], space vector machine [47], pattern mining [48, 49], and dictionary-based prediction [9]. In a sequential prediction framework, these techniques represent different parametrised predictor classes.

4.2.1 Prediction model selection
Parameters studied by spectrum occupancy modelling are (i) *channel status*, i.e prediction of the spectrum status as idle or busy, (ii) *duty cycle*, i.e. prediction of average fraction of time the spectrum channel is occupied, or (iii) *signal/power*, i.e. prediction of the power level on a specific channel. These occupancy series are modelled based on assumptions about their state space, loss function, and predictor action. For instance, channel status observation series can be modelled as an ON/OFF (2-state model) binary source model $\mathcal{X} = [0, 1]$, or more (e.g. 3-state model). Similarly, the predictor action a_t is commonly modelled as one-step ahead state prediction, i.e. $a_t = \hat{x}_t$ or as a probabilistic assignment, i.e. $a_t = p\left(x_t|x^{t-1}\right)$. Common choices for loss functions are self information, 0/1 loss and mean square error, while regret and redundancy often adopt KL-divergence. However, the loss function in each proposal is often formulated based on the intended application (e.g. throughput, sensing accuracy, or hand-off success rate). Performance comparison metrics such as secondary user's throughput, spectrum interference and wastage, and probability of error (or mean square error) are generally defined based on the probability density of the one step-ahead prediction, as well as, the prediction loss function. For example, the probability of incorrect prediction of an available spectrum hole generally describe spectrum interference or spectrum wastage [50].

Consequently, spectrum occupancy prediction modelling is essentially the selection of a class $\mathbb{M}$ of predictors or the mixture of sources from class M. The choice of the predictor class is limited by the application requirements and constraints. For example, a set of finite k*th*-order Markov models is more practical for the predictor design than the set of all arbitrary order Markov models, due to the set size. Moreover, HMM model is suitable for finite state occupancy models one step-ahead prediction given the errors in the wireless channel, while Kalman filter is a more suitable for infinite state space scenarios. Kernel density estimation is rarely proposed for on-line prediction, but can be used to construct the probability density of selected predictor class. Ultimately, the sequential predictor performance depends on the predictor set $\mathbb{M}$ "complexity" or richness, which quantifies the class type, size, and statistical regression between observations [19, 20, 23, 29].

Table 1 provides a summary of the current techniques used for spectrum prediction in dynamic spectrum access systems. The fourth column in the table presents the sample space for the observation series. Finite sets (e.g.

Table 1 Summary of spectrum prediction technique

Category	Model	Research works	State space/state dependency	Occupancy decision criteria
Memory less stochastic source models (Section 5)	Bernoulli/ binomial	[61–63, 67, 70, 102]	$x \in [0, 1, ..., S] / p(x_t)$	Channel status
	Poisson	[58, 59, 64–66, 110]		
	Exponential	[9, 58, 59, 67]	$x \in \mathcal{R} / p(x_t)$	Duty cycle
	Log-normal	[43, 68]		Signal/power
	Uniform	[69]	$x \in [0, 1, ..., S], x \in \mathcal{R} / p(x_t)$	
Finite order Markov models (Section 6)	2-state Markov chain	[58, 79–82]	$x \in [0, 1] / p(x_t) p(x_{t-1})$	Channel status
	3-state Markov chain	[59]	$x \in [0, 1, 2] / p(x_t) p(x_{t-1})$	
	High-order Markov chain	[84]	$x \in [0, 1] / p(x_t | x^{t-m}), m > 1$	
	Semi-Markov	[85, 86]	$x \in [0, 1, ...] / p(x_{s+r}, s > 0 | x_t) s \in \mathcal{R}$	Duty cycle
	Continuous time MC	[87, 88, 94]		
	Hidden Markov model	[9, 14, 89–92, 95, 96]	$x \in [0, 1, ..., S] / p(x_t | x_{t-1})$	Channel status
	Bayesian models	[97, 98, 105, 111]	$x \in \mathcal{R} / p(x_t | x_{t-1})$	Signal/power
Finite order linear regression models (Section 7)	Autoregressive	[99–101, 103]	$x \in [0, 1]$ or $x \in \mathcal{R} / p(x_t | x_{t-1}, ..., x_{t-m})$	Channel status or signal/power
	Moving average	[100]		
	ARIMA	[66]		
	Random walk	[102]		
Machine learning statistical based techniques (Subsection 4.3)	Neural networks	[51–53, 56, 57, 112]	$x \in [0, 1]$ or $x \in \mathcal{R} / p(x_t | x^{t-1})$	Duty cycle, channel status, or signal/power
	Support vector machine	[47]		
	Pattern mining	[48, 49]		

ON/OFF) or infinite set (e.g. real space $\mathcal{R}$) are presented. Additionally, state regression and dependency on previous events (e.g. first order Markov chain) are presented. Finally, occupancy series are displayed in the last column.

4.2.2 Prediction models classification

By dissociating the implementation requirements and assumptions from the stochastic components of the spectrum prediction model, the authors distinguish three major categories of parametrised predictor classes used in literature:

1. Memoryless stochastic sources classes (single source). This category contains a diverse set of parametrised sources including Bernoulli, Binomial, Poisson, exponential, uniform, and normal distributions. Such models are better suited for traffic such as *internet of things*, telemetry, and applications that use radio spectrum.
2. Finite order Markov chain class (finite source memory). The dominant choice is first order Markov chain with finite/infinite state space such as hidden Markov model, Kalman filters, and particle filters. These models are better suited for applications such as TCP/IP traffic.
3. Finite order linear regression source class. Autoregressive (AR) and moving-average (MA) models along with ARMA, and ARIMA models assume linear regression in the observation series. This set of models is also suitable for TCP/IP traffic, with the advantage of low complexity implementation.
4. Machine learning-based techniques including neural networks, support vector machines, and pattern mining can be used for massive access network scenarios.

Table 2 highlights few major advantages and disadvantages of different spectrum occupancy prediction categories. For example, stochastic memoryless modes ignores temporal correlation of the data, but suitable for low complexity single PU sparse channel usage scenarios. Similarly, finite Markov models are suitable for heavy-tail channel usage scenarios such as multimedia transfer. Markov-Bayesian mixtures can be used to model scenarios with multiple primary and secondary users. Finally, linear regression models exploit further past measurements with less complexity compared to finite state Markov models.

Figure 5 summarizes the sequential prediction theory presented in Section 3 and maps current SOP techniques. The number of mixture sources C in the replacement source assignment Q differentiates mixture models (Subsection 3.4). The figure conceptually illustrates the

modelled occupancy series as an input, where the selected mixture model produces the desired performance measure based on the selected loss function. Model classification presented in this section is displayed under mixture model framework. We present a review of current spectrum prediction techniques for each category in the next three sections. Section 5 presents single memoryless source approaches, Section 6 handles Markov-based models, while Section 7 presents linear statistical regression based prediction.

4.3 Machine learning-based techniques

Machine learning, data mining, and pattern recognition algorithms are based on existing statistical inference models. Kobayashi et al. [[24], Chapter 21] discusses the statistical aspects of machine learning. Several classification and prediction techniques are a numerical methods based on a statistical prediction model. For example, *artificial neural networks* and *HMM* are numerical solutions of Bayesian /Markov models (particularly particle filer solutions). Similarly, support vector machine are numerical solutions of linear regression models.

Artificial intelligence and machine learning in spectrum prediction generally address the *learning* of predictor class parameters. The methods improve likelihood estimation for spectrum prediction problems with large sample size. For example, neural network genetic algorithms can be used for maximum likelihood estimation of HMM parameters [51]. *Neural networks*-based techniques are presented extensively in cognitive radio networks [18, 51–55], with application on spectrum prediction presented in [56, 57]. Support vector machines [47], pattern mining [48, 49], and dictionary-based prediction [9] were suggested for spectrum prediction and user activity modelling. The surveys in [17, 18, 54] discuss artificial intelligence and machine learning applications for dynamic spectrum access.

5 Spectrum occupancy prediction with memoryless stochastic source models

In this category, the observations are assumed independent and identically distributed (i.i.d) random variables drawn from a single parametrised stochastic source. The series x^{t-1} has no conditional dependency on the prediction of $\hat{x}_t$, i.e. models fall under this category are *memoryless*. Practically, one-step ahead prediction is not possible with such models. Thus, it is often combined with time correlated assumptions (e.g. Poisson Markov chain [58, 59]) or used to estimate the stochastic source probability density function $\mathbb{Q}_w$ from a training sequence. Models adopted in SOP proposals include the following:

Bernoulli trial process is the mathematical abstraction of repeated coin tossing. The random variable x_t takes

Eltom *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:29

Page 9 of 17

Table 2 Comparison of spectrum prediction categories

Category	Advantages	Disadvantages
Memoryless stochastic source models (Section 5)	Low complexity	Limited to sparse spectrum usage
	Closed form solution for sparse spectrum usage scenarios	Limited to single primary user scenarios
	Easier/convenient model to adopt	May not describe real world channel occupancy status
Finite order Markov models (Section 6)	Expandable to various PU/SU scenarios	Require sufficient measurements for model training
	Applicable to heavy-tailed channel traffic	Complexity depends on the order of the model
	Higher accuracy with manageable number of parameters	
Finite order linear regression models (Section 7)	Expandable to various PU/SU scenarios	Expandability increases the number of model parameters
	Approximation of probabilistic model to linear equation model	Require sufficient measurements for model training

only the values 0 or 1 representing failure and success, respectively. The series $x_1, x_2 \ldots x_{t-1}$ is assumed to be independent and identically distributed Bernoulli random variables, with probability mass function parametrised by ρ [24, 60]:

$$f(k;\rho) = \rho^k(1-\rho)^{(1-k)}, \quad k \in \{0, 1\}$$

Where k is the number of trials, and ρ is the probability a certain outcome, e.g. $\rho = p(X_t = 1)$.

Binomial distribution models the probability of exactly k success in n trials, yielding the probability mass function parametrised by ρ as

$$f(k;\rho) = C_k^n \rho^k (1-\rho)^{(n-k)}, \quad C_k^n = \frac{n!}{k!\,(n-k)!}$$

Poisson distribution describes the probability of a number of k events in a time period with a constant average rate $\lambda = \frac{k}{n}$ [24, 60]:

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \{0, 1, \ldots\}$$

Exponential distribution. The interval between events in a Poisson distributed process follows the negative exponential distribution parametrised by λ, with probability density function [24, 60]:

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0.$$

In spectrum occupancy literature, memoryless sources are not often used for one-step ahead prediction. However, this class of stochastic sources is frequently used to describe primary user activity. Bernoulli process have
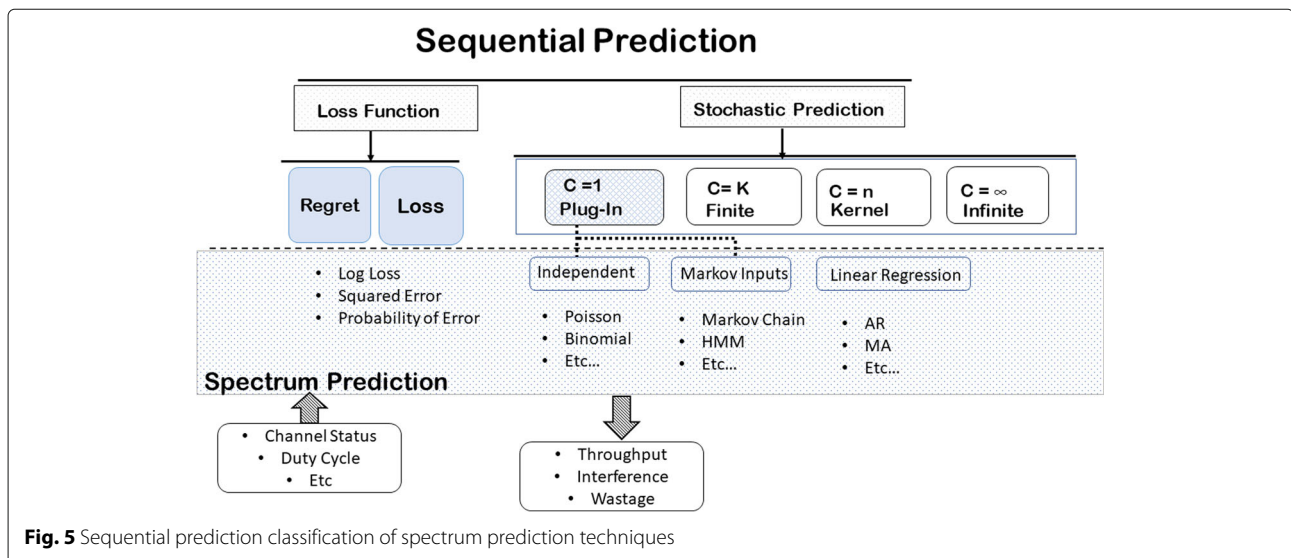

Fig. 5 Sequential prediction classification of spectrum prediction techniques

Eltom *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:29

Page 10 of 17

been proposed in [61–63] to describe ON/OFF spectrum occupancy in spectrum sensing/access proposals. Similarly, Poisson process have been proposed in [58, 64] (2 states) and [59] (3 states) to model the arrival/departure process of the primary user. Exponentially distributed duty cycle models were presented based on queuing theory in [58, 59]. Similarly, proposals in [65, 66] suggested a non-exponential service time as a result for multiple primary users scheduling. In [9], an exponential distribution to model inter-arrival time of the primary users was proposed to design a secondary contention algorithm. Joint cognitive radio spectrum sensing and prediction model in [67] proposed an exponential primary user prediction and estimated spectrum opportunity wastage and interference. Other primary user modelling efforts utilised an identical approaches with i.i.d events, but employed different probability distributions such as log-normal distribution [43, 68], uniform distribution [69], and binomial distribution [67, 70]. The choices were generally motivated by physical layer assumptions.

6 Spectrum occupancy prediction with finite order Markov models

Various Bayesian-based techniques utilise different assumptions about the observation sample space, the statistical regression, and the underlying stochastic process. The case when the probability of current event x_t only depends on the probability of previous event x_{t-1}, i.e. $p\left(x_t|x^{t-1}\right) = p(x_t|x_{t-1})$ is called *Markov property* [24]. Markov-based construction is attractive due to the desirable convergence properties of Markov chain-based models [71–74]. *Markov chain* and *partially observable Markov* models are commonly used for spectrum occupancy modelling. Markov processes also include *semi-Markov* processes such *M-order Markov chain* with dependence on m previous events, i.e. x^{t-m}, or *explicit duration* Markov chain (a form of continuous-time Markov chain), where the time spent on each state is not exponentially distributed [24, 75]. The main difference between proposals is the number of states assumed by different models and the proposal's loss function.

Bayesian Markov model General Markov-based model in estimation theory utilises a Bayesian model framework as [76]:

$$p\left(x_t|y^{t-1}\right) = \int_{\mathcal{X}} p(x_t|x_{t-1})p\left(x_{t-1}|y^{t-1}\right) dx_{t-1}$$

$$p(x_t|y^t) = \frac{p(y_t|x_t)p\left(x_t|y^{t-1}\right)}{p\left(y_t|y^{t-1}\right)}$$

$$p\left(y_t|y^{t-1}\right) = \int_{\mathcal{X}} p(y_t|x_t)p\left(x_t|y^{t-1}\right) dx_t$$

The first equation is Chapman-Kolomogrov prediction equation, the second is Bayes rule update, while the last equation is normalisation factor [76]. This model is labelled *doubly stochastic* as it accounts for measurement error in x^{t-1} by defining the observation series y^{t-1}, while x^{t-1} is defined as the latent variable series. The *latent state model* is defined by the non-linear function $[x_t = \mathbf{f}_t(x_{t-1}, \mathbf{v}_t)]$, and $\mathbf{v}_t$ an independent additive noise source. x_t is distributed based on the probability $p(x_t|x_{t-1})$ defined as latent state Markov prior. The observations are defined as the dependent variable $\left[y_t = \mathbf{h}_t(x_t, \mathbf{u}_t)\right]$, where $\mathbf{h}_t$ is a non-linear function, and $\mathbf{u}_t$ is an independent additive noise source (measurement error) [24]. The observation variable is distributed according to $p(y_t|x_t)$, defined as the observation likelihood probability. The conditional posterior probability $p\left(x_t|y^{t-1}\right)$ is recursively calculated from the prior and likelihood probabilities from an initial state distribution $p(x_0)$. The equation set simplifies the probability assignment in the form $p\left(x_t|y^{t-1}\right) = p\left(x_t|x_{t-1}, y^{t-1}\right) p\left(x_{t-1}|y^{t-1}\right)$ (Markov property). When implementing such model, the density $p\left(x_t|y^{t-1}\right)$ is either estimated using the prior/likelihood function or using kernel density estimation [36, 76].

Markov chain process is the simplest Bayesian Markov model. It is assumed to be fully observable, and finite. Markov chain process is parametrised by *transition probability* and initial state distribution. Each element in the transition matrix is the probability p_t^{ij}, i.e. the probability of being in state j at time t given the system is currently in state i at time $t-1$ [24, 76–78].

Hidden Markov model (HMM) is partially observable Markov chains, i.e. observing a Markov chain through a noisy channel [24, 75]. HMM employs two finite sample sets for latent variables $\mathcal{X}$ and observations $\mathcal{Y}$. The additional conditional probability of a system is at state i ($x_t = i$) to emit an observation ($y_t = j$) is referred to as e_{ij} or the *emission* probability. Figure 6 displays a snapshot of HMM state transition (connected lines).

Kalman filter is the optimal solution for linear Gaussian state space Markov-based models [24, 36, 76]. Non-linear predictors are often a sub-optimal variation of Kalman filter, such as extended Kalman filter, and unscented Kalman filter [24, 35, 36].

Bayesian particle filters Particle filter methods utilise Monte-Carlo Markov chain (MCMC) to approximate the conditional posterior probability assignment $p\left(x_t|y^{t-1}\right)$ or the full posterior probability $p\left(x^t|y^{t-1}\right)$. They utilise either weighted samples of a plug-in probability assignment based on prior/likelihood or a mixture model based density [76].

Fig. 6 Markov statistical prediction in hidden Markov model
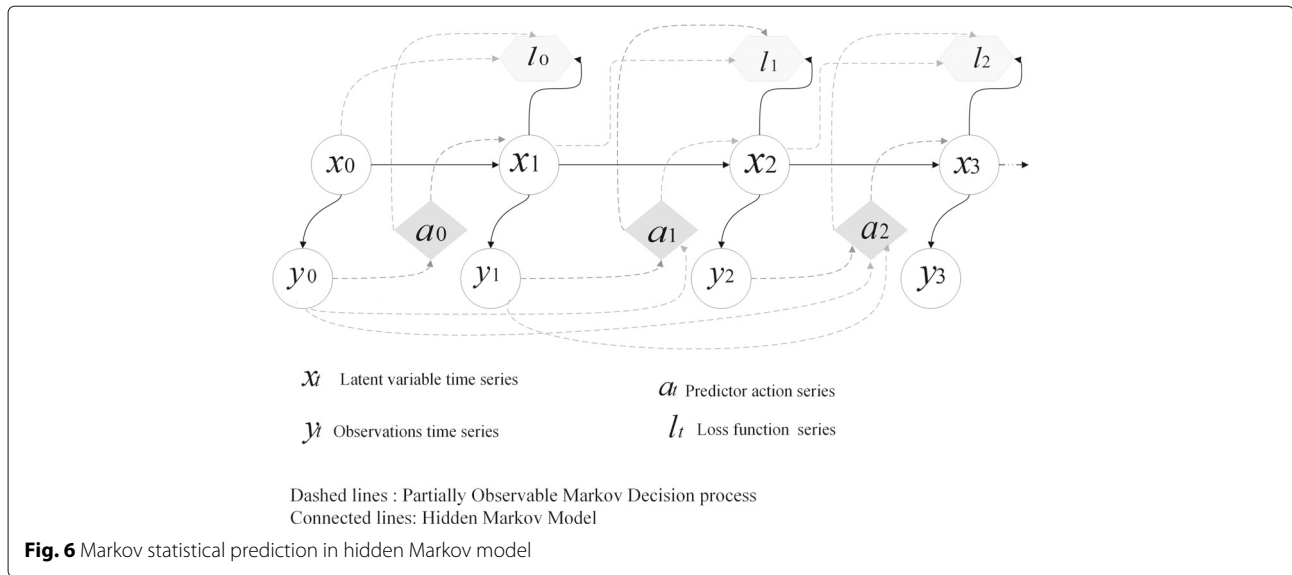
In the spectrum modelling literature, Poisson Markov chain-based proposals in [58, 59] studied primary user interference and wastage. Two-state [79–83] and three-state discrete-time Markov [59] chain have been proposed to model the primary-secondary users stochastic behaviour. Similarly, higher-order Markov chains in [84] were used to detect the primary user traffic pattern. Explicit duration semi-Markov chains with generalised distribution of duty cycle time modelled primary user's inter-arrival time in [85, 86], while continuous time Markov chain modelled primary user behaviour in [87, 88]. Moreover, hidden Markov model received wide attention in spectrum occupancy prediction literature [9, 14, 83, 89–93]. Liu et al. addressed the prediction confidence, and the error of a continuous time Markov chain model with Erlang-2 distribution model for primary user's activity [94]. K-step ahead prediction was studied in [95, 96] assuming a non-stationary HMM. Finally, works in [97, 98] utilised regularised particle filters with Kernel density estimation to model primary user activity in multi-primary and secondary user cases.

7 Spectrum occupancy prediction with finite order linear regression models

A special case of the general non-linear statistical regression for $p\left(x_t|x^{t-1}\right)$, linear regression models focus on the linear dependency between the random variables x_t, and x^{t-1} [24]. Autoregressive model AR ($p = 0$), and moving average (MA) ($q = 0$) are special cases of autoregressive moving-average. ARMA model $ARMA(p, q)$ can be written as [24, 60]

$$x_t = c + \eta_t + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \eta_{t-i}$$

Where c is a constant that can be replaced with $\mu = \mathbb{E}_x\{x_t\}$. η_t is a noise random variable that represents the uncertainty in sampling. ϕ_i, θ_i are the autoregressive, and moving average parameters. p, q are the order of the autoregressive and moving average components. Autoregressive integrated moving-average (ARIMA) process generalises the ARMA model to $ARIMA(p, d, q)$ and written as [24, 60]

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) (1 - L)^d x_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \eta_t$$

Where $L^i(x_t) = x_{t-i}$ is the time lag operator, $\Delta x_t = x_t - x_{t-1} = (1-L)x_t$ is the difference operator, and $\Delta^d x_t = (1 - L)^d x_t$ is the generalised difference operator. Setting the differencing degree $d = 0$ in ARIMA model will result in ARMA model, while setting $p = q = 0$, $d = 1$ results in a random walk model. ARMA and ARMIA assume no specific underlying stochastic process, but provides the regression between observation samples.

An autoregressive with Gaussian distributed random variables was used to model spectrum occupancy in [99–101]. Similarly, moving-average [100] and ARIMA [66] were proposed for spectrum occupancy status modelling. Random walk model was proposed in [102] to model spectrum occupancy duty cycle. Finally, an autoregressive model of decimal equivalent of a binary series model was proposed for primary user activity in [103].

8 Cooperative spectrum prediction

Spectrum prediction in single secondary user environment is *local spectrum prediction*. Consequently, *cooperative spectrum prediction* in multi-user environment was proposed to improve the collective accuracy of spectrum occupancy prediction [98]. The term *homogeneous*

Eltom *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:29

Page 12 of 17

cooperative prediction refers to the case when secondary users have identical detection performance in terms of channel conditions, e.g. signal to noise ratio. While *heterogeneous* cooperative prediction refers to the general case of non-identical secondary user detection performance. The latter scenario incorporates additional dependency on the spatial distribution [93]. Cooperative prediction fusion proposals are commonly classified into hard and soft fusion techniques.

8.1 Hard prediction fusion

For R cooperative users with a binary observation series $d_r \in \{0, 1\}$, the predicted occupancy state at each cooperative user r at time instance t is defined as $d_r = \hat{x}_r(t)$. The cooperative decision $D_R(t)$ is M-out of-N rule written as [83]

$$D_R = \begin{cases} 1 \ (\text{Busy}) & \mu \geq M \\ 0 \ (\text{Idle}) & \mu < M \end{cases}$$

$$\mu = \sum_{i=1}^{R} 1\left[d_r\right]$$

Where $1[..]$ is the indicator function. The three main rules for threshold M are: $M = 1$ the logical OR, $M = N$ the logical AND, and $M = N/2$ is the majority decision rule.

8.2 Soft prediction fusion

If the data shared by each cooperative user r at time instance t is defined as $d_r = a_r(t) = p_r\left(x_t|y^{t-1}\right)$, or $d_r = \hat{x}_k(r)$. Then, the total cooperative decision $D_R(t)$ at each time instant t can be defined as a finite weighted sum of each user's data. The soft fusion prediction D_R can be defined as

$$D_R = \sum_{k=1}^{R} w_r d_r, \quad w_r \in \{w_1, \ldots, w_R\}$$

$$d_r = p_r\left(x_t|y^{t-1}\right), \quad r \in [1, .., R]$$

Where w_r is a prior distribution on the R cooperative user. Soft fusion-based formulation on the predictive posterior probability $p_r(x_t|y_{r,1:t-1})$ can be modelled as a linear mixture model. Let a non-negative normalised weighting function be $w(\theta)$ parametrised by θ, the mixture model is defined as

$$p_w\left(x_t|y_{1:R,1:t-1}\right) = \int_{\theta} p_r\left(x_t|y_{r,1:t-1}, \theta\right) w(\theta) d\theta$$

For a finite number of user R the mixture sum is

$$p_w\left(x_t|y_{1:R,1:t-1}\right) = \sum_{r=1}^{R} w_r(\theta) \times p_r\left(x_t|y_{r,1:t-1}, \theta\right)$$

A diverse collection of techniques can be adopted for the prior w_θ selection. Equal gain, maximal ratio, and selection combining fusion are the most common of these soft fusion techniques.

8.2.1 Equal gain fusion

Equal gain combining assumes all secondary users have an equal "weight", i.e. $w_r = \frac{1}{R}, r \in \{1, 2, .., R\}$, i.e. $\theta = R$. This fusion strategy ignores the heterogeneous nature of secondary user's detection/prediction performance, as well as their spatial distribution.

$$p_w\left(x_t|y_{1:R,1:t-1}\right) = \frac{1}{R} \sum_{r=1}^{R} p_r\left(x_t|y_{r,1:t-1}\right)$$

8.2.2 Selection combining

Selection combining uses the decision of the user with the best channel condition, e.g. signal to noise ratio ρ. The method overrules decisions made by all other cooperative user and uses the best user decisions [30].

$$p_w\left(x_t|y_{1:R,1:t-1}\right) = \max_{\rho_r} p_r\left(x_t|y_{r,0:t}, \rho_r\right)$$

8.2.3 Maximal ratio fusion

Maximal ratio combining utilise SU signal to noise ratio ρ_r in the prior $w(\theta)$, i.e. $\theta = \rho_r$:

$$w_r(\rho_r) = \frac{\rho_r}{\sum_{r=1}^{R} \rho_r}$$

Cooperative fusion of channel access decisions has been studied extensively in spectrum sensing using hard and soft fusion techniques [104]. But a limited number of studies focused on applications for one step ahead prediction. In our previous work on cooperative prediction [93], we extended prediction error performance analysis for HMM predictors [83]. In [98], cooperative prediction was proposed as a coalition game. Similarly, the studies in [67] proposed a majority hard fusion based cooperative prediction for binomially distributed predictions. Finally, Saad et al. [105] proposed a beta distribution prior for a linear Gaussian kernel density estimated PU activity. The study presented a trade-off between communication cost and prediction accuracy.

9 Spectrum occupancy prediction challenges

The survey in [6] discussed the issue of occupancy modelling validity based on the type and amount of traffic pattern. The survey presented several scenarios of possible implementation issues for primary user modelling. This section extends the survey, and presents theoretical challenges for SOP models.

9.1 Validity and complexity

Spectrum occupancy observation representation is limited by state space dimensionality. Spectrum samples

Eltom *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:29

Page 13 of 17

have temporal, spectral, and spatial dependency. Proposed spectrum occupancy prediction models simplify assumption about spectral and spatial assumptions to avoid model complexity. To our knowledge, there are no multi-dimensional proposals for spectrum occupancy prediction. Moreover, the validity of any chosen model is generally questionable from dimensionality and universality perspective, as any assumption about the underlying observation process may not fit the actual behaviour. Few spectrum measurement campaigns invalidated several short-term prediction assumptions. Thus, validation through empirical spectrum campaigns is essential for any spectrum predictor design [8, 10]. For example in [106], the popular i.i.d exponential duty cycle assumption is criticised as a model for short-term prediction. A Pareto distribution was proposed for long-term prediction, but short-term prediction was deemed application dependent, and technology specific.

Moreover, common challenges in sequential prediction theory are model over-fitting, and redundancy loss convergence guarantee. Model over-fitting refers to the case when a model is too complex, that renders it sensitive to small changes in observation statistics [19, 20, 23, 29]. Model complexity constraints the applicability of the prediction mode. The complexity of a specific class of predictors, i.e. class size and statistical regression affects the predictor convergence guarantee to the desired redundancy loss bound (see redundancy-capacity theorem [20, 23, 28, 31]). Plug-in approaches simplify predictor design complexity using assumptions about the observation generating mechanism to achieve optimal predictor design. For example, a set of finite k*th*-order Markov models are more practical for predictor design compared to the set of all arbitrary order Markov models. Moreover, mixture models are more complex but allow empirical measurements-based source estimation. An example would be Dirichlet mixture process which often used to generate mixture prior distributions, but tracing convergence bounds becomes increasingly difficult for non-Gaussian mixtures for example [20, 28, 40]. Convergence bounds are calculated only for limited Bayesian mixture class/prior distribution pairs (for example, uniform prior/Epanchinkov kernel) [107].

9.2 Cooperation and contention
Cooperative spectrum prediction faces the practical issue of common control channel design [97]. The amount of data shared between users sets a trade-off between spectrum prediction accuracy and control channel capacity [97]. Common control design for cooperative spectrum prediction in a multi-primary, user's environment is yet to develop in the spectrum prediction literature. Analysis of cooperative prediction using hierarchical Dirichlet processes is an interesting proposal to model cooperative spectrum prediction, that is not explored in SOP literature [40].

Contention policy proposals for DSA systems are still under development in current literature. In single user case, reinforcement learning is suggested in some literature sources to model the spectrum occupancy [50, 108]. However, the study in [108] questions reinforcement learning as useful tool to improve spectrum occupancy modelling of their own spectrum campaign measurements. Multi-user game theory-based approaches are interesting candidates for multi-user spectrum prediction.

10 Conclusions
In this paper, we presented a comprehensive survey and classification of spectrum occupancy prediction (SOP) based on theoretical sequential prediction framework. To the best of authors' knowledge, this review on spectrum occupancy prediction in literature is the first to consolidate current techniques based on sequential prediction theoretical framework. This classification approach highlights candidate techniques suitable for SOP scenarios not extensively covered in current literature. In the paper, we presented the definition and fundamentals of statistical sequential prediction. Then, we addressed predictor loss, regret, and Bayesian methods for underlying stochastic source assignment. Based on parametric and non-parametric mixture model framework, this paper classifies spectrum occupancy modelling approaches in literature based on predictor class selection. Predictor class selection categories of memoryless sources, Markov models, and linear regression models along with machine-based techniques were detailed based on current SOP literature proposals. SOP cooperative prediction based on hard and soft fusion techniques was discussed for multi-user scenarios. Finally, spectrum predication theoretical and practical challenges were presented and highlighted candidate techniques.

Endnotes
[1] Probabilistic assumptions are made about the $\mathbb{M}$ sources prior, and under probabilistic action a_t assumptions see [19, 29, 30].

[2] The major cases are 0/1 loss function for probabilistic action a_t and 0/1 loss for ON/OFF non-stochastic observations, see [19, 29, 30] for analysis and [29, 30, 109] for Starkov codes, Hedge algorithm, and game theory approaches for sequential prediction.

Abbreviations
AR: Autoregressive model; ARIMA: Autoregressive integrated moving average; ARMA: Autoregressive integrated moving average; DSA: Dynamic spectrum access; HMM: Hidden Markov model; MA: Moving average model PU: Primary user; SU: Secondary user; SOP: Spectrum occupancy prediction

Authors' contributions
HE contributed towards Section 1 to 10. A/Prof. SK contributed to Section 2. Prof. RJE contributed to Sections 3 and 6. Prof. YCL contributed to Sections 4, 8, and 9. Dr. BR contributed to Sections 2 and 3. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]School of Engineering, RMIT University, 124 La Trobe St, 3000 Melbourne, Australia. [2]Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Melbourne, 3010 Australia. [3]University of Electronic Science and Technology of China (UESTC), Chengdu 611731, China.

References
1. IF Akyildiz, W-Y Lee, MC Vuran, S Mohanty, NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey. Comput. Netw. **50**(13), 2127–2159 (2006). http://dx.doi.org/10.1016/j.comnet.2006.05.001
2. C Baylis, M Fellows, L Cohen, RJM II, Solving the spectrum crisis: intelligent, reconfigurable microwave transmitter amplifiers for cognitive radar. IEEE Microw. Mag. **15**(5), 94–107 (2014). https://doi.org/10.1109/mmm.2014.2321253
3. YC Liang, KC Chen, GY Li, P Mahonen, Cognitive radio networking and communications: an overview. IEEE Trans. Veh. Technol. **60**(7), 3386–3407 (2011). https://doi.org/10.1109/TVT.2011.2158673
4. IF Akyildiz, W-Y Lee, MC Vuran, S Mohanty, A survey on spectrum management in cognitive radio networks. IEEE Commun. Mag. **46**(4), 40–48 (2008). https://doi.org/10.1109/mcom.2008.4481339
5. X Xing, T Jing, W Cheng, Y Huo, X Cheng, Spectrum prediction in cognitive radio networks. IEEE Wirel. Commun. **20**(2), 90–96 (2013). https://doi.org/10.1109/mwc.2013.6507399
6. Y Saleem, MH Rehmani, Primary radio user activity models for cognitive radio networks: a survey. J. Netw. Comput. Appl. **43**, 1–16 (2014). https://doi.org/10.1016/j.jnca.2014.04.001
7. Y Chen, H-S Oh, A survey of measurement-based spectrum occupancy modeling for cognitive radios. IEEE Commun. Surv. Tutorials. **18**(1), 848–859 (2016). https://doi.org/10.1109/comst.2014.2364316
8. A Al-Hourani, V Trajkovic, S Chandrasekharan, S Kandeepan, Spectrum occupancy measurements for different urban environments. 2015 Eur. Conf. Netw. Commun. (EuCNC) (2015). https://doi.org/10.1109/eucnc.2015.7194048
9. SJ Kim, GB Giannakis, in *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. Dynamic learning for cognitive radio sensing (IEEE, Piscataway, 2013), pp. 388–391. https://doi.org/10.1109/CAMSAP.2013.6714089
10. Z Chen, N Guo, Z Hu, RC Qiu, Experimental validation of channel state prediction considering delays in practical cognitive radio. IEEE Trans. Veh. Technol. **60**(4), 1314–1325 (2011). https://doi.org/10.1109/TVT.2011.2116051
11. L Yin, S Yin, W Hong, S Li, in *Military Communications Conference (MILCOM)*. Spectrum behaviour learning in cognitive radio based on artificial neural network (IEEE, Piscataway, 2011), pp. 25–30. https://doi.org/10.1109/MILCOM.2011.6127671
12. J Lee, HK Park, Channel prediction-based channel aladdress scheme for multichannel cognitive radio networks. J. Commun. Netw. **16**(2), 209–216 (2014). https://doi.org/10.1109/jcn.2014.000032
13. W Pu, IF Akyildiz, Asymptotic queuing analysis for dynamic spectrum access networks in the presence of heavy tails. IEEE J. Sel. Areas Commun. **31**(3), 514–522 (2013). https://doi.org/10.1109/JSAC.2013.130316
14. X Li, SA Zekavat, Cognitive radio based spectrum sharing: evaluating channel availability via traffic pattern prediction. J. Commun. Netw. **11**(2), 104–114 (2009). https://doi.org/10.1109/JCN.2009.6391385
15. VK Tumuluru, P Wang, D Niyato, Channel status prediction for cognitive radio networks. Wirel. Commun. Mob. Comput. **12**(10), 862–874 (2012). https://doi.org/10.1002/wcm.1017
16. S Chen, L Tong, Maximum throughput region of multiuser cognitive access of continuous time Markovian channels. IEEE J. Sel. Areas Commun. **29**(10), 1959–1969 (2011). https://doi.org/10.1109/JSAC.2011.111206
17. M Bkassiny, Y Li, SK Jayaweera, A survey on machine-learning techniques in cognitive radios. IEEE Commun. Surv. Tutorials. **15**(3), 1136–1159 (2013). https://doi.org/10.1109/surv.2012.100412.00017
18. A He, KK Bae, TR Newman, J Gaeddert, K Kim, R Menon, L Morales-Tirado, JJ Neel, Y Zhao, JH Reed, WH Tranter, A survey of artificial intelligence for cognitive radios. IEEE Trans. Veh. Technol. **59**(4), 1578–1592 (2010). https://doi.org/10.1109/tvt.2010.2043968
19. N Cesa-Bianchi, G Lugosi, *Prediction, Learning, and Games*. (Cambridge University Press, New York, 2006). https://doi.org/10.1017/cbo9780511546921
20. N Merhav, M Feder, Universal prediction. IEEE Trans. Inf. Theory. **44**(6), 2124–2147 (1998). https://doi.org/10.1109/18.720534
21. H Bolfarine, S Zacks, *Prediction theory for finite populations, Springer Series in Statistics*. (Springer, New York, 1992). https://doi.org/10.1007/978-1-4612-2904-9
22. CE Shannon, Prediction and entropy of printed english. Bell Syst. Tech. J. **30**(1), 50–64 (1951). https://doi.org/10.1002/j.1538-7305.1951.tb01366.x
23. J Rissanen, Universal coding, information, prediction, and estimation. IEEE Trans. Inf. Theory. **30**(4), 629–636 (1984). https://doi.org/10.1109/tit.1984.1056936
24. H Kobayashi, BL Mark, W Turin, *Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance*. (Cambridge University Press, New York, 2011)
25. J Ziv, A Lempel, A universal algorithm for sequential data compression. IEEE Trans. Inf. Theory. **23**(3), 337–343 (1977). https://doi.org/10.1109/tit.1977.1055714
26. JL KELLY, A new interpretation of information rate. IRE Trans. Inf. Theory. **2**(3), 25–34 (2011)
27. Kotł, W,owski, Grü,nwald, in *IEEE Information Theory Workshop (ITW), 2012*. Sequential normalized maximum likelihood in log-loss prediction (IEEE, Piscataway, 2012), pp. 547–551. https://doi.org/10.1109/ITW.2012.6404734
28. M Hutter, Convergence and loss bounds for bayesian sequence prediction. IEEE Trans. Inf. Theory. **49**(8), 2061–2067 (2003). https://doi.org/10.1109/tit.2003.814488
29. G Shafer, V Vovk, *Probability and finance: it's only a game! Wiley Series in Probability and Statistics*. (Wiley, New York, 2005). https://doi.org/10.1002/0471249696
30. PD Grnwald, IJ Myung, MA Pitt, *Advances in minimum description length: theory and applications (Neural Information Processing)*. (The MIT Press, Cambridge, 2005)
31. N Merhav, M Feder, A strong version of the redundancy-capacity theorem of universal coding. IEEE Trans. Inf. Theory. **41**(3), 714–722 (1995). https://doi.org/10.1109/18.382017

32. NN Cencov, *Statistical decision rules and optimal inference (translations of mathematical monographs), vol. 53*. (American Mathematical Society, Providence, 2000)

33. PP Vaidyanathan, The theory of linear prediction. Synth. Lect. Signal Process. **2**(1), 1–184 (2007). https://doi.org/10.2200/s00086ed1v01y200712spr003

34. PH Algoet, The strong law of large numbers for sequential decisions under uncertainty. IEEE Trans. Inf. Theory. **40**(3), 609–633 (1994). https://doi.org/10.1109/18.335876

35. EA Wan, RVD Merwe, in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*. The unscented kalman filter for nonlinear estimation (IEEE, Piscataway, 2000), pp. 153–158. https://doi.org/10.1109/asspcc.2000.882463

36. B Ristic, S Arulampalam, NJ Gordon, *Beyond the Kalman filter: particle filters for tracking applications*. (Artech house, London, 2004)

37. JGD Gooijer, RJ Hyndman, 25 years of time series forecasting. Int. J. Forecast. **22**(3), 443–473 (2006). https://doi.org/10.1016/j.ijforecast.2006.01.001

38. TM Cover, JA Thomas, *Elements of information theory*. (Wiley, New York, 2006)

39. A Goldsmith, P Varaiya, Capacity, mutual information, and coding for finite-state Markov channels. IEEE Trans. Inf. Theory. **42**(3), 868–886 (1996). https://doi.org/10.1109/isit.1994.394696

40. RM Neal, Markov chain sampling methods for Dirichlet process mixture models. J. Comput. Graph. Stat. **9**(2), 249–265 (2000). https://doi.org/10.1080/10618600.2000.10474879

41. M Dudí, SJ Phillips, RE Schapire, in *Learning Theory*. Performance guarantees for regularized maximum entropy density estimation (Springer, Berlin, Heidelberg, 2004), pp. 472–486

42. YW Teh, *Dirichlet Process*. (C Sammut, GI Webb, eds.) (Springer, Boston, 2010), pp. 280–287. https://doi.org/10.1007/978-0-387-30164-8

43. M Wellens, P Mähönen, Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model. Mob. Netw. Appl. **15**(3), 461–474 (2010). https://doi.org/10.1007/s11036-009-0199-9

44. MH Islam, CL Koh, SW Oh, X Qing, YY Lai, C Wang, Y-C Liang, BE Toh, F Chin, GL Tan, W Toh, in *2008 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2008)*. Spectrum survey in singapore: occupancy measurements and analyses (IEEE, Piscataway, 2008), pp. 1–7. https://doi.org/10.1109/crowncom.2008.4562457

45. W Tang, J Zhou, H Yu, S Li, A fair scheduling scheme based on collision statistics for cognitive radio networks. Concurr. Comput. Pract. Experience. **25**(9), 1091–1100 (2012). https://doi.org/10.1002/cpe.2879

46. C Xianfu, Z Honggang, AB Mackenzie, M Matinmikko, Predicting spectrum occupancies using a non-stationary hidden Markov model. IEEE Wirel. Commun. Lett. **3**(4), 333–336 (2014). https://doi.org/10.1109/LWC.2014.2315040

47. C Xu, H Jianwei, Evolutionarily stable spectrum access. IEEE Trans. Mob. Comput. **12**(7), 1281–1293 (2013). https://doi.org/10.1109/TMC.2012.94

48. P De, Y-C Liang, Blind spectrum sensing algorithms for cognitive radio networks. IEEE Trans. Veh. Technol. **57**(5), 2834–2842 (2008). https://doi.org/10.1109/tvt.2008.915520

49. P Huang, C-J Liu, L Xiao, J Chen, Wireless spectrum occupancy prediction based on partial periodic pattern mining. 2012 IEEE 20th Int. Symp. Model. Anal. Simul. Comput. Telecommun. Syst. **25**(7), 1925–1934 (2012). https://doi.org/10.1109/mascots.2012.16

50. S Arunthavanathan, S Kandeepan, RJ Evans, in *2013 IEEE Globecom Workshops (GC)*. Reinforcement learning based secondary user transmissions in cognitive radio networks (IEEE, Piscataway, 2013), pp. 374–379. https://doi.org/10.1109/glocomw.2013.6825016

51. J Yang, H Zhao, X Chen, in *IEEE 2nd International Conference on Computer and Communications (ICCC)*. Genetic algorithm optimized training for neural network spectrum prediction (IEEE, Piscataway, 2016), pp. 2949–2954. https://doi.org/10.1109/compcomm.2016.7925237

52. S Ni, X Bai, Z Wang, B Guo, in *IEEE International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. A new method of cognitive radio spectrum prediction research (IEEE, Piscataway, 2016), pp. 982–986. https://doi.org/10.1109/cisp-bmei.2016.7852855

53. A Agarwal, S Dubey, MA Khan, R Gangopadhyay, S Debnath, in *2016 International Conference on Signal Processing and Communications (SPCOM)*. Learning based primary user activity prediction in cognitive radio networks for efficient dynamic spectrum access (IEEE, Piscataway, 2016), pp. 1–5. https://doi.org/10.1109/SPCOM.2016.7746632

54. C Clancy, J Hecker, E Stuntebeck, T O'Shea, Applications of machine learning to cognitive radio networks. IEEE Wirel. Commun. **14**(4), 47–52 (2007). https://doi.org/10.1109/MWC.2007.4300983

55. L Gavrilovska, V Atanasovski, I Macaluso, LA DaSilva, Learning and reasoning in cognitive radio networks. IEEE Commun. Surv. Tutorials. **15**(4), 1761–1777 (2013). https://doi.org/10.1109/surv.2013.030713.00113

56. DC Karia, BK Lande, RD Daruwala, Performance analysis of HMM- and ANN-based spectrum vacancy predictor behaviour for cognitive radios. Int. J. Ad Hoc Ubiquit. Comput. **11**(4), 206–213 (2012). https://doi.org/10.1504/ijahuc.2012.050439

57. S-S Gu, S-N Yu, A chaotic neural network-based algorithm for relational structure matching. IEEE 2004 Int. Conf. Mach. Learn. Cybern. **6**, 3328–3333 (2004). https://doi.org/10.1109/icmlc.2004.1380353

58. MH Rehmani, AC Viana, H Khalife, S Fdida, SURF: A distributed channel selection strategy for data dissemination in multi-hop cognitive radio networks. Comput. Commun. **36**(10), 1172–1185 (2013). https://doi.org/10.1016/j.comcom.2013.03.005

59. S Bayhan, F Alagöz, Distributed channel selection in CRAHNs: A non-selfish scheme for mitigating spectrum fragmentation. Ad Hoc Netw. **10**(5), 774–788 (2012). https://doi.org/10.1016/j.adhoc.2011.04.010. Special Issue on Cognitive Radio Ad Hoc Networks

60. DP Bertsekas, JN Tsitsiklis, *Introduction to probability, Athena Scientific books*. (Athena Scientific, Belmont, 2002), pp. 9–16. https://doi.org/10.1017/cbo9780511996504.005

61. A Banaei, CN Georghiades, in *2009 IEEE International Conference on Communications*. Throughput analysis of a randomized sensing scheme in cell-based ad-hoc cognitive networks (IEEE, Piscataway, 2009), pp. 1–6. https://doi.org/10.1109/icc.2009.5199524

62. J Gambini, O Simeone, U Spagnolini, Y Bar-Ness, Y Kim, in *2008 IEEE International Conference on Communications*. Cognitive radio with secondary packet-by-packet vertical handover (IEEE, Piscataway, 2008), pp. 1050–1054. https://doi.org/10.1109/icc.2008.205

63. M Derakhshani, T Le-Ngoc, Learning-based opportunistic spectrum access with adaptive hopping transmission strategy. IEEE Trans. Wirel. Commun. **11**(11), 3957–3967 (2012). https://doi.org/10.1109/twc.2012.091812.111873

64. P Thakur, A Kumar, S Pandit, G Singh, SN Satashia, in *IEEE Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. Performance improvement of cognitive radio network using spectrum prediction and monitoring techniques for spectrum mobility (IEEE, Piscataway, 2016), pp. 679–684. https://doi.org/10.1109/pdgc.2016.7913208

65. M Khabazian, S Aissa, N Tadayon, Performance modeling of a two-tier primary-secondary network operated with IEEE 802.11 DCF mechanism. IEEE Trans. Wirel. Commun. **11**(9), 3047–3057 (2012). http://doi.org/10.1109/twc.2012.071612.110010

66. Z Wang, S Salous, Spectrum occupancy statistics and time series models for cognitive radio. J. Signal Process. Syst. **62**(2), 145–155 (2011). https://doi.org/10.1007/s11265-009-0352-5

67. J Zhang, G Ding, Y Xu, F Song, in *IEEE 8th International Conference on Wireless Communications & Signal Processing (WCSP)*. On the usefulness of spectrum prediction for dynamic spectrum access (IEEE, Piscataway, 2016), pp. 1–4. https://doi.org/10.1109/wcsp.2016.7752555

68. S Joshi, P Pawelczak, D Cabric, J Villasenor, When channel bonding is beneficial for opportunistic spectrum access networks. IEEE Trans. Wirel. Commun. **11**(11), 3942–3956 (2012). http://doi.org/10.1109/twc.2012.092512.111730

69. W Wang, T Lv, T Wang, X Yu, in *2010 IEEE 72nd Vehicular Technology Conference - Fall*. Primary user activity based channel allocation in cognitive radio network (IEEE, Ottawa, 2010), pp. 1–5. https://doi.org/10.1109/vetecf.2010.5594260, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5594260&isnumber=5594061

70. J Yang, H Zhao, Enhanced throughput of cognitive radio networks by imperfect spectrum prediction. IEEE Commun. Lett. **19**(10), 1738–1741 (2015). https://doi.org/10.1109/lcomm.2015.2442571

Eltom *et al. EURASIP Journal on Wireless Communications and Networking* (2018) 2018:29

Page 16 of 17

71. RD Smallwood, EJ Sondik, The optimal control of partially observable Markov processes over a finite horizon. Oper. Res. **21**(5), 1071–1088 (1973). https://doi.org/10.1287/opre.21.5.1071

72. D Blackwell, in *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes Held at Liblice Near Prague from November*. The entropy of functions of finite-state Markov chains, vol. 28 (Czechoslovak Academy of sciences, Czech Republic, 1957), pp. 13–20

73. T Kaijser, A limit theorem for partially observed Markov chains. Ann. Probab. **3**(4), 677–696 (1975). https://doi.org/10.1214/aop/1176996308

74. J Marroquin, S Mitter, T Poggio, Probabilistic solution of ill-posed problems in computational vision. J. Am. Stat. Assoc. **82**(397), 76–89 (1987). https://doi.org/10.2307/2289127

75. LR Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE. **77**(2), 257–286 (1989). https://doi.org/10.1109/5.18626, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=18626&isnumber=698

76. MS Arulampalam, S Maskell, N Gordon, T Clapp, A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. IEEE Trans. Signal Proc. **50**(2), 174–188 (2002). https://doi.org/10.1109/78.978374

77. D Haussler, A Barron, SCCRL University of California, *How well do Bayes methods work for on-line prediction of [[+ or -]1] values?, Technical reports*. (University of California, Santa Cruz, Computer Research Laboratory, California, 1992)

78. D Barber, *Bayesian time series models*. (Cambridge University Press, New York, 2011)

79. L Csurgai-Horváth, J Bito, in *Proceedings of the 2011 11th International Conference on Telecommunications (ConTEL)*. Primary and secondary user activity models for cognitive wireless network (IEEE, Piscataway, pp. 301–306

80. S Bayhan, F Alagöz, A Markovian approach for best-fit channel selection in cognitive radio networks. Ad Hoc Netw. **12**, 165–177 (2014). https://doi.org/10.1016/j.adhoc.2011.08.007

81. AW Min, KG Shin, Exploiting multi-channel diversity in spectrum-agile networks. IEEE Conf. Comput. Commun (2008). https://doi.org/10.1109/infocom.2007.256

82. Q Zhao, L Tong, A Swami, Y Chen, Decentralized cognitive mac for opportunistic spectrum access in ad-hoc networks: A pomdp framework. IEEE J. Sel. Areas Commun. **25**(3), 589–600 (2007). https://doi.org/10.1109/jsac.2007.070409

83. H Eltom, S Kandeepan, B Moran, RJ Evans, in *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*. Spectrum occupancy prediction using a hidden Markov model IEEE, Piscataway, 2015), pp. 1–8. https://doi.org/10.1109/icspcs.2015.7391772

84. Y Li, Y-N Dong, H Zhang, H-T Zhao, H-X Shi, X-X Zhao, in *IEEE 10th International Conference on Computer and Information Technology (CIT)*. Spectrum usage prediction based on high-order Markov model for cognitive radio networks (IEEE, Piscataway, 2010), pp. 2784–2788. https://doi.org/10.1109/cit.2010.464

85. J Riihijärvi, J Nasreddine, P Mähönen, in *European Wireless Conference (EW)*. Impact of primary user activity patterns on spatial spectrum reuse opportunities (IEEE, Piscataway, 2010), pp. 962–968. https://doi.org/10.1109/ew.2010.5483445

86. M Wellens, J Riihijarvi, P Mahonen, in *IEEE Annual Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops*. Modelling primary system activity in dynamic spectrum access networks by aggregated on/off-processes (IEEE, Piscataway, 2009), pp. 1–6. https://doi.org/10.1109/SAHCNW.2009.5172946

87. S Wang, J Zhang, L Tong, A characterization of delay performance of cognitive medium access. IEEE Trans. Wirel. Commun. **11**(2), 800–809 (2012). https://doi.org/10.1109/twc.2012.010312.110765

88. L Jiao, E Song, V Pla, FY Li, Capacity upper bound of channel assembling in cognitive radio networks with quasistationary primary user activities. IEEE Trans. Veh. Technol. **62**(4), 1849–1855 (2013). https://doi.org/10.1109/tvt.2012.2236115

89. SD Barnes, BT Maharaj, Prediction based channel allocation performance for cognitive radio. AEU - Int. J. Electron. Commun. **68**(4), 336–345 (2014). https://doi.org/10.1016/j.aeue.2013.09.009

90. L Meliá Gutiérrez, S Zazo, JL Blanco-Murillo, I Pérez-Álvarez, A García-Rodríguez, B Pérez-Díaz, HF spectrum activity prediction model based on HMM for cognitive radio applications. Phys. Commun. **9**, 199–211 (2013). https://doi.org/10.1016/j.phycom.2012.09.004

91. T Nguyen, BL Mark, Y Ephraim, Spectrum sensing using a hidden bivariate Markov model. IEEE Trans. Wirel. Commun. **12**(9), 4582–4591 (2013). https://doi.org/10.1109/twc.2013.072513.121864

92. A Saad, B Staehle, R Knorr, in *IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. Spectrum prediction using hidden Markov models for industrial cognitive radio (IEEE, Piscataway, 2016), pp. 1–7. https://doi.org/10.1109/wimob.2016.7763231

93. H Eltom, S Kandeepan, YC Liang, B Moran, RJ Evans, in *2016 IEEE International Conference on Communications Workshops (ICC)*. HMM based cooperative spectrum occupancy prediction using hard fusion (IEEE, Piscataway, 2016), pp. 669–675. https://doi.org/10.1109/iccw.2016.7503864

94. C-H Liu, D Cabric, Prediction of Erlang-2 distributed primary user traffic for dynamic spectrum access. IEEE Wirel. Commun. Lett. **4**(5), 481–484 (2015). https://doi.org/10.1109/lwc.2015.2442249

95. SH Sohn, HMM-based adaptive frequency-hopping cognitive radio system to reduce interference time and to improve throughput. KSII Trans. Internet Inf. Syst. **4**(4), 475–490 (2010). https://doi.org/10.3837/tiis.2010.08.002

96. Y Zhao, Z Hong, G Wang, J Huang, in *IEEE 25th International Conference on Computer Communication and Networks (ICCCN)*. High-order hidden bivariate Markov model: A novel approach on spectrum prediction (IEEE, Piscataway, 2016), pp. 1–7. https://doi.org/10.1109/icccn.2016.7568528

97. SS Dias, MGS Bruno, Cooperative target tracking using decentralized particle filtering and RSS sensors. IEEE Trans. Signal Proc. **61**(14), 3632–3646 (2013). https://doi.org/10.1109/tsp.2013.2262276

98. X Xing, T Jing, W Cheng, Y Huo, X Cheng, T Znati, Cooperative spectrum prediction in multi-PU multi-SU cognitive radio networks. Mob. Netw. Appl. **19**(4), 502–511 (2014). https://doi.org/10.1007/s11036-014-0507-x

99. D Dash, A Sabharwal, Paranoid secondary: waterfilling in a cognitive interference channel with partial knowledge. IEEE Trans. Wirel. Commun. **11**(3), 1045–1055 (2012). https://doi.org/10.1109/TWC.2012.012412.110348

100. B Canberk, IF Akyildiz, S Oktug, Primary user activity modeling using first-difference filter clustering and correlation in cognitive radio networks. IEEE/ACM Trans. Netw. **19**(1), 170–183 (2011). https://doi.org/10.1109/tnet.2010.2065031

101. Z Wen, T Luo, W Xiang, S Majhi, Y Ma, in *ICC Workshops - 2008 IEEE International Conference on Communications Workshops*. Autoregressive spectrum hole prediction model for cognitive radio systems (IEEE, Beijing, 2008), pp. 154–157. https://doi.org/10.1109/ICCW.2008.34, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4531882&isnumber=4531848

102. D Willkomm, S Machiraju, J Bolot, A Wolisz, Primary user behavior in cellular networks and implications for dynamic spectrum access. IEEE Commun. Mag. **47**(3), 88–95 (2009). https://doi.org/10.1109/mcom.2009.4804392

103. A Eltholth, in *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*. Forward Backward autoregressive spectrum prediction scheme in cognitive radio systems (IEEE, Piscataway, 2015), pp. 1–5. https://doi.org/10.1109/ICSPCS.2015.7391770

104. K Sithamparanathan, A Giorgetti, *Cognitive radio techniques: spectrum sensing, interference mitigation, and localizatio. Artech House mobile communications library* (Artech House, Boston, 2012)

105. W Saad, Z Han, HV Poor, T Basar, JB Song, A cooperative bayesian nonparametric framework for primary user activity monitoring in cognitive radio networks. IEEE J. Sel. Areas Commun. **30**(9), 1815–1822 (2012). https://doi.org/10.1109/JSAC.2012.121027

106. M Lopez-Benitez, F Casadevall, Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks. IEEE Trans. Veh. Technol. **62**(5), 2091–2104 (2013). https://doi.org/10.1109/tvt.2013.2238960

107. VA Epanechnikov, Non-parametric estimation of a multivariate probability density. Theory Probab. Appl. **14**(1), 153–158 (1969). https://doi.org/10.1137/1114019

108. I Macaluso, D Finn, B Ozgul, LA DaSilva, Complexity of spectrum activity and benefits of reinforcement learning for dynamic channel selection. IEEE J. Sel. Areas Commun. **31**(11), 2237–2248 (2013). https://doi.org/10.1109/JSAC.2013.131115

109. J Rissanen, Strong optimality of the normalized ML models as universal codes and information in data. IEEE Trans. Inf. Theory. **47**(5), 1712–1717 (2001). https://doi.org/10.1109/18.93091

110. F Hou, X Chen, H Huang, X Jing, in *2016 16th International Symposium on Communications and Information Technologies (ISCIT)*. Throughput performance improvement in cognitive radio networks based on spectrum prediction, (2016), pp. 655–658. https://doi.org/10.1109/iscit.2016.7751715

111. H Li, RC Qiu, in *IEEE Global Telecommunications Conference (GLOBECOM)*. A graphical framework for spectrum modeling and decision making in cognitive radio networks (IEEE, Piscataway, 2010), pp. 1–6. https://doi.org/10.1109/GLOCOM.2010.5683361

112. IA Akbar, WH Tranter, in *Proceedings of 2007 IEEE SoutheastCon*. Dynamic spectrum aladdress in cognitive radio using hidden Markov models: Poisson distributed case (IEEE, Piscataway, 2007), pp. 196–201. https://doi.org/10.1109/SECON.2007.342884