# Spectrum Prediction in Dynamic Spectrum Access Systems

A thesis submitted in fulfilment of the requirements for the degree of

*Doctor of Philosophy*

**Hamid Eltom**

B.Sc (Honours) Electrical and Electronics
University of Khartoum

School of Engineering
College of Science, Engineering and Health
RMIT University

February 2018

# Declaration

I declare that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university, and that to the best of my knowledge it does not contain any materials previously published or written by another person except where due reference is made in the text.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

This research was supported under Australian Research Council's Discovery Projects funding scheme (Cognitive Radars for Auto-mobiles, DP150104473).

Hamid Eltom
February 2018

*A modest contribution for a better future.*

*For: Najat Hamid*

*To: Najat Hamid*

# Acknowledgements

I would first like to thank my thesis advisor A/Professor Kandeepan Sithamparanathan. The door to Prof. Sithamparanathan office was always open whenever I ran into a trouble spot or had a question about my research or writing. He/She consistently allowed this paper to be my own work, but steered me in the right the direction whenever he thought I needed it. I would like to thank my fellow doctoral students for their feedback, cooperation and of course friendship.

I would like to thank my friends for accepting nothing less than excellence from me. Last but not the least, I would like to thank my family (My wife Amira, my sister Salma, and my father) for supporting me spiritually throughout writing this thesis and my life in general.

This research was supported under Australian Research Council's Discovery Projects funding scheme (Cognitive Radars for Auto-mobiles, DP150104473)

# Table of contents

# List of tables

# List of figures

# List of Abbreviations

**CR** Congitive radio

**CFAR** Constant false alarm

**DSA** dynamic spectrum access

**ED** Energy detector

**EG** equal gain

**HF** hard fusion

**HMM** hidden markov mode

**IoT** intrent of things

**MRC** maximum ratio combining

**PU** primary user

**SOP** spectrum occupancy prediction

**SC** selection combining

**SU** secondary user

**SF** soft fusion

**WRAN** Wireless Regional Access Networks

**SNR** signal to noise ratio

# Notations and Symbols

Table 1 Notations and Symbols

| | |
|---|---|
| $\mathbf{X}$ | Vector or matrix |
| $\mathbb{E}[\cdot]$ | Expectation operator |
| $\mathbb{V}[\cdot]$ | Variance operator |
| $\det(\cdot)$ | Determinant of a matrix |
| $\mathbf{I}$ | Identity matrix |
| $\mathbf{1}$ | Column vector of ones |
| $[\cdot]^{\mathrm{T}}$ | Transpose operator |

# Abstract

Despite the remarkable foreseen advancements in maximizing network capacities, the in-expansible nature of radio spectrum exposed outdated spectrum management techniques as a core limitation. Fixed spectrum allocation inefficiency has generated a proliferation of dynamic spectrum access solutions to accommodate the growing demand for wireless, and mobile applications.
This research primarily focuses on spectrum occupancy prediction which equip dynamic users with the cognitive ability to identify and exploit instantaneous availability of spectrum opportunities. The first part of this research is devoted to identifying candidate occupancy prediction techniques suitable for SOP scenarios are extensively analysed, and a theoretical based model selection framework is consolidated. The performance of single user Bayesian/Markov based techniques both analytically and numerically. Understanding performance bounds of Bayesian/Markov prediction allows the development of efficient occupancy prediction models. The third and fourth parts of this research investigates cooperative decision and data-based occupancy prediction. The expected cooperative prediction accuracy gain is addressed based on the single user prediction model. Specifically, the third contributions provide analytical approximations of single user, as well as cooperative hard fusion based spectrum prediction. Finally, the forth contribution shows soft fusion is superior and more robust compared to hard fusion cooperative prediction in terms of prediction accuracy. Throughout this research, case study analysis is provided to evaluate the performance of the proposed approaches. Analytical approaches and Monte-Carlo simulation are compared for the performance metric of interest. Remarkably, the case study analysis confirmed that the statistical approximation can predict the performance of local and hard fusion cooperative prediction accurately, capturing all the essential aspects of signal detection performance, temporal dependency of spectrum occupancy as well as the finite nature of the network.

# Publications

- **H. Eltom**, S. Kandeepan, B. Moran, and R. J. Evans, “Spectrum occupancy prediction using a hidden Markov model” 2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS), Dec 2015 [2].
- **H. Eltom**, S. Kandeepan, Y. C. Liang, B. Moran, and R. J. Evans, “HMM based cooperative spectrum occupancy prediction using hard fusion,” 2016 IEEE International Conference on Communications Workshops (ICC), May 2016 [3].
- **H. Eltom**, S. Kandeepan, R. J. Evans, Y. C. Liang, and B. Ristic,"Statistical Spectrum Occupancy Prediction for Dynamic Spectrum Access: A Classification ", EURASIP Journal on Wireless Communications and Networking, February 2018 [4].
- **Hamid Eltom**, Sithamparanathan Kandeepan, Y.C. Liang, and Robin J. Evans,"Cooperative Soft Fusion for HMM based Spectrum Occupancy Prediction", Submitted to IEEE Communications letters.
- **Hamid Eltom**, Sithamparanathan Kandeepan, Y.C. Liang, and Robin J. Evans,"An approximation of stationary posterior distribution of HMM based Spectrum Occupancy Prediction", Submitted to IEEE Communications letters.
- Hourani, A., Evans, R., Sithamparanathan, K., Moran, W., **Eltom, H.** " Stochastic Geometry Methods for Modelling Automotive Radar Interference In: IEEE Transactions on Intelligent Transportation Systems, 2017 [5] .

# Chapter 1

# Introduction

## 1.1 Background

The ubiquity of mobile phones and Wi-Fi access points overemphasises the impending age of Internet of things (IoT), which intuitively projects an exponential increase in demand for wireless data traffic. The seemingly augmenting mobile services subscription rate as well as the proliferation of mobile media consumption overshadows the efforts by the wireless communication industry to fulfil the requirements for next generation communication systems. The anticipated (1000x) demand factor in data traffic, which was set as a target requirement for (5G) networks, manifests the axiom of exponential increase in wireless data traffic. Advanced receiver design, multi-antenna techniques accompanied by cooperative and heterogeneous networks deployments, actualized the current innovative state of wireless communication development to address the increasing demand. However, despite the remarkable foreseen advancements in maximizing network capacities, the *in-expansible* nature of radio spectrum exposed outdated spectrum management techniques as a core limitation. Spectrum scarcity is the paradigm that describes the artificial problem in spectrum sharing regulation policies created by inefficient allocation of frequency bands [1, 6, 7].

Fixed spectrum allocation inefficiency has generated a proliferation of dynamic spectrum access solutions to accommodate the growing demand for wireless, and mobile applications. Software-defined radios (or cognitive radio) networks which offer a dynamic spectrum access management policy of fixed allocated licensed bands, could potentially solve the artificial spectrum scarcity problem. Dynamic Spectrum Access

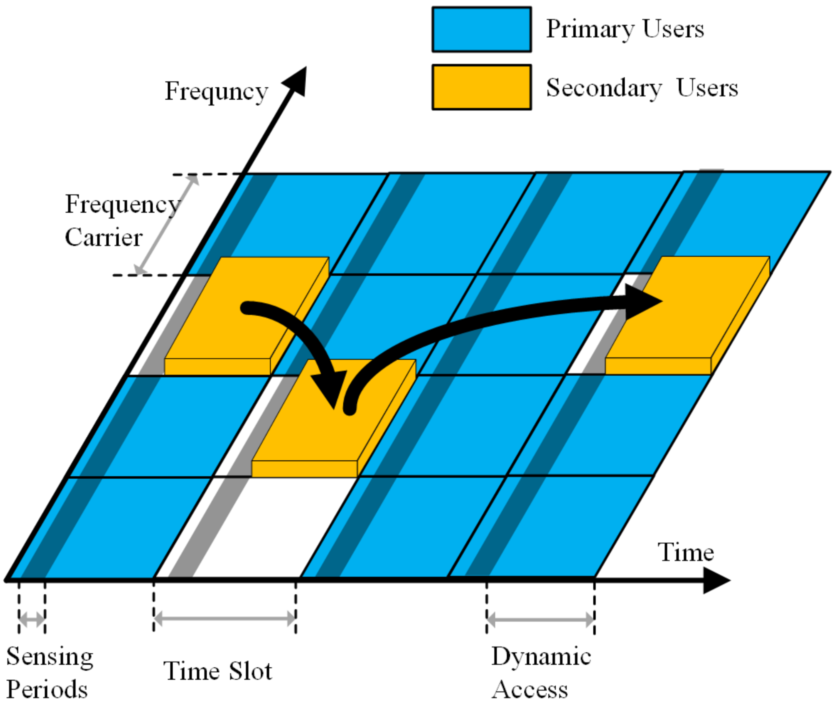


Fig. 1.1 An example of spectrum sensing and access in a typical DSA time-slotted system.

(**DSA**) systems typically consist of licensed *Primary* users (**PU**'s), and opportunistic *Secondary* users (**SU**'s). 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.1.1). 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) [6, 8]. 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 preform spectrum *Sensing*, *Decision*, *Sharing*, and *Mobility* [8, 9]. Spectrum occupancy prediction (**SOP**) models were proposed in DSA literature to optimize cognitive cycle functional processing time [10]. SOP models add agility, and adaptability to cognitive radio functions to optimize processes such as periodic spectrum sensing scheduling, and channel selection in spectrum decision [8]. Similarly, SOP models allow the implementation of a proactive spectrum mobility strategy based on predicted occupancy patterns which avoids collisions with incumbent primary users [10, 11]. The resource efficiency that spectrum prediction potentially adds to dynamic spectrum access, emphasizes the importance of this cognitive radio enabling process.

## 1.2 Motivation

Despite the crucial role of spectrum prediction in enabling cognitive radio to efficiently utilize spectrum opportunities, spectrum prediction literature is scarce. Current literature on spectrum prediction in DSA systems lacks a consolidated framework for model selection, as well as prediction performance analysis. Additionally, candidate prediction techniques suitable for SOP scenarios are not extensively classified nor immediately identified. Current framework of prediction model selection is application dependent where SOP proposals restrict the model validity to a specific technology scenario. While several proposals simulated Bayesian/Markov based prediction models for different wireless technologies, the abstract ability of these models to provide accurate occupancy modelling and prediction is still not intuitively clear [10]. In other words, the functional relationship between prediction performance measures and model parameters are not explicitly studied, formulated nor approximated.

In this thesis, performance analysis of spectrum prediction is the primary focus. In particular, this research is devoted to explore the performance of Bayesian/Markov based techniques (namely, Hidden Markov Model **HMM**) both analytically and numerically. Understanding performance bounds of HMM based prediction allows the development of efficient SOP models. An extensive review of sequential prediction in cognitive radio is pursued to identify the statistical framework of prediction model selection. Then, numerical recursive techniques and statistical approximations of single, and multi-user spectrum prediction performance analysis are proposed. Section-1.3 provides an overview of spectrum occupancy, while Subsection-1.4 provides an introduction into the proposed HMM model.

## 1.3 Spectrum Occupancy Prediction Overview

The motivation for SOP models is to minimize the accumulated *time delay* due to cognitive cycle serial functional processing. By predicting the channel status in advance, more processing time becomes available for spectrum sensing, decisions, and mobility (Fig.1.2) [10]. Performance gains of SOP can is manifested in the context of:

- Wireless Regional Access Networks **WRAN** (802.22); TV UHF/VHF bands attracted attention as an ideal candidate to provide high speed data communication, due to its appealing radio characteristics (low noise, reasonable antenna size and line of sight). As analogue TV bands are being unused with the dominance of digital TV; these UHF/VHF bands represent an opportunity to provide valuable growth capital for service providers [12]. 802.22 **WRAN** provides two methods of obtaining spectrum occupancy, 1) through geo-location and database, and 2) through spectrum sensing. Spectrum occupancy prediction is a potentially valuable asset in optimizing 802.22 sensing and dynamic spectrum access abilities. Through additional layer of machine learning, cognitive radio networks are able to opportunistically utilize the spatial/ temporal spectrum opportunities to provide data access in WRAN white TV spaces.

- Consumer Cognitive Radar (DSA cognitive radar) [7, 13–15]: Cognitive radar is a concept developed by S. Haykin as the future of fully adaptive and efficient radar that adjusts its parameters for faster and more accurate surveillance [14]. Though dynamic spectrum access was not initially proposed since the intended application was predominantly military focused; consumer radar adopting both DSA and cognitive radar concepts is a new realization of the cognitive radar[15]. Consumer radar with DSA capabilities can be realized in vehicular and/or aerial scenarios. Such radar would make an efficient use of limited shared spectrum bands, while providing intelligent collision avoidance and environment awareness. One feasible scenario involves collision avoidance for vehicle highway scenario [1]. The motivation in this thesis to address this particular application stems from the ARC project for automotive radar interference mitigation/avoidance [5] (76-77 GHZ band set by the Australian Communication Authority for automotive radars [16]).

### 1.3.1 Spectrum Occupancy Classification

Unless specified otherwise , in the cognitive DSA system design such as 802.22 [12], spectrum occupancy state is the result of the stochastic sampling process of radio environment. However, typical stochastic sampling intuitively carries stochastic randomness gained from combination of sampling process accuracy and/or underlying stochastic activity patterns [17]. Spectrum occupancy prediction models target

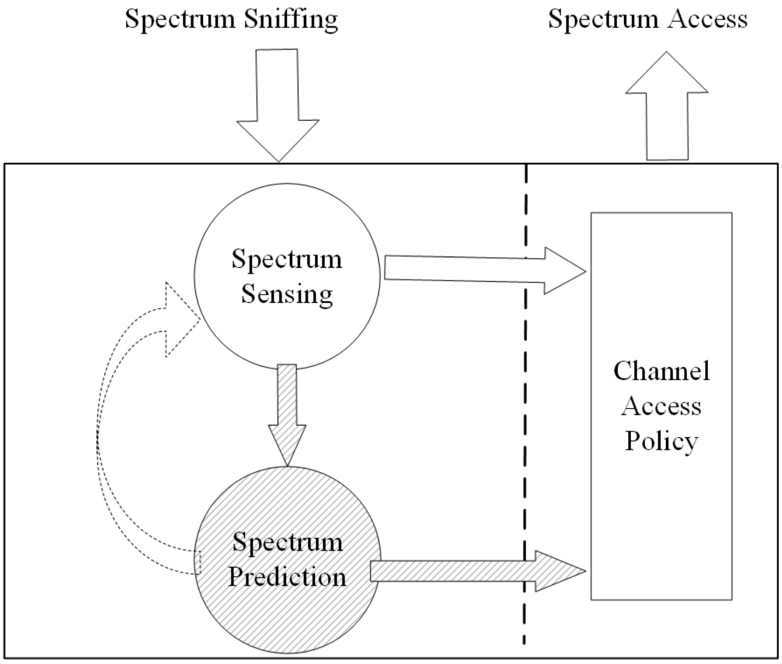


Fig. 1.2 Spectrum prediction in dynamic spectrum access framework

the usage of spectrum bands utilized by both incumbent users, and opportunistic users.

SOP models broadly target 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* [1, 18]. In the DSA literature, spectrum occupancy prediction techniques address prediction either *explicitly* [19–21], or *implicitly*. Implicit approaches present SOP models as primary/secondary user's occupancy/activity models. Statistical SOP models proposed for spectrum occupancy analysis include Poisson processes [22, 23], Bayesian prediction [24, 25], and linear regression [26, 27]. Machine learning based techniques have also been proposed for model *learning* including neural networks, time regression, and space vector machines [10, 28, 29]. The framework of current research into spectrum occupancy can be divided into the following generic categories:

- Measurement campaign modelling: An empirical research conducted in specific scenarios (indoor, outdoor) to collect real life data using an antenna that covers specific frequency band. Statistical analysis, averaging and maximization are conducted to generate an approximate statically density functions and/ or simple statistical description of power, channel occupancy, etc. Though such modelling is able to capture the real life data, it is however, riddled with inaccuracy, spatial and temporal dependency[18].

- Statistical primary user modelling and prediction: Statistical modelling aims to find the best suitable stochastic procedure that describes a specific primary user(s) mode of operation and its channel utilization. Given such statistical estimation, prediction of possible presence of absence in future instances is possible [10, 11].

Chapter-2 elaborates on spectrum prediction classification and presents our published framework and review on spectrum prediction in dynamic spectrum access systems. The basis of the proposed framework is the sequential statistical prediction introduced in next the section as well as Chapter-2.

### 1.3.2 Statistical Prediction

Statistical prediction in it's simplest form is: "Given an observation sequence $x_{1:t-1}$ up to time instant $t-1$, and before the symbol at time $t$ is revealed the predictor guesses the next value $x_t$ based on the previous $t-1$ observations ". The observations are assumed to follow a stationary stochastic process, and thus statistical properties can be derived from the past observations and an effective rule can be postulated from these statistics. The predictor measures the loss function between the actual value and the predicted value. The loss function measures the quality of the prediction and provides the input to compare prediction rules [17, 30]. Based on the main components of the statistical prediction problem presented in Chapter-2, HMM (in the next subsection) is introduced in Section-1.4, and more details are presented in Chapter-3.

### 1.3.3 Bayesian Based Spectrum Prediction

Bayesian-based prediction techniques provide powerful, and flexible tools to learn and adapt to the radio environment. Secondary users within the cognitive radio network, collect sensing information, and utilize statistical correlation, to infer possible future states of the primary user usage patterns. Bayesian *Mixture* models scenario are extensively studied in information and coding theory [31–33]. Bayesian algorithms are minimax optimal, and are universal under self information loss functions [34–36]. The algorithms perform well under both probabilistic and deterministic non-stochastic settings [31, 34, 35, 37, 38]. Additionally, Markov-based construction is attractive due to the desirable convergence properties of Markov chain based models [39–42]. *Markov chain*, and *partially observable Markov* models are commonly used for spectrum occupancy modelling. Chapter-3 considers single user prediction scenarios are based on HMM system model.

### 1.3.4 Cooperative spectral prediction

Spectrum prediction in single secondary user environment is commonly known as *local spectrum prediction*. Consequently, *cooperative spectrum prediction* in multi-user environment was proposed to improve the collective accuracy of spectrum occupancy prediction [43, 44]. Cooperative fusion of secondary user's decisions has been studied extensively in DSA based solutions to address diverse optimization problems. A multitude of decision fusion techniques were used such as hard and soft for temporal and, spatial fusion for local node decisions [45]. Combined decision fusion and Bayesian estimation has been also suggested for decentralized tracking [46].

In spectrum occupancy prediction literature, cooperative spectrum prediction fusion was only studied in handful of papers such as [20] where a coalition based game theory approach was implanted for a multi-primary/secondary users environment. The study showed general improvements when using cooperative prediction, but the results lacked fine details of improvement in terms of dependency on traffic load, and/or detection accuracy. Chapter-4 and Chapter-5 focus on hard and soft fusion of spectrum prediction decisions, respectively.

## 1.4 Mathematical Preliminaries for Spectrum Occupancy Prediction

The flow chart in Fig-1.3 highlights the sequence of spectrum occupancy prediction process. Spectrum sensing provides occupancy observations for local spectrum prediction. Given the selected prediction model (Chapter-2), the model parameters are estimated based on a training sequence. Then, Bayesian methods are used to estimate the probability that the next spectrum opportunity is available. Finally, using a predefined threshold the prediction decision is made and passed to spectrum decision function. This section provides background information on spectrum sensing, hidden Markov models, and maximum likelihood estimation *Baum Welch* algorithm

[47]. Background and related contents contained in this section are used for system models in Chapter-2, and statistical analysis of spectrum prediction in Chapter-3 and Chapter-5.

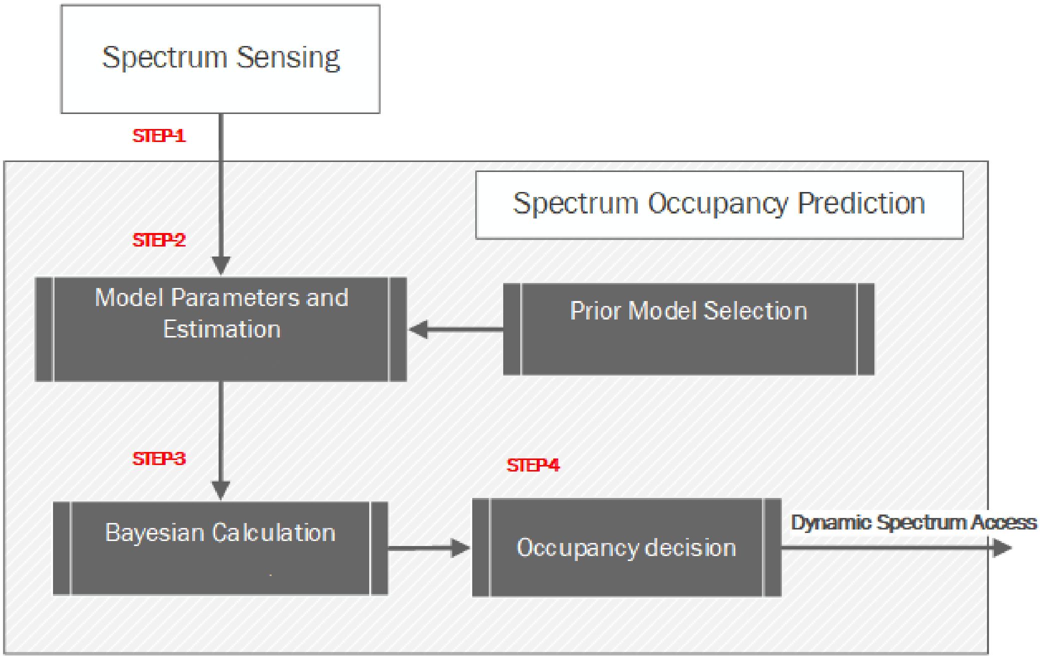


Fig. 1.3 Spectrum occupancy prediction flowchart

### 1.4.1 Some Theoretical Preliminaries

The theories and mathematical background in this subsection are the basis of statistical analysis for SOP models in Chapter-3 and Chapter-5, as well as Section-1.4.3.

To estimate the probability of error in estimation of a random variable, Fano inequality serves as bound for error probability, and definitions of Markov chain stationary distribution, Cesáro mean, and Entropy rate are needed to derive such approximation.

**Theorem 1.4.1** **(Cesáro Mean).** *Cesaro mean for convergent sequences states regarding the arithmetic mean states [48]: let* $a_n \rightarrow \mathbf{A}$*, let* $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$*, then* $\lim_{n\rightarrow\infty} b_n = \mathbf{A}$.

**Theorem 1.4.2** **(Entropy Rate).** *Entropy rate of a stochastic process* $\mathbf{X}_n$ *is defined by* $\mathbf{H}(\mathcal{X}) = \lim_{n\rightarrow\infty} \frac{1}{n}\mathbf{H}(\mathbf{X^n})$ *[38]. For a stationary process the limit exists and equals* *the other notion of entropy rate defined as* $\mathbf{H}^{'}(\mathcal{X}) = \lim_{n\to\infty} \mathbf{H}(\mathbf{X_n}|\mathbf{X^{n-1}})$.

**Theorem 1.4.3 (Fano Inequality).** *let* $\boldsymbol{X}$ *be a random variable with finite outcomes in* $\mathcal{X}$, *let* $\hat{\mathbf{X}} = g(\mathbf{Y})$ *be an estimated value of* $\boldsymbol{X}$ *for some deterministic function of* $g$ *that takes values in* $\mathcal{X}$, *then the loose probability of error bound is stated as [38]*

$$p_e \triangleq p_e(\hat{\mathbf{X}} \neq \mathbf{X}) \geq \frac{\mathbf{H}(\mathbf{X}|\mathbf{Y}) - 1}{\log|\mathcal{X}|}$$

*or more strongly as*

$$\mathbf{H}(Ber(p_e)) + p_e \log(|\mathcal{X} - 1|) \geq \mathbf{H}(\mathbf{X}|\mathbf{Y})$$

*Where* $\mathbf{H}(\mathbf{X}|\mathbf{Y}) = \mathbf{E}_{p(x,y)} \log_2 \frac{1}{p(x|y)}$, *and* $Ber(p_e)$ *is error Bernoulli random variable.*

**Distance measure** Following [31], *instantaneous Distance*, and *total distance* between probability distributions $P, Q_w$ are defined as:

$$d_t(x_t|x_{0:t-1}) := \sum_{x_{t-1}} P(x_t|x_{0:t-1}) \ln \frac{P(x_t|x_{0:t-1})}{Q_w(x_t|x_{0:t-1})}, \quad D_n := \sum_{t=1}^{n} \mathbb{E}[d_t].$$

Where $d_t$ is the instantaneous KL divergence. While $D_n$ is the total distance counterpart. $D_n$ is chosen as the distance measure in Section-1.4.

**Beta-Bernoulli Distribution** Bernoulli random variable $X$ takes only the values 0 or 1 representing failure and success, respectively. The probability mass function parametrised by $\rho$ the probability of success is:

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

If $\rho$ is assumed to be drawn from a random distribution, then the conjugate prior distribution is then a *Beta* distribution given by:

$$p(\rho|\alpha,\beta) \sim Beta(\alpha,\beta) \quad \alpha,\beta > 0$$
$$p(\rho|\alpha,\beta,a,b) = \frac{(\rho)^{\alpha-1}(1\rho)^{\beta-1}}{\mathbf{B}(\alpha,\beta)}$$
$$\mathbf{B}(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$$

where $B(\alpha, \beta)$ is the standard beta function. The posterior predictive distribution is the compound distribution given by :

$$f(k|\alpha, \beta) = \frac{\mathbf{B}(k + \alpha, 1 - k + \beta)}{\mathbf{B}(\alpha, \beta)}$$

The Beta-Bernoulli distribution is a special case of the Beta-Binomial distribution [49, 50] for the number of successes $k$ in $n$ trials :

$$f(k|n, \alpha, \beta) = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k-1)} \frac{\mathbf{B}(k + \alpha, n - k + \beta)}{\mathbf{B}(\alpha, \beta)}$$

Beta-Bernouli and Beta-binomial are used in Chapter-3/Chapter-4 for the proposed approximation of local, and cooperative prediction error distribution, respectively.

### 1.4.2 Spectrum Sensing in Dynamic Spectrum Access

Spectrum sensing provides a method of using the spectrum more efficiently. Spectrum sensing enables cognitive radio devices to access, and monitor free sections of radio spectrum. Cognitive radio must relay on spectrum sensing to keep monitoring the spectrum to avoid undue interference. Sensing functions must then be able to detect other transmissions, and identify the sources.

#### Cognitive Radio Spectrum Sensing Basics

Spectrum sensing provides the instantaneous occupancy status of every spectrum opportunity. Consequently, spectrum sensing algorithms must accommodate several considerations to effectively coexist in a multi-user environment. Sensing must be continuous to avoid causing interference to primary users, but also to identify alternative spectrum availabilities. Sensing cognitive ability to recognize the transmission source and type in order to identify spurious interference sources is crucial element of the functionality.

Spectrum sensing can be also performed in a cooperative manner where several radio units to improve sensing robustness to hidden sources. However, parameters such as sensing bandwidth and transmission type identification are the some of the key challenges to sensing algorithm design. Additionally, accuracy and window timing control the efficiency of spectrum sensing utilization of available spectrum opportunities. Signal detection techniques used in spectrum sensing includes energy detectors, cyclostationary, matched filter and Bayesian detectors. Basic energy detector is described in subsection-1.4.2 along with correct/incorrect detection probabilities. Each secondary user prediction performance relies on spectrum sensing accuracy. The results in Chapter-3 and Chapter-5 are presented against sensing detection probabilities, as well as HMM model parameters (subsection-1.4.3) [45].

### Spectrum Detection Accuracy

The received signal $z_{r,t}$ is detected by a complex baseband equivalent of an energy detector. The two hypothesis of *present*, and *absent* signals for an observation period of $T$ or the equivalent of $N$ samples for each time slot $t$. Test statistics $\Phi_{z,r,t}$ for time slot $t$ for a large number of samples $N$ is assumed to be Gaussian distributed (Central limit Theorem) [51]. The test statistic of the analogue energy detector can be given as:

$$\Phi_{z,r,t} \sim \mathcal{N}(\sum_{i=1}^{N} \mathbb{E}[|z_{r,t}^i|^2], \sum_{i=1}^{N} \mathbb{V}[|z_{r,t}^i|^2]) \tag{1.1}$$

$$\begin{aligned} \mathbb{E}[|z_{r,t}^i|^2] &= \sigma_{w_r}^2 + \theta |s_t^i|^2 \\ \mathbb{V}[|z_{r,t}^i|^2]) &= (\sigma_{w_r}^2)^2 + \theta \sigma_w^2 |s_t^i|^2 \end{aligned} \tag{1.2}$$

where $s_t$ is the transmitted signal, channel noise is assumed as a zero mean Gaussian noise $w_r$ with variance $\sigma_{w_r}^2$, and $\theta = [0, 1]$ the null and alternative hypothesis, respectively.

$$\rho_r = \frac{1}{N} \frac{\sum_{i=1}^{N} |s_t^i|^2}{\sigma_{w_r}^2} \tag{1.3}$$

Hypothesis testing on the detection statistics yields the the series $y_{r,t}$ i.e., the occupancy perceived by each SU. The uncertainty around spectrum sensing performance is quantified by the probability of correct detection $P_d(r)$, and the probability of false alarm, and $P_f(r)$ can be defined using central limit theorem approximation for large number of samples $N$ [45]:

$$
\begin{aligned}
P_f(r) &\approx Q(\frac{\lambda - N\sigma_w^2}{\sqrt{N}\sigma_w^2}) \\
P_d(r) &\approx Q(\frac{\lambda - \sqrt{N}\sigma_w^2(1+\rho_r)}{\sqrt{N}\sigma_w^2(1+\rho_r)})
\end{aligned} \tag{1.4}
$$

The probability of detection $P_d = 1 - P_m$ is the probability of successfully observing the channel correctly as busy where $P_m$ is the miss-detection probability, and the probability of false alarm $P_f$ is the probability of observing occupied channel while the primary user is idle. For constant false alarm based approaches (**CFAR**), the threshold $\lambda$ in is calculated, for a large number of samples $N$ using inverse Q-function

### 1.4.3 Hidden Markov Model

This section describes the estimation and training of hidden models analysed in Chapter-3. HMM is the proposed system model for single/multi user spectrum prediction occupancy as seen by the secondary user, which is used in in Chapter-3, Chapter-4, and Chapter-5. Markov-based construction is attractive due to the desirable convergence properties of Markov chain based models [39–42]. *Markov property* describes the case when the probability of current event $x_t$ only depends on the probability of previous event $x_{t-1}$ i.e., $p(x_t|x^{t-1}) = p(x_t|x_{t-1})$ [48]. The Discrete-Time HMM can be fully defined as follows:

$$
\begin{aligned}
\lambda &\triangleq (\mathbf{P}, \mathbf{E}, v) \\
v &\triangleq p(x_0 = i), \quad : i \in \{1, ..\mathbf{K}\}. \\
\mathbf{P} &\triangleq p(x_t = j|x_{t-1} = i) \quad : j \in \{1, ..\mathbf{K}\}. \\
\mathbf{E}_r &\triangleq p(y_{r,t} = k|x_t = j) \quad : k \in \{1, ..\mathbf{L}\}.
\end{aligned} \tag{1.5}
$$

Where $\mathbf{P}$ is the hidden state transition matrix, $\mathbf{L}$ is the total number of hidden states, $\mathbf{E}_r$ is the observation emission matrix, $\mathbf{K}$ is the total number of possible observation states, and $v$ is the initial state vector.

#### Event Space and transition matrix

The two state of a primary user activity in a frequency channel are *Idle and Busy*. Define $\{x_t : t \geq 1\}, x \in \{0, 1\}$ as the channel state at time slot $t$ as irreducible stationary Markov chain. The transition probability matrix between states $i, j \in [0, 1]$ i.e. $p(x_t = j|x_0, .., x_{t-1} = i) = p(x_t|x_{t-1} = i$ is given by [48]:

$$\mathbf{P} = [p_{ij}] = \begin{bmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{bmatrix} \tag{1.6}$$

Where $\mathbf{P}$ is a transition probability matrix for all time instants $1..t$.

### Observation Space, and the emission matrix

In a two state hidden Markov model, observations are independent and identically distributed random variables. The two hypothesis of a secondary user's perceived occupancy are *present*, and *absent*. The relation between the observation sequence $y_{1:t}$, and the hidden event space is the emission matrix that represents the probability of an observation is an outcome of a specific hidden state. The emission matrix is designed as follows :

$$\mathbf{E} = [e_{ij}] = \begin{bmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{bmatrix} \tag{1.7}$$

Secondary user's model is a hidden Markov model characterized by the transition $\mathbf{P}$, and emission $\mathbf{E}_r$ matrices, as well as the initial state distribution $p(x_0)$. The observation sequence $y_{r,t}$ is characterised by the emission matrix $\mathbf{E}_r$ of each SU which maps Markov chain based PU channel activity $x_{1:t}$ to the perceived SU spectrum occupancy $y_{1:t}$:

$$\mathbf{E}_r = \begin{bmatrix} 1 - P_f(r) & P_f(r) \\ 1 - P_d(r) & P_d(r) \end{bmatrix} \quad r \in \{1, 2, .., R\} \tag{1.8}$$

### HMM Estimation and Forward-Backward algorithm

For a finite state space HMM, local prediction utilizes forward backward algorithm to estimate the joint posterior probability of primary user activity and secondary user's observation sequences $p(\hat{x}_{0:t}, y_{1:t-1})$ [52]. Define $\hat{x}_t$, $x \in [0, 1]$ as the *predicted* state value by SU:

$$p(\hat{x}_{0:t}, y_{1:t-1}) = p(x_0) \left[ \prod_{n=1}^{t} p(x_n|x_{n-1}) \right] \prod_{n=1}^{t-1} p(y_n|x_n) \tag{1.9}$$

The prediction problem can be formulated using the Bayesian notion $p(\hat{x}_t|y_{1:t-1})$

as the probability of the next state given a vector of past observations. Using the Bayesian definition of joint posterior probability relation:

$$p(\hat{x}_t|y_{1:t-1}) = \sum_{\hat{x}_t=i} p(x_t|x_{t-1})p(x_t|y_{1:t-1}) \quad : i \in \{1,2\}. \tag{1.10}$$

The last term in the equation above, represents the predictive posterior probability at time instant $t$ given the observation sequence. This posterior probability can be calculated using the Forward-Backward algorithm. The forward probability $\alpha_t(i) = p(y_{1:t-1}, x_t = i)$ is calculated recursively for a observation vector $y_{1:t}$:

$$\alpha_{t-1}(i) = p(y_{1:t-1}, x_{t-1} = i), \quad 1 \leq t \leq T-1 \tag{1.11}$$

$$\alpha_1(i) = p(x_0 = i)p(y_t|x_t = i) \qquad : i, j \in \{1,2\}$$

$$\alpha_t(j) = p(y_t|x_t = j)\left[\sum_{i=1}^{2} \alpha_{t-1}(i)p(x_t = j|x_{t-1} = i)\right] \tag{1.12}$$

$$p(y_{1:t}) = \sum_{i=1}^{2} \alpha_T(i) \tag{1.13}$$

While the backward probability $\beta_t(i) = p(y_{t+1:t}, x_t = i)$ the probability of observing all future events from this state. Since the initial state is assumed as given (i.e. the prior probability of this state = 1). The backward probability is calculated recursively for a observation vector $y_{t+1:1}$:

$$\beta_t(i) = p(y_{r,t+1:T}, x_t = i), \quad t = T-1, T-2, ..., 1 \tag{1.14}$$

$$\beta_T(i) = 1$$

$$\beta_t(j) = \sum_{i=1}^{2} \beta_{t+1}(j)p(x_t = j|x_{t-1} = i)p(y_{t+1}|x_t = j) \tag{1.15}$$

The Forward-Backward algorithm computes the expectation of how often each transition/emission is used, and repeats until convergence.

Staring with initial estimates of the transition and emission matrices, **Baum-Welch** training iteratively recalculates each matrix probabilities from the *training* sequence using (Equations-1.14 and Equation-1.11), then uses the outcome to estimate the state transition matrices. Baum-Welch algorithm is special case of the Expectation Maximization (**EM**) algorithm. HMM posterior probability is non-stationary random variable, and is a function of the observation series. Thus, analytical expression of the stationary conditional predictive posteriori (and hence error probability in Subsection-1.4.4) is counter-intuitive, and hard to obtain in a closed form for arbitrary HMM models [35, 53]. In Summary, HMM prediction follows the general steps (Fig.1.4):

- HMM Training: the observation sequence is used to train the HMM, and re-estimate the model parameters. The transition, and the emission probabilities are calculated using Baum-Welch algorithm.
- HMM decoding: to estimate the hidden state corresponding to the observed sequence. Forward algorithm is used to calculate the corresponding posterior probability of the hidden state given the observed sequence.
- HMM prediction decision is hypothesis testing of the SU predicted value $\hat{x}_t$
- Instantaneous error function calculation of the SU predicted occupancy value $\hat{x}_t$ against the actual system state $x_t$ (Subsection-1.4.4).

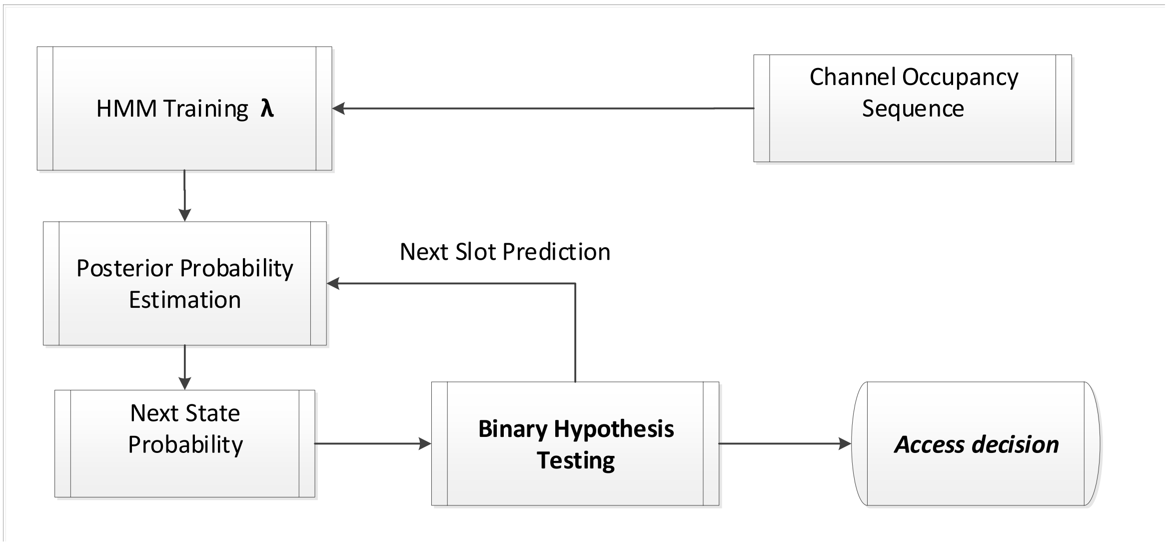


Fig. 1.4 HMM Prediction Flow chart

### 1.4.4 Mean Prediction Error

Under the two state ON/OFF model, prediction error is a Bernoulli random variable ($E_t$, where $E_t = \hat{x}_t \oplus x_t, E_t \in \{0, 1\}$). $\pi_t$ the error in prediction at time instant $t$, and the mean prediction error $\bar{pi}_e$ are defined as in 1.16.

$$\begin{aligned} \pi_t &= P(E_t = 1) \\ \bar{\pi}_e &= \lim_{t \to \infty} \mathbf{E}(\pi_t) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (\pi_i) \end{aligned} \tag{1.16}$$

The existence and equality in (1.16) requires the convergence of the quantity $p(\hat{x}_t|y_{1:t-1})$ according to *Cesáro Mean* (Theorem-1.4.1). Additionally, define the local (prediction) miss-detection $\bar{\pi}_d$, and false alarm probability $\bar{\pi}_f$ as:

$$\bar{\pi}_f = \bar{\pi}_e \, p(x_t = 0)$$
$$\bar{\pi}_d = (1 - \bar{\pi}_e) \, p(x_t = 1)$$

to indicate mean prediction error for busy and idle channels, respectively. Additionally, the conditional entropy rate of prediction probability (Theorem-1.4.2) is given by:

$$\mathcal{H}(\hat{x}_t|y_{1:t-1}) = \lim_{t \to \infty} \mathbb{E}_{p(\hat{x}_t, y_{1:t-1})} \log_2 \frac{1}{p(\hat{x}_t|y_{1:t-1})} \tag{1.17}$$

$$\mathcal{H}(\hat{x}_t|y_{1:t-1}) = \lim_{t \to \infty} \mathbb{H}(\hat{x}_t|y_{1:t-1}) \tag{1.18}$$

Using *Fano Inequality* (Theorem-1.4.3), the mean prediction error $\bar{pi}_e$ can be loosely bounded [38]:

$$\bar{\pi}_e \geq \frac{\mathcal{H}(\hat{x}_t|y_{1:t-1}) - 1}{\log(|\mathcal{R}|)} \tag{1.19}$$

Where $p(\hat{x}_t|y_{1:t-1})$, or using the tighter Fano bound (Theorem-1.4.3):

$$\mathcal{H}(\bar{\pi}_e) \geq \mathcal{H}(\hat{x}_t|y_{1:t-1})$$

The entropy rate for finite Markov processes is studied by Blackwell in [40] using a probability distribution on a Borel set of measures expressed using a integral that is hard to evaluate. A closed form expression for (1.16) requires the calculation of the limiting distribution of $p(\hat{x}_t|y_{1:t-1})$. Such expression requires the calculation of convergence of non-stationary Markov chain. Thus, mean prediction error required the complete characterisation of the predictive posterior probability ($p(\hat{x}_t|y_{1:t-1})$).

## 1.5 Research Questions and Contribution

Based on the literature review conducted in this work as presented in Chapter-2, three research questions were formulated. In general the literature review presented can be summarized into the following:

- Existing methods for statistical prediction.
- Existing work on spectral occupancy modelling.
- Existing limited work on spectrum occupancy prediction.

Based on the literature review the gaps in the research conduced so far and the scope for improving the techniques were identify, which are formulated as research questions as given below:

### Research Question 1: How accurate is Bayesian spectrum prediction with information uncertainty and error? What are the mean prediction error characteristics?

Under this research question, spectrum prediction under the family of Bayesian models is investigated. The performance of a secondary user prediction is studied against the error introduced by wireless channel, and sensing/sampling process. These errors can be quantified as detection error, and false alarm probability. The relation is studied for different scenarios of primary user activity patterns as well as channel conditions. Monte-Carlo simulations are utilized to investigate mean prediction error for HMM based spectrum prediction models. Then, a practical approximation of mean prediction error is constructed as a function of HMM model parameters. The expected deliverables outcome of this research question will include:

- A practical and tractable model selection framework based on sequential prediction theory.
- Numerical performance analysis of single user HMM based spectrum prediction.
- Numerical approximation of HMM prediction error as a function of model parameters.

### Research Question 2: How can multi-user cooperation improve the spectral prediction performance? How well Hard fusion based prediction fares compared to local spectrum prediction?

Under this research question, spectrum prediction with cooperative hard decision fusion rules is investigated. The impact of multi-user cooperation is studied against the error introduced by wireless channel. The relation is studied for different scenarios of primary user activity patterns as well as channel conditions. Monte-Carlo simulations are utilized to investigate mean prediction error for hard fusion based spectrum prediction models. Then, construct a practical approximation of mean prediction error is constructed as a function of HMM model parameters. The expected deliverables will out of this research question will include:

- Numerical performance analysis of cooperative Hard fusion based spectrum prediction.
- Numerical approximation of cooperative hard fusion based prediction as a function of model parameters.

### Research Question 3: How can multi-user soft combining improve the spectral prediction performance? How well soft fusion based prediction fares compared to local spectrum prediction?

Under this research question, spectrum prediction for cooperative soft fusion of spectrum decisions is investigated. The relation is studied for different scenarios of primary user activity patterns as well as channel conditions. Monte-Carlo simulations are utilized to investigate mean prediction error for soft fusion based spectrum prediction models. Then, alternative fusion rules based on HMM model parameters are proposed. The expected deliverables will out of this research question will include:

- Numerical performance analysis of cooperative Hard fusion based spectrum prediction.
- Propose alternative soft fusion based techniques based on HMM model parameters.

### 1.5.1 Contribution Summary

In this thesis, case study analysis is provided to evaluate the performance of the proposed approaches in different applications in wireless communication and cognitive radio. All the work in this research has been peer-reviewed and published, or submitted for publication. In summary, the findings of this thesis focuses on two major streams; firstly, single user (local) occupancy prediction model selection and performance is addressed. Secondly, cooperative spectrum prediction where the performance of decision (hard) based as well as data (soft) fusion is addressed. The main contributions of our work are presented in Chapter-2 - Chapter-5. Remarkably, the case study analysis confirmed that the statistical approximation is able to predict the performance of local and hard fusion cooperative prediction accurately, capturing all the essential aspects of signal detection performance, temporal dependency of primary user activity as well as the finite nature of the network.

In the first contribution, a consolidated framework based on sequential prediction theory, as well as a survey on current SOP models proposed in literature is presented. Based on an in-depth review of sequential prediction as well as spectrum occupancy, it is identified that prediction model selection is not instantly clear in SOP literature. 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, the fundamentals of statistical prediction are reviewed. Then, based on the stochastic mixture model framework, parametric and non-parametric approaches for underlying stochastic source assignment are reviewed. Secondly, spectrum occupancy prediction is described in terms of the stochastic class assignment. Mixture model formulation is extended to cooperative spectrum occupancy prediction using decision (Hard), and data (Soft) fusion techniques. Finally, theoretical and practical challenges of sequential spectrum occupancy prediction implementation are presented.

In the second contribution, performance analysis of HMM based occupancy prediction is addressed using Monte-Carlo simulation techniques. Firstly, the prediction error of one step-ahead (single time slot) prediction against the channel detection errors, as well as primary user's state transition probability is addressed to assess the model accuracy. Prediction error is also investigated against the observation sequence length to examine the correlation between prediction accuracy, and the number of samples required to calculate the next state probability. The prediction formulation is further examined for two step-ahead prediction assuming an incorrect one step-ahead prediction was made. Consequently, a new recursive equation to estimate HMM prediction performance as a function of channel detection errors is proposed based on HMM posterior probability. Finally, a new generalized Beta-Bernoulli approximation of the predictive posterior probability for local HMM based SOP is presented which provides a tractable expression of HMM based prediction performance.

The third contribution of our work put forth performance analysis of cooperative hard fusion based spectrum prediction. Specifically, hard fusion techniques are analysed for cooperative prediction based on channel detection errors to assess prediction gains of hard fusion. The performance analysis of local spectrum prediction using Monte-Carlo simulation techniques is further extended to hard fusion cooperative prediction. In particular, an analysis of secondary user's mean prediction error is presented in terms of primary user's activity pattern, and spectrum sensing errors. We utilize Bayesian filtering, and known information theory inequalities, to express cooperative prediction error bounds. Finally, a new generalized Beta-Binomial approximation of the predictive posterior probability for cooperative hard fusion based SOP is presented which provides a tractable expression of prediction performance.

The fourth contribution address soft decision based fusion for cooperative SOP. Monte-Carlo simulation performance analysis of local SOP is extended to soft fusion cooperative prediction . Soft fusion techniques are compared to local spectrum prediction, as well as benchmarked against hard fusion techniques. In particular, soft fusion superiority in terms of robustness as well as prediction accuracy is identified. Accordingly, alternative soft fusion techniques are proposed based on local prediction model parameters. The alternative techniques conceptually attempt to avoid common control channel requirements, while providing identical performance to known soft fusion techniques.
Throughout this research, a case study analysis is presented to evaluate the performance of the proposed approaches. In particular, analytical approaches and Monte-Carlo simulation results of the performance metric of interest are compared. Remarkably, the case study analysis confirmed that the statistical approximation is able to predict the performance of local and hard fusion cooperative prediction accurately, capturing all the essential aspects of signal detection performance, temporal depen-

dency of spectrum occupancy as well as the finite nature of the network.

### 1.5.2 Publications

- **H. Eltom**, S. Kandeepan, B. Moran, and R. J. Evans, “Spectrum occupancy prediction using a hidden Markov model” 2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS), Dec 2015 [2].

- **H. Eltom**, S. Kandeepan, Y. C. Liang, B. Moran, and R. J. Evans, “HMM based cooperative spectrum occupancy prediction using hard fusion,” 2016 IEEE International Conference on Communications Workshops (ICC), May 2016 [3].

- **H. Eltom**, S. Kandeepan, R. J. Evans, Y. C. Liang, and B. Ristic,"Statistical Spectrum Occupancy Prediction for Dynamic Spectrum Access: A Classification ", EURASIP Journal on Wireless Communications and Networking, February 2018 [4].

- **Hamid Eltom**, Sithamparanathan Kandeepan, Y.C. Liang, and Robin J. Evans,"Cooperative Soft Fusion for HMM based Spectrum Occupancy Prediction", Submitted to IEEE Communications letters.

- **Hamid Eltom**, Sithamparanathan Kandeepan, Y.C. Liang, and Robin J. Evans,"An approximation of stationary posterior distribution of HMM based Spectrum Occupancy Prediction", Submitted to IEEE Communications letters.

- Hourani, A., Evans, R., Sithamparanathan, K., Moran, W., **Eltom, H.** " Stochastic Geometry Methods for Modelling Automotive Radar Interference In: IEEE Transactions on Intelligent Transportation Systems, 2017 [5].

## 1.6 Thesis Structure

In Chapter-2 background information on cognitive radio, statistical prediction theory, and Bayesian mixture models are provided. Firstly, the fundamentals of statistical prediction are reviewed. Then, based on stochastic mixture models, parametric and non-parametric stochastic approaches for underlying stochastic source modelling are reviewed. Secondly, spectrum occupancy prediction is described in terms of these techniques. Cooperative spectrum occupancy prediction is studied for both decision (Hard), and data (Soft) fusion. Finally, theoretical and practical challenges of sequential spectrum occupancy prediction implementation are presented. This chapter introduces the first contribution in the form a survey on statistical prediction in cognitive radio literature [4].

In Chapter-3, local prediction performance for an HMM based predictors is presented. The prediction error performance dependency on HMM parameters is presented [2]. The analytical formulation of prediction error is formulated in a recursive equation [3] in conjunction with research question 2. The chapter also propose a new generalized Beta-Bernoulli approximation of the predictive posterior probability for local HMM based SOP models.

In Chapter-4, we present a contribution related to research question 2. In this chapter, we further extend the numerical performance analysis of local spectrum prediction, and address possible gains from cooperative spectrum prediction. In particular, an analysis of secondary user's mean prediction error is presented in terms of primary user's activity pattern, and spectrum sensing errors. We utilise Bayesian filtering, and known information theory inequalities, to express mean prediction error for single secondary user [3]. Then, the analysis of *Hard Fusion* based cooperative spectrum prediction is presented to highlight possible improvements of cooperative spectrum prediction.

In Chapter-5, we present a contribution related to research question 3. Soft decision based fusion for cooperative SOP is presented based on the local prediction model. The soft fusion based techniques are benchmarked based on hard fusion performance presented in Chapter-4. Alternative soft fusion techniques are proposed based on local prediction model parameters in Chapter-3. The chapter discusses prediction error analytical approximations for local prediction in Chapter-3, and cooperative prediction in Chapter-4. Finally, contribution summary and conclusion are presented in Chapter-6 along with future work.

# Chapter 2

# Prediction Model Classification

Our contribution in this chapter is a consolidated top-down classification of spectrum occupancy prediction. SOP taxonomy is presented 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, the fundamentals of statistical prediction are reviewed. Then, based on the stochastic mixture model framework, parametric and non-parametric approaches for underlying stochastic source assignment are reviewed. Secondly, spectrum occupancy prediction in terms of the stochastic class assignment. Mixture model formulation is extended to cooperative spectrum occupancy prediction using is formulated for decision (Hard), and data (Soft) fusion techniques. Finally, theoretical and practical challenges of sequential spectrum occupancy prediction implementation are presented.

In this chapter, Section-2.1- and Section-2.2 provide the background and problem components for sequential prediction problem. A brief review of empirical and statistical based approaches for SOP models is presented in Section-2.3. Then, a review of current spectrum occupancy techniques is presented in Section-2.4, Section-2.5. Lastly, the challenges in spectrum occupancy prediction are listed in Section-2.9. This chapter introduces the first contribution in the form a survey on statistical prediction in cognitive radio literature. The contribution of this chapter is published in the European Association for Signal Processing **EURASIP** journal on wireless communications and networking [4].

## 2.1 Background

Prediction theory asks the question: *Is it possible to forecast the short term evolution of an event? And if possible, how can we quantify the performance of this forecast?*, and quantify the prediction accuracy [35, 54]. *Sequential prediction* is deeply embedded in statistics [17], information theory [36, 55], machine learning [36, 48, 55, 56], source coding theory [56], and gambling [57] among many other disciplines. The term *prediction* in 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 [36]. A well known definition of the sequential prediction problem is [34–36, 54]:

> *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 $a_t = p_t(x_t|x_{0:t-1})$ 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 [34, 35].

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

[1]Probabilistic assumptions are made about the $\mathbb{M}$ sources prior, and under probabilistic action $a_t$ assumptions see [54, 58, 59].

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 $\mathbb{M}$ [31, 35]. 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 [54, 58, 59]. The universal predictor is not necessarily a member of $\mathbb{M}$ [59], but can be created as a mixture of predictor set $\mathbb{M}$ [33]. Universal prediction formulation can be summarised as:

> *Let $\mathbb{M}$ be an indexed set of arbitrary predictors. There exist prediction strategies for each sequence $x_{0:t-1}$ 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"* [54, 59].

In classic statistics [60, 61], the underlying stochastic process is assumed to be known. However, more recent approaches recognise unknown or non-existent underlying stochastic process [35]. Merhav *et. al* targeted a universal definition of prediction with emphasis on the *universality* of the predictor, while addressing the *triviality* of some predictor classes. Predictor choice is a trade-off between an optimal predictor that fits a specific set of observations, versus a universal predictor that accounts for all possible sets of observations. Prediction problem can formulated by two defining characteristics. The first is the underlying (*Known, Unknown, or non-existent*) stochastic process [54]. The second characteristic is the loss function that measure the accuracy of the prediction [38].

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 $k$th-order Markov models, or a set of auto-regressive models of order $p$ [35, 36, 54, 58]. 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 [35, 36, 54, 58]. 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 [35] 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 [31, 33, 35, 36]). However, they are by no means the only available methods, nor perform well for all arbitrary loss functions [54, 58, 59].[2]

## 2.2 Statistical Prediction

In broad terms, a sequential predictor is either fitted to the observation 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 [35, 36, 54, 58]. Statistical prediction is commonly presented under either probabilistic or deterministic settings. Prediction loss function, regret, and redundancy are discussed in Subsection-2.2.3, while Subsection-2.2.4 provides an overview of Bayesian based techniques.

### 2.2.1 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 $x_{0:t}$ [17, 60, 61]. Accordingly, optimal prediction is formulated as the minimisation of the expected value of the predictor loss function [31, 32, 34–36]. For example, if $\{X_t\}$ is an arbitrary parametrised random source, the action $a_t = \hat{x}_t$ is set as one step-ahead prediction, and the loss function is the squared distance $l(a_t, x_t) = (a_t - x_t)^2$ then the optimal predictor will always choose the conditional mean as it's predicted value. One of the most well known techniques that utilises this approach is the Kalman filter [48, 63, 64] (see Section-2.5). 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 $\mathbb{P}$ [58, 59].

[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 [54, 58, 59] for analysis, and [58, 59, 62] for Starkov codes, Hedge algorithm, and game theory approaches for sequential prediction.

### 2.2.2 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 *Auto-regressive* linear models (see Section-2.6) are commonly used for deterministic settings prediction. The second approach seeks a universal deterministic predictor. 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. Finally under most loss function assumptions, predictor design techniques for deterministic and probabilistic settings are *dual*, but may diverge for different loss functions (e.g. 0/1 loss function) [54, 59].

### 2.2.3 Loss Function and Regret

One step-ahead prediction commonly seeks the estimated state value at the next prediction slot $a_t = \hat{x}$. 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 choices for for next event prediction loss function, 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 [58, 59].

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 $n$ [34, 35]. 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 $\mathbf{H}(\mathbb{P})$ [34, 35]. In other words, no additional loss due to the use of $\mathbb{Q}$ instead of $\mathbb{P}$ [58, 59]. KL-divergence is commonly used to measure performance distance, and can be defined by the *cross entropy* between $\mathbb{P}$ and $\mathbb{Q}$ as:

$$d_t(x^{t-1}) := \log \frac{\mathbb{P}(x_t|x^{t-1})}{\mathbb{Q}(x_t|x^{t-1})}$$
$$D_n := \sum_{t=1}^{n} \mathbb{E}_{p(x_t|x^{t-1})}\{d_t\}$$

$\mathbb{E}\{..\} = \sum_{x^t \in \mathcal{X}^t} \mathbb{P}(x^t)[..]$, $d_t$ is the instantaneous Kullback-Leibler (**KL**) divergence, and $D_n$ is the total distance counterpart [35, 38]. Other possible choices for distance between $\mathbb{P}$ and $\mathbb{Q}$ are absolute, squared, Hellinger, and absolute divergence distances [31].

### 2.2.4 Bayesian Methods for Source Assignment

Bayesian *Mixture* models with self-information (entropy) loss are extensively studied in information and coding theory [31–33]. Bayesian algorithms are minimax optimal, and are universal under self information loss functions [34–36]. These methods perform well under both probabilistic and deterministic non-stochastic settings [31, 34, 35, 37, 38]. 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 $\{w(\theta), \theta \in \mathbb{M}\}$ [34]. 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 $\mathbb{M}$ [35, 36]. Using a non-negative normalised weighting function $w(\theta)$. The mixture model density function is defined as:

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

Upper, and lower loss bounds for Bayesian mixtures are defined using Minimax and Maximin approaches [34, 35]. However, the challenge in such models is the appropriate choice of the weights $w(\theta)$, i.e the prior distribution of the parameter $\theta \in \mathbb{M}$. Mixture models generally differ in terms of the size of the predictor index class $C$, stochastic class type $P_\theta$, and mixture prior $w(\theta)$. Different mixture models can be grouped into the four generic approaches:

### 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 $\theta$. 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 [35, 48]. However, plug-in approaches are heuristic and lack theoretical justification [35, 36].

### 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).

$$\mathbb{Q}_w(x^t) = \sum_{i=1}^{C} P_{\theta_i}(x^t) w(\theta_i)$$

Expectation-Maximisation *EM* algorithm is used to estimate the parameter set $\theta \in \mathbb{M}$ [48, 65, 66].

### 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:

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

$h > 0$ 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 [48, 65, 66].

### Infinite mixture models

When the class $\mathbb{M}$ size is infinite, the prior distribution on $\theta$ 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 $\alpha$ is concentration parameter, and $G$ is the distribution over $\theta \in \mathbb{M}$. Samples of $\theta_t$ 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 [65–67].

## 2.3 Empirical Spectrum Occupancy Prediction

Spectrum occupancy prediction models broadly target 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* [1, 18]. Measurements on spectrum occupancy show that spectrum prediction is necessary to improve spectrum utilization efficiency (Fig-2.2). The common motivation for SOP techniques is to minimise the accumulated *time delay* due to cognitive cycle processing. By predicting the channel status in advance, more processing time becomes available for spectrum sensing, decisions, and mobility [10]. SOP models address *prediction* either *explicitly*[20, 24, 68], or *implicitly*. Implicit approaches present SOP models as primary/secondary user activity models. In this review, both implicit, and explicit formulations are addressed as *statistical* SOP models. Statistical SOP models proposed for spectrum occupancy analysis include Poisson processes [22, 23], Bayesian prediction [24, 25], and linear regression [26, 27]. Machine learning based techniques have also been proposed for model *learning* including neural networks, time regression, and space vector machines [10, 28, 29]. The surveys in [10, 11] provide a good taxonomy of primary user's activity model collection. This review abstracts and consolidate SOP models in DSA systems, and extends the aforementioned works.

The flow chart in Fig-2.1 highlights the temporal sequence of spectrum occupancy prediction process presented in this section. The flowchart is reused from Chapter-1, where it was used to highlight spectrum sensing/prediction interrelation. This section focuses on model selection, while the next three sections address selected model classes. Current spectrum occupancy prediction techniques are presented 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.

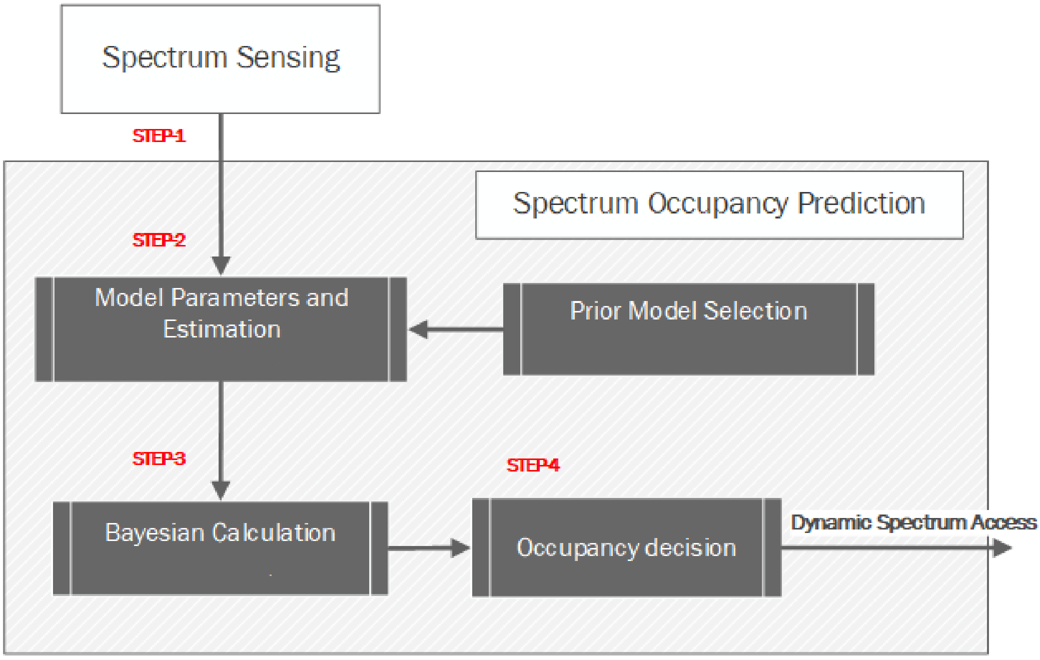


Fig. 2.1 Spectrum occupancy prediction flowchart (this figure is repeated for the convenience of the reader)

### 2.3.1 Spectrum Measurement campaigns

A spectrum measurement campaign is an empirical data collection conducted for specific scenarios (indoor/outdoor) to collect spectrum occupancy samples on pre-selected frequency bands (television white bands/cellular bands). Statistical analysis, and estimation are conducted to generate an approximate statistical description of average power, or channel occupancy. The campaign captures real life spectrum occupancy scenarios, it's 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 [1, 18, 69, 70]. Campaigns in Hong Kong in [70], and Melbourne [1] assessed spectrum occupancy patterns for a large section the of radio spectrum. The survey by Chen el. [18] provides an intensive review of several measurement campaigns for selected wireless communication technologies.

Fig.2.2 presents raw spectrogram results of spectrum monitoring experiment conducted in three different urban environments in Melbourne metropolitan [1]. The spectrum campaign addressed spectral allocation for cognitive radio Device-to-Device communications and small cell networks. The spectrum occupancy is quantized 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 ranges 402 – 460 MHz, and 520 – 820 MHz (vacated analogue TV band) are suitable candidates for DSA applications [1].

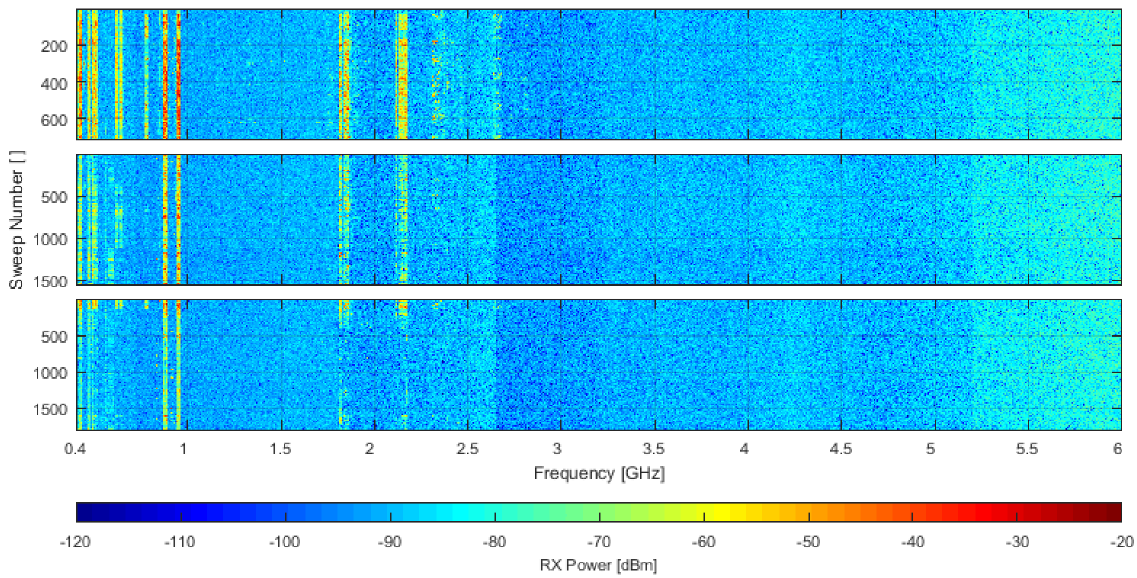


Fig. 2.2 Power measurement campaign sample for Melbourne LTE system measurements [1]

### 2.3.2 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 [20, 71], Poisson processes [22, 23], Bayesian models [24, 25], neural networks [10, 44, 68], linear regression [26, 27], space vector machine [72], pattern mining [73, 74], and dictionary based prediction [24]. In a sequential prediction framework, these techniques represent different parametrised predictor classes. Models generally takes several categories with major literature focus on:

- Bayesian Models (including HMM) [25, 27, 71, 75–83]
- Queuing theory and Poisson Process [22, 23]
- ON/OFF models [84]
- Dictionary based [19]
- Neural networks [21, 26, 72, 85]
- Linear Regression [25, 73]
- Space Vector Machine[86]
- Pattern Mining [87]

These models assume a specific probability distribution based on the application e.g. (inter-arrival time as Poisson process [22]) and attempts to identify its parameters. Other models learn the information from training data and/or in an on-line manner[10, 17, 29].

### Prediction Model Selection

Parameters studied by spectrum occupancy modelling are 1) *channel status* i.e. prediction of the spectrum status as idle or busy, 2) and *duty cycle* i.e. prediction of average fraction of time the spectrum channel is occupied, or 3) 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. $a_t = p(x_t|x^{t-1})$. 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 users's throughput, spectrum interference and wastage,

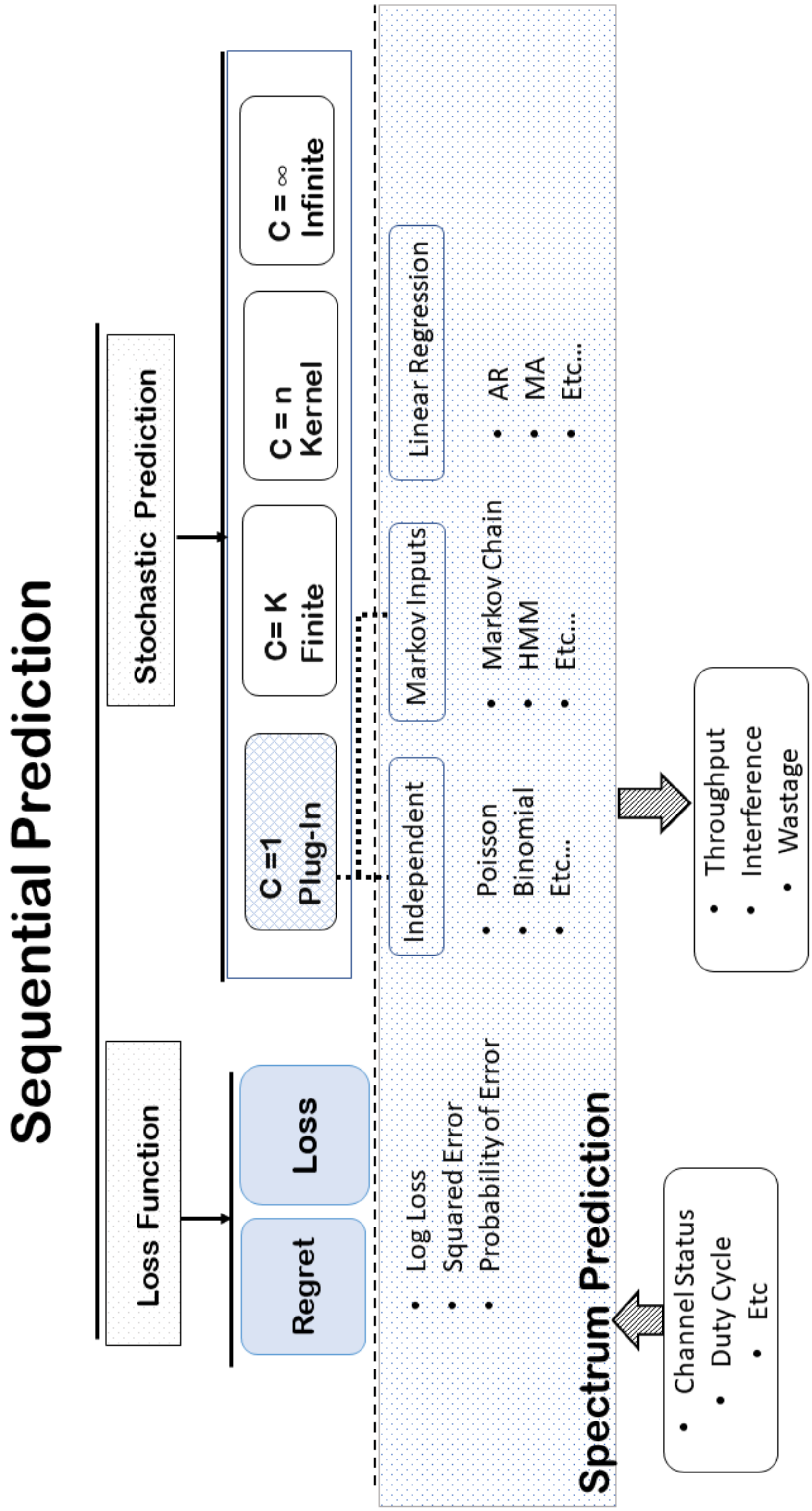


Fig. 2.3 Sequential prediction classification of spectrum prediction techniques

and probability of error (or mean square error) are generally defined based on the probability density of the one step-ahead prediction. For example, the probability of incorrect prediction of an available spectrum hole generally describe spectrum interference or spectrum wastage [88].

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 [35, 36, 54, 58].

Table.2.1 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. 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.

### Prediction Models classification

By dissociating the implementation requirements and assumptions from the stochastic components of the spectrum prediction model, the author's distinguish four 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. Auto-regressive (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.2.2 highlights few major advantages, and disadvantages of different spectrum occupancy prediction categories. For example, stochastic memoryless modes ignores temporal correlation of the observation series, but could be 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. Similarly, 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.

Fig.2.3 summarizes the sequential prediction theory presented in Section-2.2, and maps current SOP techniques available in literature. The number of mixture sources $C$ in the replacement source assignment $Q$ differentiate mixture models (Subsection-2.2.4). The figure conceptually illustrate the modelled occupancy series as an input, where the selected mixture model produces the desired performance measure based on the selected loss function. A review of current spectrum prediction techniques is presented for each category in the next three sections. Section-2.4 presents single memoryless source approaches, Section-2.5 handles Markov based models, while Section-2.6 presents linear statistical regression based prediction.

Table 2.1 Summary of spectrum prediction technique

| Category | Model | Research Works | State Space / State Dependency | Occupancy Decision Criteria |
|---|---|---|---|---|
| Memoryless stochastic source models (Section-2.4) | Bernoulli/ Binomial | [89–91],[92–94] | $x \in [0, 1, .., S]/p(x_t)$ | Channel status |
| | Poisson | [95–100] | | |
| | Exponential | [24, 93, 95, 96] | $x \in \mathcal{R}/p(x_t)$ | Duty cycle |
| | Log-Normal | [69, 101] | | Signal/Power |
| | Uniform | [102] | $x \in [0, 1, .., S]\,, x \in \mathcal{R}/p(x_t)$ | |
| Finite order Markov models (Section-2.5) | 2-State Markov chain | [95, 103–106] | $x \in [0, 1]/p(x_t\|p(x_{t-1})$ | Channel status |
| | 3-state Markov chain | [96] | $x \in [0, 1, 2]/p(x_t\|p(x_{t-1})$ | |
| | High-order Markov chain | [107] | $x \in [0, 1]/p(x_t\|x^{t-m})\,, m > 1$ | |
| | Semi-Markov | [108, 109] | $x \in [0, 1, ...]/p(x_{s+t}\,, s > 0\|x_t)\; s \in \mathcal{R}$ | Duty cycle |
| | Continuous time MC | [110–112] | | |
| | Hidden Markov model | [24, 25, 76, 79, 81, 82, 113, 114] | $x \in [0, 1, ..., S]/p(x_t\|x_{t-1})$ | Channel status |
| | Bayesian models | [115, 116], [43, 46] | $x \in \mathcal{R}/p(x_t\|x_{t-1})$ | Signal/Power |
| Finite order linear regression models (Section-2.6) | Auto regressive | [117–120] | $x \in [0, 1]\; or\; x \in \mathcal{R}/p(x_t\|x_{t-1}, .., x_{t-m})$ | Channel status, or Signal/Power |
| | Moving average | [118] | | |
| | ARIMA | [98] | | |
| | Random walk | [92] | | |
| Machine learning statistical based techniques (Subsection-2.7) | Neural networks | [75, 78, 121–124] | $x \in [0, 1]\; or\; x \in \mathcal{R}/p(x_t\|x^{t-1})$ | Duty cycle, Channel status, or Signal/Power |
| | Support vector machine | [72] | | |
| | Pattern mining | [73, 74] | | |

Table 2.2 Comparison of Spectrum prediction categories

| Category | Advantages | Disadvantages |
|---|---|---|
| Memoryless stochastic source models (Section-2.4) | Low complexity<br>Closed form solution for sparse spectrum usage scenarios<br>Easier/convenient model to adopt | Limited to sparse spectrum usage<br>Limited to single primary user scenarios<br>May not describe real world channel occupancy status |
| Finite order Markov models (Section-2.5) | Expandable to various PU/SU scenarios<br>Applicable to heavy-tailed channel traffic<br>Higher accuracy with manageable number of parameters | Require sufficient measurements for model training<br>Complexity depends on the order of the model |
| Finite Order Linear Regression Models (Section-2.6) | Expandable to various PU/SU scenarios<br>Approximation of probabilistic model to linear equation model | Expandability increases the number of model parameters<br>Require sufficient measurements for model training |

## 2.4 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 $x_{1:t-1}$ 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 [95, 96]) or used to estimate the stochastic source probability density function $\mathbb{Q}_w$ from a training sequence. Models adopted in SOP proposals include:

**Bernoulli trial process** is the mathematical abstraction of repeated coin tossing. The random variable $x_t$ takes only the values 0 or 1 representing failure and success, respectively. The series $x_1, x_2...x_{t-1}$ is assumed to be independent, and identically distributed Bernoulli random variables, with probability mass function parametrised by $\rho$ [48, 125]:

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

Where $k$ is the number of trials, and $\rho$ 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 $\rho$ 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 $\lambda = \dfrac{k}{n}$ [48, 125]:

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

**Exponential distribution:** The interval between events in a Poisson distributed process follows the negative exponential distribution parametrised by $\lambda$, with probability density function [48, 125]:

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

In spectrum occupancy literature, memoryless sources are not used as often for one-step ahead prediction. However, this class of stochastic sources is frequently used to describe primary user activity. Bernoulli process have been proposed in [89–91] to describe ON/OFF spectrum occupancy in spectrum sensing/access proposals. Similarly, Poisson process have been proposed in [95, 99] (2 States), and [96] (3 states) to model the arrival/departure process of the primary user. Exponentially distributed duty cycle models were presented based on queuing theory in [95, 96]. Similarly, proposals in [97, 98] suggested a non-exponential service time as a result for multiple primary users scheduling. In [24] an exponential distribution to model inter-arrival time of the primary users was proposed to design a secondary user contention algorithm. Joint cognitive radio spectrum sensing and prediction model in [93] 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 [69, 101], uniform distribution [102], and binomial distribution [93, 94]. The choices were generally motivated by physical layer assumptions.

## 2.5 Finite order Markov Models

In this category, 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. $p(x_t|x_{1:t-1}) = p(x_t|x_{t-1})$ is called *Markov property* [48]. Markov-based construction is attractive due to the desirable convergence properties of Markov chain based models [39–42]. *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. $x_{1:t-m}$, or *Explicit duration* Markov chain (a form of continuous-time Markov chain), where the time spent on each state is not exponentially distributed [48, 52]. 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[126]:

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

The first equation is Chapman-Kolomogrov prediction equation, the second is Bayes rule update, while the last equation is the normalisation factor [126]. This model is labelled *doubly stochastic* as it accounts for measurement error in *observing* $x_{1:t-1}$ by defining the observation series $y_{1:t-1}$, where $x_{1:t-1}$ 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 $[y_t = \mathbf{h}_t(x_t, \mathbf{u}_t)]$, where $\mathbf{h}_t$ is a non-linear function, and $\mathbf{u}_t$ is an independent additive noise source (measurement error)[48]. The observation variable is distributed according to $p(y_t|x_t)$, defined as the observation likelihood probability. The conditional posterior probability $p(x_t|y_{1:t-1})$ 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 $p(x_t|y_{1:t-1}) = p(x_t|x_{t-1}, y_{1:t-1})p(x_{t-1}|y_{1:t-1})$ (Markov property). When implementing such model, the density $p(x_t|y_{1:t-1})$ is either estimated using the prior/likelihood function, or using kernel density estimation [64, 126].

**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 $t-1$ [48, 126–128].

**Hidden Markov Model** **HMM** is partially observable Markov chains i.e. observing a Markov chain through a noisy channel [48, 52]. 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. Fig-2.4 displays a snapshot of HMM state transition (connected lines).

**Kalman filter** is the optimal solution for linear Gaussian state space Markov based models [48, 64, 126]. Non-linear predictors are often a sub-optimal variation of Kalman filter, such as extended Kalman filter, and unscented Kalman filter [48, 63, 64].

**Bayesian Particle Filters** Particle filter methods utilise Monte-Carlo Markov chain (MCMC) to approximate the conditional posterior probability assignment $p(x_t|y_{1:t-1})$, or the full posterior probability $p(x_{1:t}|y_{1:t-1})$. They utilise either weighted samples of a plug-in probability assignment based on prior/likelihood, or a mixture model based density [126].

In spectrum modelling literature, Poisson Markov chain based proposals in [95, 96] studied primary user interference and wastage. Two-state[2, 103–106], and three state discrete-time Markov [96] chain have been proposed to model the primary-secondary users stochastic behaviour. Similarly, higher-order Markov chains in [107] were used to detect the primary user traffic pattern. Explicit duration semi-Markov chains with generalised distribution of duty cycle time modelled primary users's inter-arrival time in [108, 109], while continuous time Markov chain modelled primary user behaviour in [110, 111]. Moreover, hidden Markov model received wide attention in spectrum occupancy prediction literature [2, 3, 24, 25, 76, 79, 82, 82, 113]. 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 [112]. K-step ahead prediction was studied in [81, 114] assuming a non-stationary HMM. Finally, works in [43, 46] utilised regularised particle filters with Kernel density estimation to model primary user activity in multi-primary and secondary user cases.

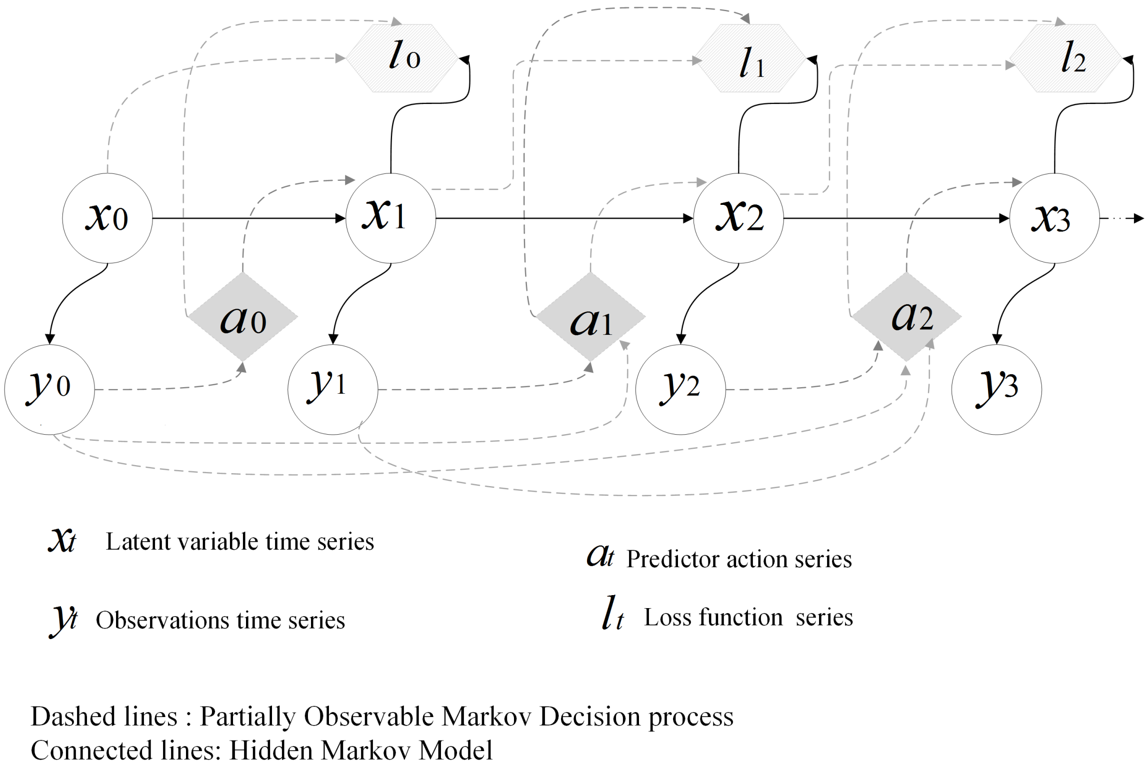


Fig. 2.4 Statistical prediction using Markov based models

## 2.6 Finite Order Linear Regression Models

This category is a special case of the general non-linear statistical regression for $p(x_t|x_{1:t-1})$. Linear regression models focus on the linear dependency between the random variables $x_t$, and $x_{1:t-1}$ [48]. Auto-regressive model $AR$ ($p = 0$), and Moving Average $MA$ ($q = 0$) are special cases of Auto-Regressive Moving-Average **ARMA**. ARMA model ($ARMA(p, q)$) can written as [48, 125]:

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

Where $c$ is a constant that can be replaced with $\mu = \mathbb{E}_x\{x_t\}$. $\eta_t$ is a noise random variable that represents the uncertainty in sampling. $\phi_i, \theta_i$ are the auto-regressive, and moving average parameters. $p, q$ are the order of the autoregressive, and moving average components. Auto-Regressive Integrated Moving-Average ($ARIMA$) process generalises the ARMA model to $ARIMA(p, d, q)$, and written as [48, 125]:

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

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 auto-regressive with Gaussian distributed random variables was used to model spectrum occupancy in [117–119]. Similarly, moving-average [118], and ARIMA [98] were proposed for spectrum occupancy status modelling. Random walk model was proposed in [92] to model spectrum occupancy duty cycle. Finally, an auto-regressive model of decimal equivalent of a binary series model was proposed for primary user activity in [120].

## 2.7 Machine Learning based Techniques

This category introduced in Section-2.3 covers techniques that utilize machine learning based algorithms for spectrum prediction which is was not covered directly under the categories described in previous sections. Machine learning, data mining, and pattern recognition algorithms are based on existing statistical inference models. Kobayashi *et. al* [[48], 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 [122]. *Neural networks* based techniques are presented extensively in cognitive radio networks [29, 122–124, 129, 130], with application on spectrum prediction presented in [78, 121]. Support vector machines [72], pattern mining [73, 74], and Dictionary based prediction [24] were suggested for spectrum prediction and user activity modelling. The surveys in [28, 29, 129] discuss artificial intelligence, and machine learning applications for dynamic spectrum access networks.

## 2.8 Extending Occupancy Prediction Formulation

The action $a_t$ in the sequential prediction definition is not limited to the next state event $\hat{x}_t$, or probabilistic representation $p_t(x_t|x_{1:t-1})$. A spectrum access based loss function assignment for the single user's prediction action $a_t$ transforms the prediction problem to a decision problem. A popular choice for spectrum prediction/access problems is *reinforcement learning* based algorithms. Reinforcement learning is a trial and error learning procedure that awards/punishes the agent based on the outcome of trial. The agent learns system dynamics, and constructs the action policy through interaction with environment. In a single agent scenario, a Markov decision policy based algorithm is commonly adopted [131] (See Fig-2.4). While in the multi-agent scenario a stochastic game is formulated [132].

In spectrum occupancy modelling, reinforcement learning was proposed to model dynamic spectrum spectrum access policy. Research proposals in spectrum prediction/access based on reinforcement learning focused on optimising access to vacant spectrum opportunities [88, 98, 133]. Wang *et. al* [98] studied a jamming scenario during spectrum access under a spectrum sharing game. Studies in [134, 135] developed secondary user access algorithm based on an auction game. In [136, 137] a partially observable Markov model formulation was developed to model spectrum sensing/access. A limited number of proposals [136, 138–141] considered utilising both observation series $x_{1:t}$ and action series $a_t$ as inputs to spectrum prediction/access algorithm.

## 2.9 Spectrum Occupancy Prediction Challenges

The survey in [11] discussed the issue of occupancy modelling validity based on the type, and amount of traffic pattern. The work presented several scenarios of possible implementation issues for primary user modelling. This section extends the results and address theoretical challenges for SOP model implementation.

### 2.9.1 Validity and Complexity

Valid spectrum occupancy observation representation is limited by state space dimensionality. Spectrum samples have temporal, spectral and spatial dependency. Proposed spectrum occupancy prediction models simplifies the assumptions about spectral, and spatial dimensions 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 occupancy pattern. Few spectrum measurement campaigns invalidated several short term prediction assumptions. Thus, validation through empirical spectrum campaigns is essential for any spectrum predictor design [1, 20]. For example in [142], 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 [35, 36, 54, 58]. Model complexity limits the applicability of the prediction model. 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 [31, 33, 35, 36]). 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 based source estimation. For example, Dirichlet mixture process is often used to generate prior distributions, however tracing convergence bounds becomes increasingly difficult [31, 35, 65]. Convergence bounds are calculated only for limited Bayesian mixture class/prior distribution pairs (for example, uniform prior/Epanchinkov kernel) [143].

### 2.9.2 Cooperation and Contention

Cooperative spectrum prediction faces the practical issue of common control channel design [46]. The amount of data shared between users sets a trade-off between spectrum prediction accuracy, and control channel capacity [46]. Common control design trade-off for cooperative spectrum prediction in a multi-primary user environment is yet to fully develop in 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 [65].

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 [88, 144]. However, the study in [144] 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.

## 2.10 Summary

Based on parametric and non-parametric mixture model framework, this chapter 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 learning 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. The contribution of this chapter is accepted for publication in the European Association for Signal Processing **EURASIP** journal on wireless communications and networking [4].

# Chapter 3

# Local Spectrum Occupancy Prediction

This chapter presents single user occupancy prediction model based on hidden Markov model. Local spectrum prediction using hidden Markov model is proposed and elaborated Section-3.2. Then, we present an analytical approximation of HMM prediction error performance. The approximation is motivated by works in [35, 37, 145] on the recursive posterior estimation of finite state Markov machines. Finally in the case study analysis, the prediction error of one step-ahead (single time slot) prediction against the channel detection errors, as well as primary user's state transition probability is presented to assess the model accuracy. Prediction error is also investigated against the observation sequence length, to examine the correlation between prediction accuracy, and the number of samples required to calculate the next state probability. We further examine the prediction of two step-ahead prediction assuming an incorrect one step-ahead prediction was made.

The contribution of this chapter is partially published in the conference proceedings of IEEE International Conference on Signal Processing and Communication Systems **ICSPCS'2015** [2]. The remaining contribution of this chapter is submitted to IEEE communication letters (Subsection-1.5.2). The contribution proposed a new generalized Beta-Bernoulli approximation of the predictive posterior probability for local HMM based SOP described in this chapter.

## 3.1 Background and Related Work

Current SOP literature explores prediction techniques such as Bayesian models (including HMM) [19, 25, 76, 79, 82], neural networks [21, 44, 85], linear regression [26, 27], space vector machine [72], and pattern mining [73]. Bayesian-based prediction techniques provide a powerful, and a flexible tool to measure and model RF environment. Secondary users within the cognitive radio network collect sensing information, and utilize statistical correlation to infer possible future latent states of the primary user usage patterns.

Hidden Markov Model **HMM** lends its name to two defining properties: Markov property that the current state $x_t$ is dependent on past state $x_{t-1}$, and the state space is assumed hidden from the observer "$r$" with only the observation $y_{r,t}$ available at time instant $(t)$ [52]. Hidden Markov model is the basis of local spectrum prediction model in this chapter. Subsection-1.4.3 introduced HMM model and paramters. The next Subsection-3.1.1 provide the related work in spectrum occupancy research using HMM based models. The remaining Subsection presents the background work for HMM posterior estimation (HMM prediction performance) presented in Section-3.3.

### 3.1.1 HMM for spectrum occupancy modelling

HMM has been used to model the channel occupancy in several studies for local spectrum prediction. The proposal in [82] presented a modified bivariate HMM to improve sensing performance assuming Gaussian distributed observations. The results presented *dwelling* time of k-steps ahead performance comparison against standard HMM assumptions. In [81, 82] authors studied the performance of dwelling time prediction for a non-Stationary HMM proposal. Similarly, the work in [29] presented HMM prediction accuracy for different training algorithms, and used it to analyse secondary user throughput. The paper in [79] simulated combined HMM, and frequency hopping performance for secondary user throughput. The efforts in [75, 76] simulated the prediction error for multiple slots, considering a modified HMM which accounted for the delay in obtaining observations. While several research papers simulated these models under different traffic and channel conditions, the HMM prediction accuracy is not immediately clear. The functional relationship between HMM prediction error and the state transition, and/or channel error is not intuitively available in literature.

Due to the non-stationary nature of state estimation in HMM, the stationary state distribution is often estimated numerically [37, 145]. The works in [2] defined the minimum prediction error (ideal sensing scenario) as lower limit for HMM based SOP performance, and addressed local prediction single user error. The work in [3] extended the mean prediction error, and presented a binomial approximation as a lower limit of hard fusion prediction error. However, the analytical expression of the stationary conditional predictive posterior is counter-intuitive, and hard to obtain in closed form for arbitrary HMM predictors [35, 146]. Thus, local HMM predictor performance is not available in a closed analytical form.

### 3.1.2 HMM posterior probability estimation

Estimation and analysis of the finite Markov Machines **FSM** (including HMM) posteriori distribution is not available in a closed form for arbitrary process assumptions. Goldsmith el.[37] studied the capacity of finite state Markov process with independent inputs. The work proposed an algorithm to recover the capacity estimation loss due to the conventional methods using the combination of memoryless channel coding, and interleaving. Earlier work by Mushkin el. [145] focused on the same question for the spacial case of Gilbert-Elliot Channel. Each pioneer paper provided a recursive estimation equations of the posteriori probability of channel state. In FSM literature, the works in [31, 147] addressed the convergence to lower bound of FSM prediction. The proposals of Arimoto-Blahut based methods [148], and run-length error statistics[149] presented efficient recursive techniques for FSM lower bound convergence of arbitrary FSM models. Additionally, recursive formulations of FSM conditional posterior under various assumptions are presented in works such as Gilbert-Elliot channel [35, 150], FSM channels with i.i.d inputs [37], and non i.i.d input [53].

Fundamentally, sequential predictor stochastic probability assignment can be constructed as a mixture model without loss of optimality by confining the predictor distribution selection to a specific class (family of distributions) [4, 35] (Section-2.2). Thus, HMM stationary posterior can be constructed from the same family of Markov chain stationary posterior. In Section-3.3, we present an approximation of HMM prediction error performance. The approximation is motivated by works in [35, 37, 145] on the recursive posterior estimation of finite state Markov machines.

## 3.2 Local SOP Model Formulation

In this section, we formulate the *local* spectrum prediction based on hidden Markov introduced in Chapter-1. The model describes the primary user activity model (Subsection-3.2.1), channel propagation characteristics (Subsection-3.2.2), occupancy detection (Subsection-3.2.3), and local prediction models (Subsection-3.2.4).

In a discrete slotted time system, $R$ secondary users are assumed independently and identically distributed according to a Poisson point process in Euclidean space $\mathbb{R}^2$. The wireless channel between the primary and secondary users is modelled by generalised path-loss, log-normal shadowing, and additive white Gaussian noise model [46]. PU activity is modelled as a fully observable Markov chain, while SU observation model is a partially observable Markov chain (HMM). Given an observation sequence, HMM prediction follows the general steps shown in Chapter-1 (Fig.1.4), and described in the proposed prediction algorithm (Algorithm-1). It worth noting that in the case study analysis section (3.4), the training phase has been bypassed to solely measure the accuracy of HMM spectrum prediction apart from any error that might be introduced by Baum-Welch training algorithm (Subsection-1.4.3) [52].

### 3.2.1 Primary User Activity Model

Define $\{x_t : t \geq 1\}, x \in \{0, 1\}$ as the channel state due to PU activity at time slot $t$ due to primary user's activity as irreducible stationary Markov chain. The transition probability matrix between states $i, j \in [0, 1]$ i.e. $p(x_t = j|x_0, .., x_{t-1} = i) = p(x_t|x_{t-1} = i$ is given by [48]:

$$\mathbf{P} = [p_{ij}] = \begin{bmatrix} \theta & 1-\theta \\ 1-\mu & \mu \end{bmatrix} \tag{3.1}$$

Where $\mathbf{P}$ is a transition probability matrix for all $t$, and $(\mu, \theta)$ represent the probability of remaining in busy and idle states, respectively. Primary user's occupancy model based on two state fully observable Markov chain is parametrised by $\{v, \mathbf{P}\}$.

**Algorithm 1** Proposed Prediction Algorithm

**Begin While** (*Next Slot Prediction* is NOT NULL)

**Step-1**: *HMM Training*: Train the model using $y_{r,1:t-1}$ .

Estimate the model parameters $(\hat{\mu}, \hat{\theta}, \hat{P}_m(or\ \hat{P}_d), \hat{P}_f\ )$, using Baum-Welch algorithm [47] (Subsection1.4.3).

**Step-2**: *HMM posterior probability calculation*: Using Forward-Backward algorithm (Subsection1.4.3) [52] estimate the posterior joint probability of the primary user occupancy $p(x_t|y_{r,1:t-1})$ using $(\hat{\mu}, \hat{\theta}, \hat{P}_m, \hat{P}_f)$

**Step-3:** *Spectral Occupancy Prediction:* calculate the hidden state corresponding to the observed sequence $p(\hat{x}_{r,t}|y_{r,1:t-1})$ (Subsection1.4.3).

**Step-4:** *Spectral Occupancy Decision:* Decide the most probable next occupancy state $(\hat{x}_{r,t}$ against threshold $\zeta$.

**Step-5:** Pass $\hat{x}_{r,t}$ to Dynamic access decision mechanism (Equation-3.12).
add $\hat{x}_{r,t}$ to state sequence $y_{r,1:t-1}$, and go to **Step-2** for 2-steps ahead prediction

**End While**

### 3.2.2 Channel Propagation Model

The power measurement by $r$th secondary user $\Phi_{z,r,t}(dB)$ at time slot $t$ is modelled using power-distance decay law, and log normal shadowing [45]

$$\Phi_{z,r,t}(dB) = 10 \times \log_{10} \frac{1}{T} \int_T |s_t(\tau)|^2 d\tau + g_r(d) + \mathcal{X} \tag{3.2}$$

where $s_t$ is the transmitted symbol for channel state $x_t$, and $X$ is Log-normal distributed random variable with zero mean, and variance $\sigma_v^2(dB)$ [46]. The $g_r(d)$ is a generalized power decay model, given the relative distance $d$ between the primary and secondary users [151]

$$g_r(d) = 10 \times \log_{10}(\frac{1}{||d||^\alpha}) \tag{3.3}$$

where $\alpha$ is the path-loss exponent, and $d$ is the distance between primary and secondary users.

### 3.2.3 Occupancy Detection Model

The received signal $z_{r,t}$ is detected by a complex baseband equivalent of energy detector test is used by each SU. The two hypothesis of *present*, and *absent* signals for an observation period of $T$ or the equivalent of $N$ samples for each time slot $t$. Test statistics $\Phi_{z,r,t}$ for time slot $t$ for a large number fo samples $N$ is assumed to be Gaussian distributed (Central limit Theorem) [45].

$$\Phi_{z,r,t} \sim \mathcal{N}(\sum_{n=1}^{N} \mathbb{E}[|z_{r,t}|^2], \sum_{n=1}^{N} \mathbb{V}[|z_{r,t}|^2]) \tag{3.4}$$

The channel Noise is assumed to be Gaussian distributed with zero mean and variance $\sigma_w^2$, and the signal to noise ratio $\rho_r = \sigma_{z,r}^2/\sigma_w^2$ where $\sigma_{z,r}^2 = \sum_{n=1}^{N} \mathbb{V}[|z_{r,t}|^2]$. Hypothesis testing on the detection statistics yield the the series $y_{r,t}$ i.e the occupancy perceived by each SU. The uncertainty around spectrum sensing performance is quantified by the probability of correct detection $P_d(r)$, and the probability of false alarm, and $P_f(r)$ can be defined using central limit theorem approximation for large number of samples $N$ as described in Equation-1.8.

The probability of detection $P_d = 1 - P_m$ is the probability of observing the channel correctly as busy, while the probability of false alarm $P_f$ is the probability of observing occupied channel while the primary user is idle.

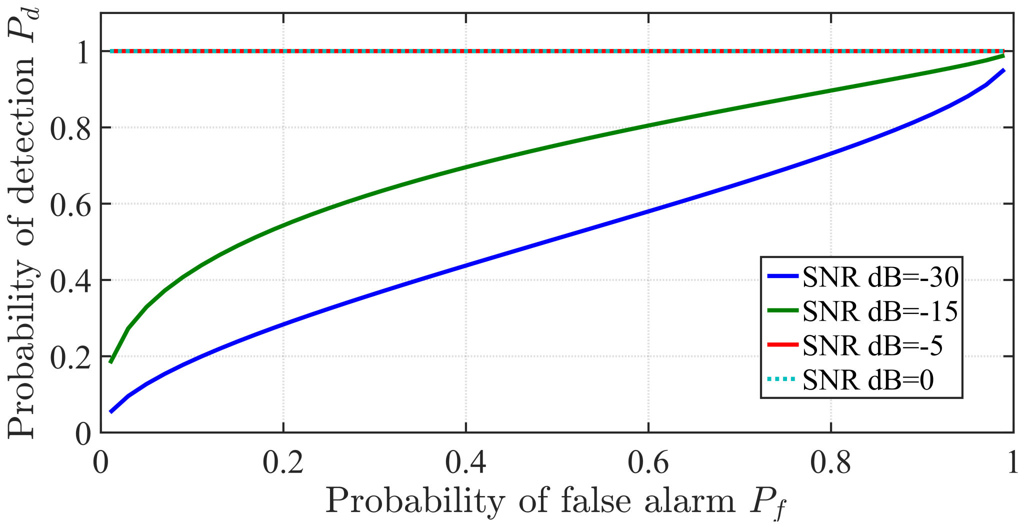


Fig. 3.1 Energy Detector ROC Curve for different values of $\rho_r$ $(N = 1000)$

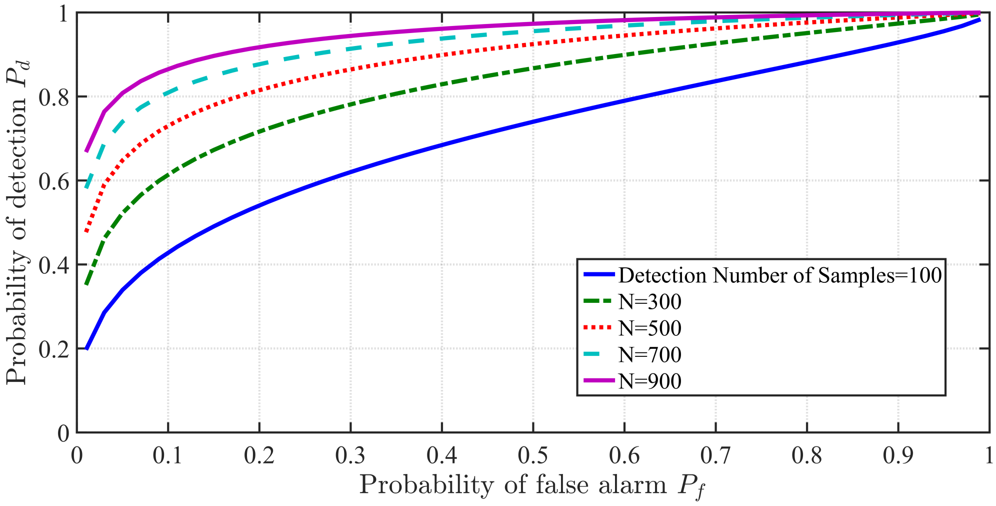


Fig. 3.2 Energy Detector ROC Curve for different values of $N$ ( SNR $\rho_r$ dB = -10)

Secondary user's model is HMM with transition $\mathbf{P}$, Emission $\mathbf{E}_r$ matrices, and initial state distribution $p(x_0)$. Using constant false alarm based approach (**CFAR**), the threshold $\lambda$ is calculated, for a large number of samples $N$ using inverse Q-function. Given PU channel activity $x_{1:t}$ to the SU spectrum perceived occupancy $y_{r,1:t}$ where:

$$x_{1:t} = x_1, .., x_t, \; x \in [0, 1] \tag{3.5}$$
$$y_{r,1:t} = y_{r,1}, ..y_{r,t}, \; r \in R, y \in [0, 1]$$

The emission matrix $\mathbf{E}_r$ of each SU that maps Markov chain based PU channel activity to the observations $y_{r,1:t}$:

$$\mathbf{E}_r = \begin{bmatrix} 1 - P_f(r) & P_f(r) \\ 1 - P_d(r) & P_d(r) \end{bmatrix} \quad r \in \{1, 2, .., R\} \tag{3.6}$$

### 3.2.4 Local Spectrum Prediction Model

Consequently, define $\hat{x}_{r,1:t}$ as the *predicted* state sequence by $r$th SU based on $y_{r,1:t}$ observations, at time instant $t$, then HMM Bayesian model is a two step process *prediction* and *Update* [126]:

$$\hat{x}_{r,1:t} = \hat{x}_1..\hat{x}_t, \quad x \in [0, 1] \tag{3.7}$$

1. *Prediction*: Chapman-Kolomogrov equation

$$p(\hat{x}_{r,t}|y_{r,1:t-1}) = \int_{\mathcal{X}} p(x_t|x_{t-1})p(\hat{x}_{r,t-1}|y_{r,1:t-1})dx_{t-1} \tag{3.8}$$

2. *Update*: Bayes rule

$$p_r(\hat{x}_{r,t}|y_{r,1:t}) = \frac{p(y_{r,t}|x_t)p(x_t|y_{r,1:t-1})}{p(y_{r,t}|y_{r,1:t-1})} \tag{3.9}$$

To insure the probability is normalized to 1, the normalisation factor is:

$$p_r(y_{r,t}|y_{r,1:t-1}) = \int_{\mathcal{X}} p(y_{r,t}|x_t)p(x_t|y_{r,1:t-1})dx_t$$

The observation sequence by $r$th SU is a function of the PU activity, channel propagation and signal detection model i.e. $y_{r,1:t} \sim p(y_{r,1:t}) = f(t, \rho_r|P_d, P_f, \mu, \theta)$. For a finite state space HMM, local prediction is utilizes forward backward algorithm to estimate the joint posterior probability of primary user activity and secondary user's observation sequences$p(\hat{x}_{r,0:t}, y_{r,1:t-1})$ [2, 3, 52].

$$p(\hat{x}_{r,0:t}, y_{r,1:t-1}) = p(x_0)\left[\prod_{n=1}^{t} p(x_t|x_{t-1})\right]\prod_{n=1}^{t-1} p(y_{r,t}|x_t) \tag{3.10}$$

The prediction problem can be formulated using the Bayesian notion $p(\check{x}_t|y_{1:t-1})$ as the probability of the next state given a vector of past observations. Using the Bayesian definition of joint probability relation above:

$$p(\hat{x}_{r,t}|y_{r,1:t}) = \sum_{x_{r,t}=i} p(x_t|x_{t-1})p(\hat{x}_{r,t}|y_{r,1:t-1}) \quad : i \in \{1, 2\}. \tag{3.11}$$

The last term in the equation above, represents the predictive posterior probability at time instant $t$ given the observation sequence. This probability can be calculated using the Forward-Backward algorithm described in Section1.4 (*practically only the forward probability is required*). Finally, local one step-ahead prediction decision (Equation-3.7) is produced using binary hypothesis testing on test statistics based on threshold $\zeta$:

$$\hat{x}_{r,t} = \begin{cases} 0 & \text{Available} \quad p(\hat{x}_{r,t}|y_{r,1:t-1}) \geq \zeta \\ 1 & \text{Occupied} \quad p(\hat{x}_{r,t}|y_{r,1:t-1}) < \zeta \end{cases} \tag{3.12}$$

The test statistics $p_r(\hat{x}_{r,t}|y_{r,1:t-1})$ (i.e. predictive posterior probability) describes the *confidence* in next state prediction given the observation sequences up to current state. The probability is non-stationary random variable and a function of individual SU observation series. However, analytical expression of the stationary conditional predictive posteriori (and hence error probability) is counter-intuitive, and hard to obtain in a closed form for arbitrary FSM models [35, 53]. Alternatively, prediction error performance can be analysed using numerical approximation of local test statistics to calculate single user mean prediction error (Subsection-1.4.4) [2, 3].

## 3.3 Theoretical Analysis

In this section, we present mean prediction error approximation based on the state posterior probability for the system presented in Section-3.2. The work is motivated by the pioneer work in [35, 37, 145] on the recursive posterior estimation performance of finite state Markov machines. The theoretical analysis addresses the stationary predictive posterior distribution of **HMM** based SOP. A new generalized Beta-Bernoulli approximation of the predictive posterior probability for local HMM based spectrum predictors is proposed. Subsequently for cooperative prediction, a new generalized Beta-Binomial **GBB** approximation of HMM decision fusion is also presented in Chapter-4. The results show that the GBB approximation is no worse than a loose upper bound of HMM stationary predictive posterior probability. The approximation effectively captures with high accuracy the minimum and maximum values of HMM stationary distributions for constant false alarm based predictors.

Ultimately, HMM conditional predictive posterior (and mean perdition error) can be approximated based on the same distribution family of Markov chain without loss of optimality due to confining the distribution selection to the same predictor class **MC** (See Minimax Universality in [35]). The theoretical analysis results summary:

- Generalized Beta-Bernoulli approximation of the predictive posterior probability for HMM based spectrum predictors.
- Mean prediction error model, and numerical estimations for local spectrum prediction.

The definition and entropy of mean prediction error are presented in Subsection-1.4.4. Subsequently, ideal sensing scenario as a baseline performance is presented in Subsection-3.3.1. Subsection-3.3.2 traces HMM posterior probability in non-ideal system model described in Section-3.2. Finally, a generalized Beta-Bernoulli approximation is proposed for mean prediction error in Subsection-3.3.3 based on the recursive calculations presented in Subsection-3.21.

### 3.3.1 Ideal Sensing (Markov chain)

In this subsection we address mean prediction error for the special case of ideal sensing or Markov chain model. Under ideal sensing assumption i.e ($P_f = 0, P_d = 1$) **HMM** collapses to a Fully observable Markov chain. The marginal state probability in Chapman-Kolomogrov equation after dropping the $r$ subscript (All users receive the same sequence)*:

$$p_r(x_t) = \int_{\mathcal{X}} p_r(x_t|x_{t-1})p_r(x_{t-1})dx_{t-1} \tag{3.13}$$

Consequently, given the transition probability $p(x_t|x_{t-1})$ of a two state Markov chain, the stationary distribution of the latent state process $\hat{p}(x_t)$ is a Bernoulli random variable [38]. The probability mass function is given by:

$$\begin{aligned} v_1 = p(x_t = 0) &= \frac{(1-\mu)}{(1-\mu)+(1-\theta)} \\ v_2 = p(x_t = 1) &= \frac{(1-\theta)}{(1-\mu)+(1-\theta)} \end{aligned} \tag{3.14}$$

which satisfies the condition [38]:

$$\mathbf{p}(x_t) = \mathbf{p}(x_{t-1})\mathbf{P} \tag{3.15}$$

where the row vector $\mathbf{p}(x_t) = [v_1 \quad v_2]$. The mean prediction error can be calculated using the entropy rate of the Markov chain defined by the transition matrix, and it's stationary distribution.

$$\mathcal{H}(\pi_e) \geq \mathcal{H}(x_t|x_{t-1})$$
$$\mathcal{H}(x_t|x_{t-1}) = \mathbb{H}(Ber(1-\theta))v_1 + \mathbb{H}(Ber(1-\mu))v_2$$

Where

$$\mathbb{H}(Ber(x)) = -p(x)\log_2 p(x) - p(1-x)\log_2 p(1-x) \tag{3.16}$$

is the entropy of a Bernoulli distributed random variable, and $v_i$ is the stationary distribution of state $i \in [0,1]$. Fig.3.3 shows the prediction error entropy against the probability to remain in busy/idle states $\mu, \theta$, for all the values when ($\mu = \theta$) under perfect sensing. The result is trivial as prediction error depend on the primary user's uncertainty in remaining in busy/idle state manifested by the pair $((1-\mu), (1-\theta))$. Practically, the result presents two significant limitation for HMM based predictors, the first being the lower bound on the spectrum prediction error, as the system completely defined by $(\mu, \theta)$. Thus, the secondary user's best possible performance is completely dependent on the primary user's statical distribution, and transition matrix barring measurement errors. The second outcome, shows that under perfect sensing cooperative hard combining (Chapter-4) has no impact on improving the prediction error. This outcome is trivial, as all secondary users will have the exact prediction of the observable Markov chain.

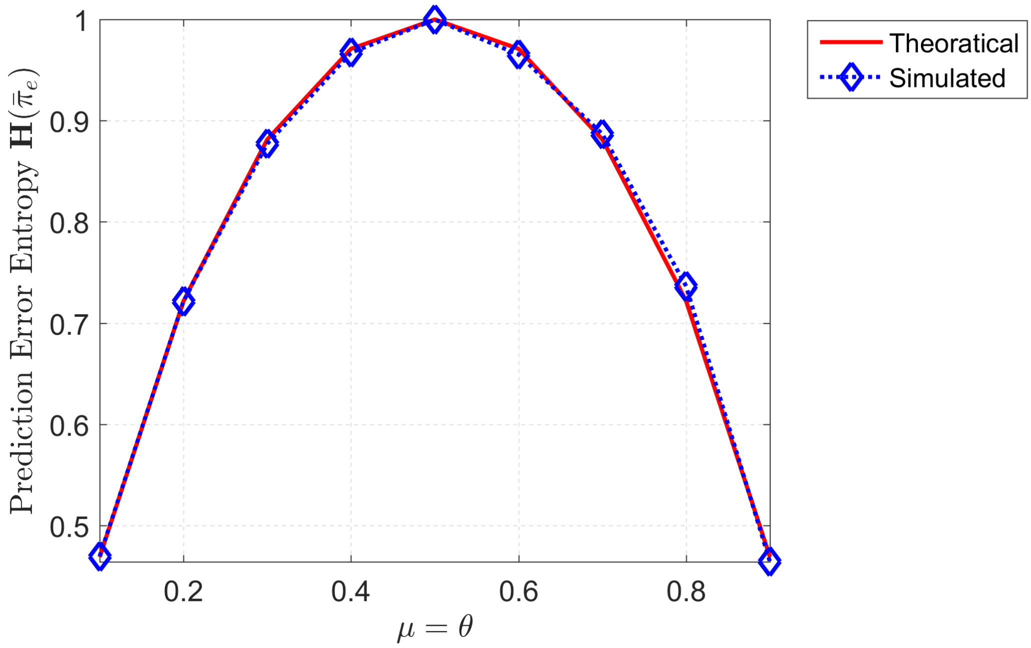


Fig. 3.3 Prediction Error Entropy

Finally, Bernoulli stationary distribution of Markov chain infers the validity of a similar approximation for non-ideal sensing case. The next subsection provides theoretical analysis of HMM posterior probability, and provide the theoretical background for Beta-Binomial approximation for mean prediction error in subsection-3.3.3.

### 3.3.2 Non-ideal Sensing (HMM)

In this subsection we address the general case of non-ideal sensing i.e. the system model described in Section-3.2. The bounds for mean prediction error presented in Subsection-1.4.4 concluded that mean prediction error required the complete characterisation of HMM predictive posterior probability. The posterior probability $p(x_{r,t}|y_{1:t-1})$ can be calculated recursively, and is defined as non-stationary Markov chain *state information variable* [37, 146]. Goldsmith *et al.*[37] and Kaijser [41] postulated that under limited assumptions the probability $p(x_t|y_{t-1})$ defined as *State Information Variable* is distributed as non-stationary Markov chain:

$$p(\psi_{r,t+1}|\psi_{r,t}) \triangleq p(\psi_{r,t+1} = \alpha|\psi_{r,t} = \beta) = \sum_{x_t \in \mathcal{X}} 1[y_{r,t} : \hat{f}(\beta, y_{r,t}) = \alpha].p(y_{r,t}|\psi_{r,t} = \beta). \tag{3.17}$$

where $\psi_{r,t}, \psi_{r,t+1} \in (0,1)$, and is given by

$$\psi_{r,t} = p(x_{r,t}|y_{r,1:t-1}) \tag{3.18}$$

$$\psi_{r,t+1} = \frac{\psi_{r,t}\mathbf{D}(y_{r,t})\mathbf{P}}{\psi_{r,t}\mathbf{D}(y_{r,t})\bar{\mathbf{1}}} \triangleq \hat{f}(\psi_{r,t}, y_{r,t})$$

$$p(\psi_{r,t+1} = \alpha\psi_{r,t} = \beta) = \begin{cases} 1-\beta & : \alpha = \hat{f}(y_{r,t} = 1, \beta) \\ \beta & : \alpha = \hat{f}(y_{r,t} = 2, \beta) \end{cases} \tag{3.19}$$

Where $\mathbf{P}$ is the transition matrix, $\mathbf{D}$ is a matrix where the $j$th diagonal term is $p_r(y_{r,t} = j|x_t)$, and $\bar{1}$ is a unit column vector. The posterior probability is a function of the signal to noise ratio $\rho_r$, given **HMM** transition and emission probabilities i.e. $y_{r,t}|\psi_{t-1} = f(\rho_r|P_d, P_f, \mu, \theta)|\psi_{t-1})$.

For the system model in Subsection-3.2, the number of states is limited to *2* which to be substituted in Equation-3.18. Prediction error can be calculated instantaneously, and the mean limiting prediction error can be computed using Equation-3.21 derived from Equation-3.18. Equation-3.21 generates posterior probability (state information variable) samples based on a non-stationary Markov chain. However, the stationary distribution of non-stationary Markov chain can be approximated with a binomial process with non-constant success probability.

The next subsection addresses the numerical approximation of local prediction error based on Beta-Binomial distribution.

$$\psi_{r,t}(i) = \left[1 - \breve{\psi}, \breve{\psi}\right] \quad : i \in \{0, 1\}. \tag{3.20}$$

$$f(y_{r,t}, \psi_{r,t}) \triangleq \begin{cases} P_f + (1-\mu)(P_m - P_f) + (\mu + \theta - 1)(\breve{\psi} - P_f)(1 - P_m)/(1 - \breve{\psi}) & : y_{r,t} = 0 \\ P_f + (1-\mu)(P_m - P_f) + (\mu + \theta - 1)(\breve{\psi} - P_f)P_m/\breve{\psi} & : y_{r,t} = 1 \end{cases}$$

### 3.3.3 Generalized Beta-Bernoulli approximation of local prediction error

Define $[a, b]$ the minimum and maximum values for $\psi_{r,t}$, respectively. The bounds represent two extreme cases for SU predictive posterior value based on it's respective detection performance $P_m, P_f$, as well as limit the probability space of $\psi_{r,t}$:

$$\begin{aligned} a \leq \psi_{r,t} \leq b \\ a = \psi_{r,t}\,, P_d = 1 \\ b = \psi_{r,t}\,, P_d = 0 \end{aligned} \tag{3.21}$$

Based on the inference in Subsection-3.3.1, HMM stationary distribution can be approximated by Bernoulli process parametrized by the instantaneous *non-stationary* predictive posterior probability:

$$\hat{p}_r(\hat{x}_{r,t}|\psi_{r,t}) \sim Bernoulli(\psi_{r,t}) \tag{3.22}$$

Where each secondary user's prediction $\hat{x}_{r,t}$ in Eq-3.12, and $\zeta$ is the detection threshold. Then, the posterior probability $\psi_{r,t}$ is generalized Beta distribution with hyper-parameters $\alpha\,, \beta$ in a probability space defined by $[a, b]$ (setting $a = 0\,, b = 1$ bears the standard Beta distribution [50]):

$$\begin{aligned} \hat{p}(\psi_{r,t}|\alpha, \beta, a, b) &\sim Beta(\alpha, \beta, a, b) \quad \alpha, \beta > 0 \\ \hat{p}(\psi_{r,t}|\alpha, \beta, a, b) &= \frac{(\psi_{r,t} - a)^{\alpha-1}(b - \psi_{r,t})^{\beta-1}}{\mathbf{B}(\alpha, \beta)(b - a)^{\alpha+\beta-1}} \\ \mathbf{B}(\alpha, \beta) &= \int_0^1 t^{\alpha-1}(1 - t)^{\beta-1} dt \end{aligned} \tag{3.23}$$

Where $\mathbf{B}(\alpha, \beta)$ is the standard Beta function. The proposed generalised Beta-Bernoulli approximation selects the stationary distribution from the same family as

Markov [35], while Beta distribution is the conjugate prior of Bernoulli Distribution [48]. The proposed prior distribution models the uncertainty in the Bernoulli parameter $\psi_{r,t}$ for each user $r$ and time instant $t$. Subsequently, The marginal predictive posterior distribution of **HMM** can be created by integrating the posterior probability $\psi_{r,t}$ over the parameter likelihood $L(\psi_{r,t}|\rho_r, x_t)$. The compound predictive posterior probability is then given by generalized Beta-Bernoulli distribution:

$$\hat{p}(x_{r,t+1}|\alpha,\beta) = \int_a^b \hat{p}_r(x_{r,t+1}|\psi_{r,t})\hat{p}(\psi_{r,t}|\alpha,\beta,a,b)d\psi_{r,t} \tag{3.24}$$

$$\hat{p}(x_{r,t+1}|\alpha,\beta) \sim Beta - Bernoulli(\alpha,\beta)$$

$$f(x_{r,t+1}|\alpha,\beta) = \frac{\mathbf{B}((\hat{x}_{r,t+1}+\alpha),(1-\hat{x}_{r,t+1}+\beta))(b-a)}{\mathbf{B}(\alpha,\beta)} \tag{3.25}$$

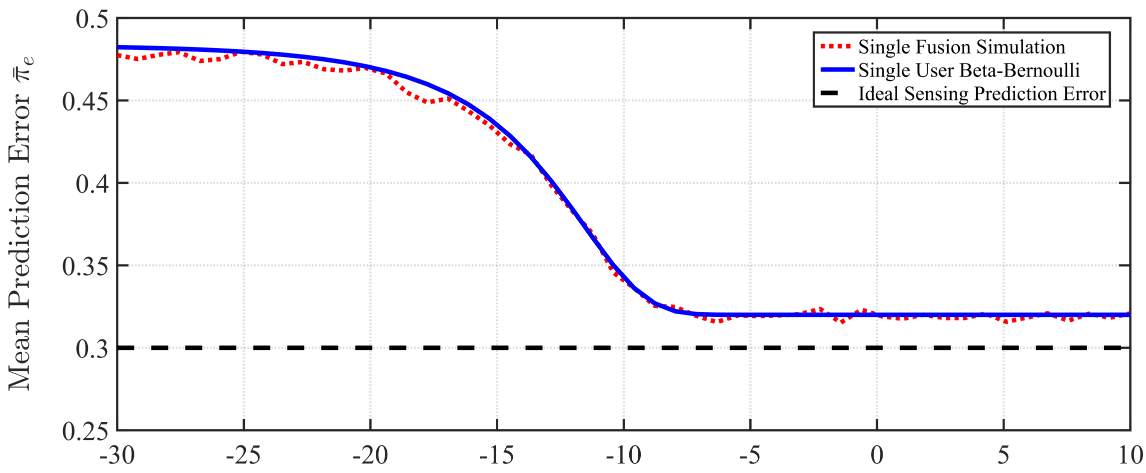


Fig. 3.4 Beta Bernoulli Approximation of HMM prediction error

Figure-3.4 displays the Beta-Binomial approximation in this subsection, along with simulated values for mean prediction error in local spectrum prediction. Ideal sensing error values present the lower bound for performance under all channel conditions e.g. signal to noise ratio values or $P_m, P_f$ values. The case study analysis in Section-3.4 provides an in-depth look at mean prediction error with respect to the model, and scenario parameters.

## 3.4 Case Study Analysis

The contribution of this section is the simulation and local prediction error performance analysis. Mean prediction error is presented as a function of state transition, and channel conditions. Specifically, we model the prediction error of one step-ahead (single time slot) prediction against the channel detection errors, as well as primary user's state transition probability. Prediction error is also investigated against the observation sequence length, to examine the correlation between prediction accuracy, and the number of samples required to calculate the next state probability. We further examine the prediction error of two step-ahead, based on incorrect one step-ahead prediction.

This case study addresses the two scenarios of *ideal sensing*, and *non-ideal sensing*. The first case is considered when $P_m = P_f = 0$, while any arbitrary values of $P_m, P_f$ represents the error introduced by spectrum sensing/detection functions. The accuracy of spectrum prediction is measured by $\bar{\pi}_e$ the prediction error. In general, asymptotic calculation of the stationary distribution is computationally expensive, and requires optimized methods such as particle filter state estimation [126]. However, the ideal sensing prediction error $\bar{\pi}_e$ when $P_m = P_f = 0$ is calculated based on $\mu, \theta$ only. The results for non-ideal sensing are calculated by recursively running Monte-Carlo simulation, and calculating the arithmetic mean $\bar{\pi}_e = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n}(\pi_i)$ (Mean Prediction Error).

Table 3.1 Local Prediction simulation parameters

| Variable | Definition | Value |
|---|---|---|
| **K** | Number of Discrete HMM Hidden States | 2 |
| **L** | Number of Discrete HMM Observed States | 2 |
| $P_m$ | Probability of Miss-Detection | [0,1] |
| $P_f$ | Probability of false alarm | [0,1] |
| $\mu$ | Probability to remain in busy State | (0,1) |
| $\theta$ | Probability to remain in idle State | (0,1) |

This case study investigates mean prediction error dependency on:

1. **Prediction error $\bar{\pi}_e$ as a function of $N$**
   To measure the correlation between the prediction accuracy on the number of observation events $N$, i.e. to measure to convergence of mean prediction error.

2. **Prediction error $\bar{\pi}_e$ as a function of $P_m$ and $P_f$**
   To measure the correlation between the prediction accuracy, and the detection accuracy represented by .

3. **Prediction error $\bar{\pi}_e$ as a function of $\mu, \theta$**
   To measure the correlation between the prediction accuracy and state dwelling probability represented by $\lambda$ and $\theta$.

4. **Prediction error $\bar{\pi}_e$ of successive slots**.
   To measure the prediction error, given an error occurred in predicting the previous slot. The goal is to identify the accumulated error of successive slots prediction.

The result shown in Fig.3.5 displays the mean prediction error of a single slot $\bar{\pi}_e$ as a function of the length of observation sequence. The probability to remain in a busy or an idle state is chosen as $\mu = \theta = 0.7$, and detection Error $P_m = P_f = 0.1$. The simulation results are averages of different observation sequence lengths. The non-stationary nature is reflected on oscillating average mean prediction error. However, as the length $N \to \infty$ the average prediction error converges to a constant value. As $n \to \infty, \bar{\pi}_e = Const$, the mean prediction error value is completely defined by the values of $\mu, \theta, P_m, P_f$. Thus, the Non-stationary Markov chain of the probability $\mathbf{P}(x_t|y_{1:t-1})$ converges to a stationary distribution.

In Fig. 3.6, the prediction error $\bar{\pi}_e$ in single slot is presented as a function of the probability to remain in a busy state $\mu$ when $P_f = P_m = 0.1$. The results intuitively showed that prediction error increases as the transitional probabilities approaches the equally likely for each state (uniform distribution). Thus, as values of $\mu, \theta$ are further away from *0.5*, the channel is more predictable and mean prediction error depends on $\mu, \theta$ for any fixed pair of $P_f, P_m$. Mean prediction error approaches it's maximum value when both or either state transition approaches *0.5*. In Fig.3.7 the case where $\mu = \theta$ represents uniform stationary distribution case of Markov transition probabilities. The mean prediction error is following a linear function of $\mu, \theta$ with maximum error at $\mu = \theta = 0.5$. This also infers that limiting prediction error distribution is completely defined by the set $\mu, \theta, P_m, P_f$.

Prediction detection Fig-3.9, and prediction false alarm Fig-3.10 display mean prediction error of a single slot as a function of the detection error probability $P_m, P_f$

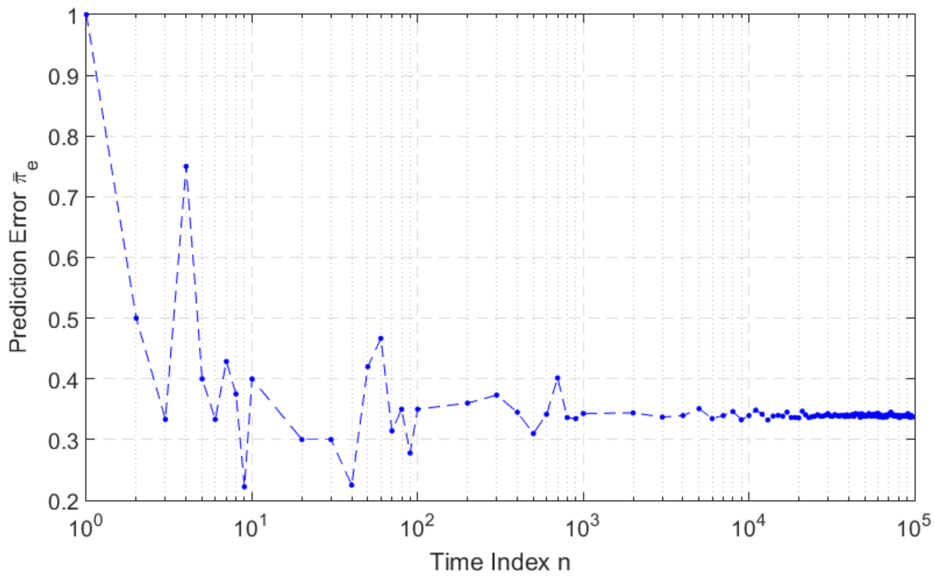


Fig. 3.5 Prediction Error as a function of observation sequence length

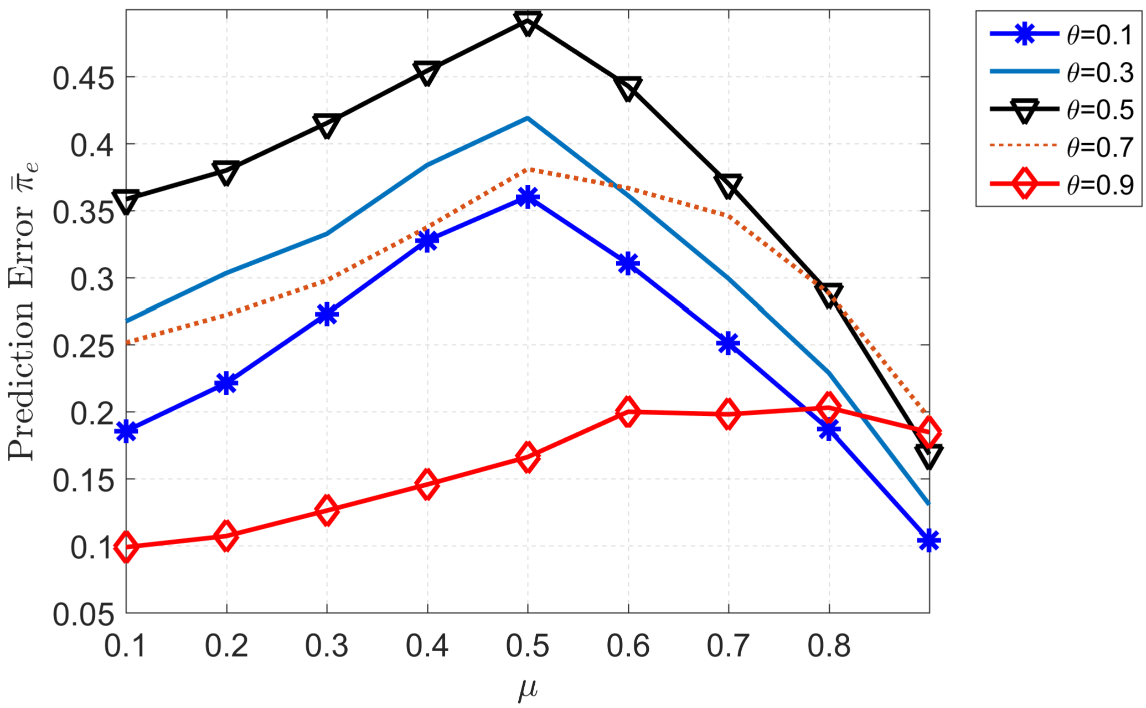


Fig. 3.6 Single User prediction Error Vs. Channel occupancy probability.

where $\mu = \theta = 0.7$. It infers that the mean prediction error is an increasing linear function of $P_m$ when the difference $P_m - P_f \geq 0$, and decreasing linear function when $P_m - Pf \leq 0$. When $P_m, P_f$ are grater than or less than *0.5*, HMM predictor per-

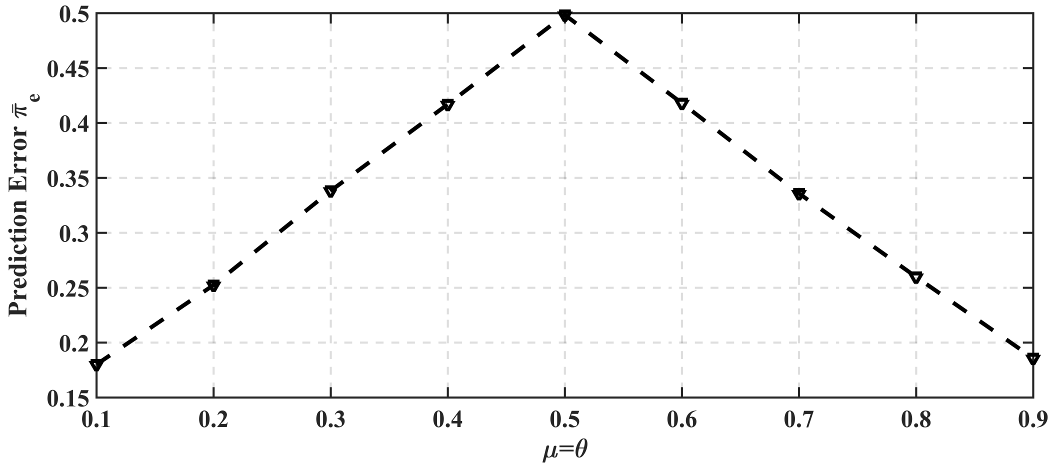


Fig. 3.7 Prediction Error Vs probability to remain in specific state, Equal Stationary Distribution case

formance is symmetrical. Fig.3.8 displays mean prediction error for selected samples of $(P_m, P_f)$ when $(\mu = \theta = 0.7)$ i.e. $v_1 = v_2 = 0.5$. The mean prediction error$(\bar{\pi}_e)$ approaches 0.5 i.e. no information can be deducted from the stochastic source when the values $(P_m + P_f = 1)$. In practice, secondary users have no control over the values of $(\mu, \theta)$ i.e. the primary user's statistics. Thus, only Fig.3.8 is available to secondary users, and so the secondary user's spectrum prediction accuracy can be evaluated relative to spectrum sensing errors $P_m, P_f$. Finally, Fig.3.9 and Fig.3.10 present the partial prediction detection and false alarm errors described in Subsection-1.4.4. The results shows how prediction detection against sensing miss-detection probability are weighted components of mean prediction error.

The cost of making a incorrect prediction in the first slot prediction on the accuracy of subsequent slot prediction is shown to be excessively high in both Fig.3.12, and Fig.3.11. The mean prediction error is plotted against the transition probabilities in Fig.3.12, and in Fig.3.11 against detection probabilities. The prediction error remains higher than $0.5$, and increase/decrease in proportion to the difference $(P_m - P_f)$ in contrast to single slot prediction. If both $\mu, \theta$ are either below or above $0.5$, the mean prediction error remain excessively high. The results infers that first order HMM mean prediction error for successive slots is excessively high given the sequential prediction is based on erroneous value. The inference in these results is that first order HMM might no be suitable for state duration prediction to predict that state of $\tau$ slots.

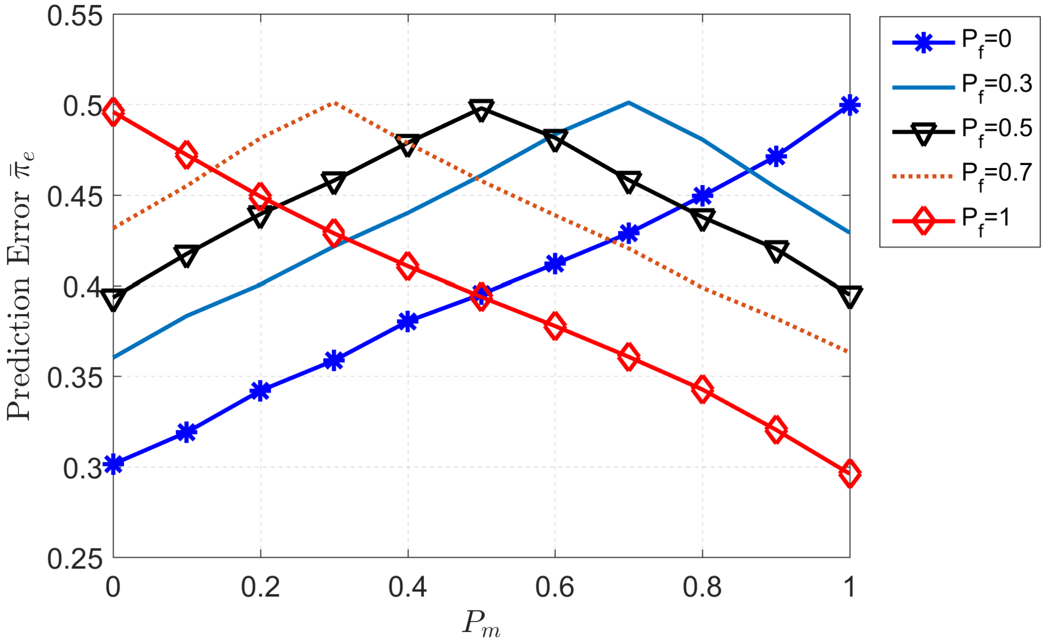


Fig. 3.8 Single User prediction Error Vs. Channel Errors

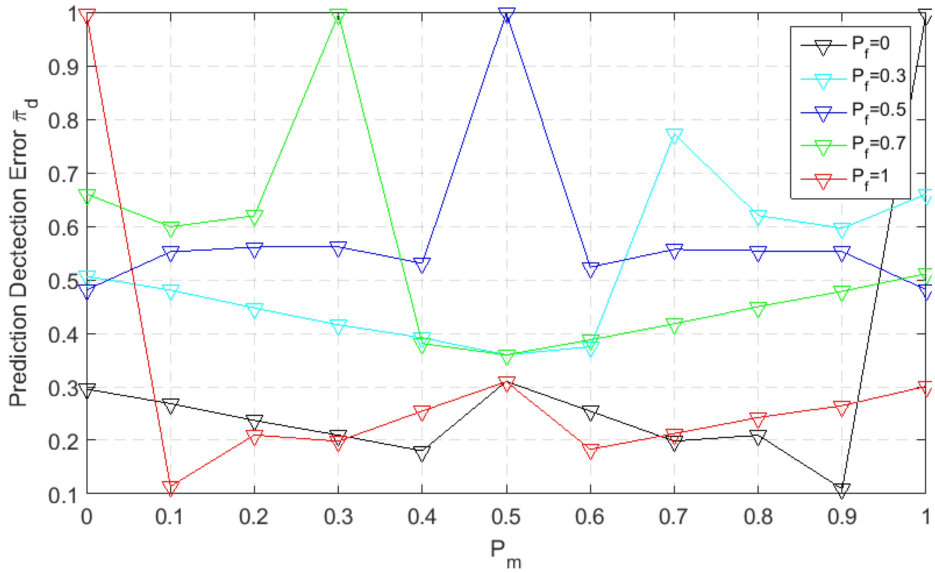


Fig. 3.9 Prediction Detection $\bar{\pi}_d$ vs Channel Errors ($P_m$ for different $P_f$ values)

**Results Summary** The mean prediction error is shown -through simulation- to converge as $N \leftarrow \infty$ to a distribution defined completely by the set $\mu, \theta, P_m, P_f$. However, mean prediction error remains dependent on $\mu, \theta$ even in $P_m = P_f = 0$ i.e. *perfect sensing* which can be relatively unacceptable for many prediction applications. In other words, even if the Markov chain is ideally observable, the mean prediction

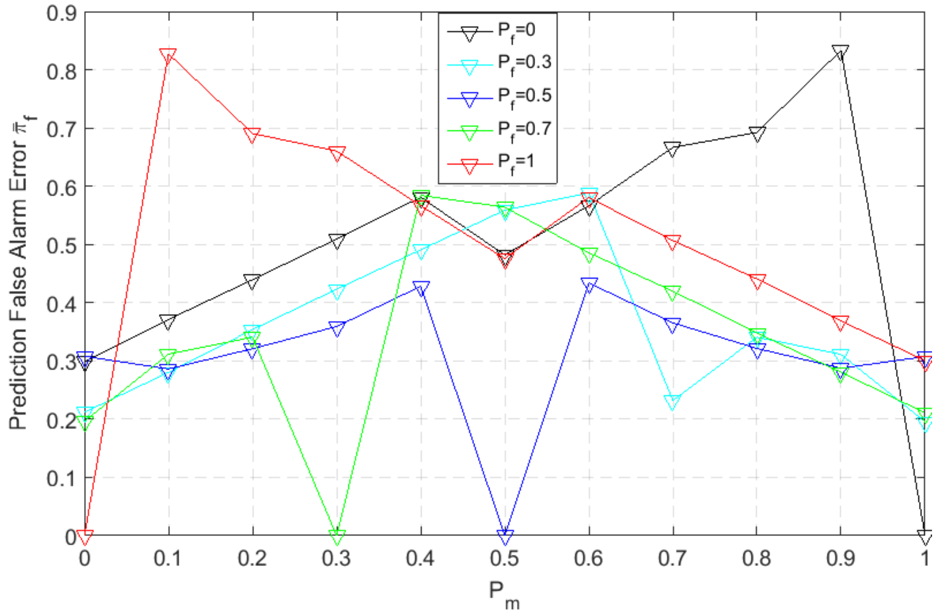


Fig. 3.10 Prediction False alarm $\bar{\pi}_f$ vs Channel Errors ($P_m$ for different $P_f$ values)

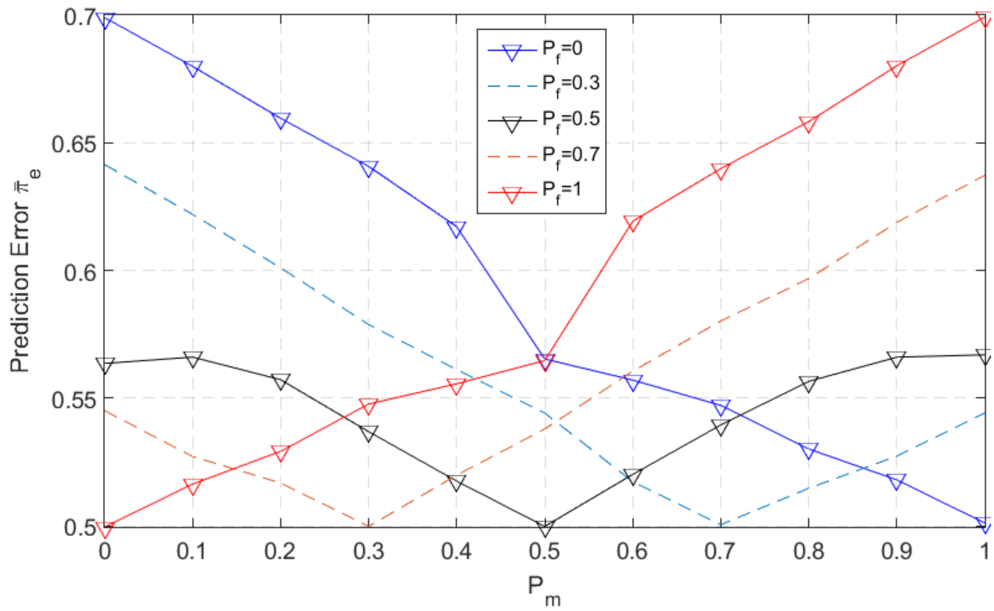


Fig. 3.11 Prediction Error for second Slot (Given error in the first prediction) VS miss-detection probability

error remains high as a function of $\mu, \theta$. Additionally, two step-ahead prediction has higher error probability in when based on wrong one step-ahead decision of the previous prediction. Thus, an explicit duration HMM is a more suitable candidate for such prediction scenarios with a large prediction span.

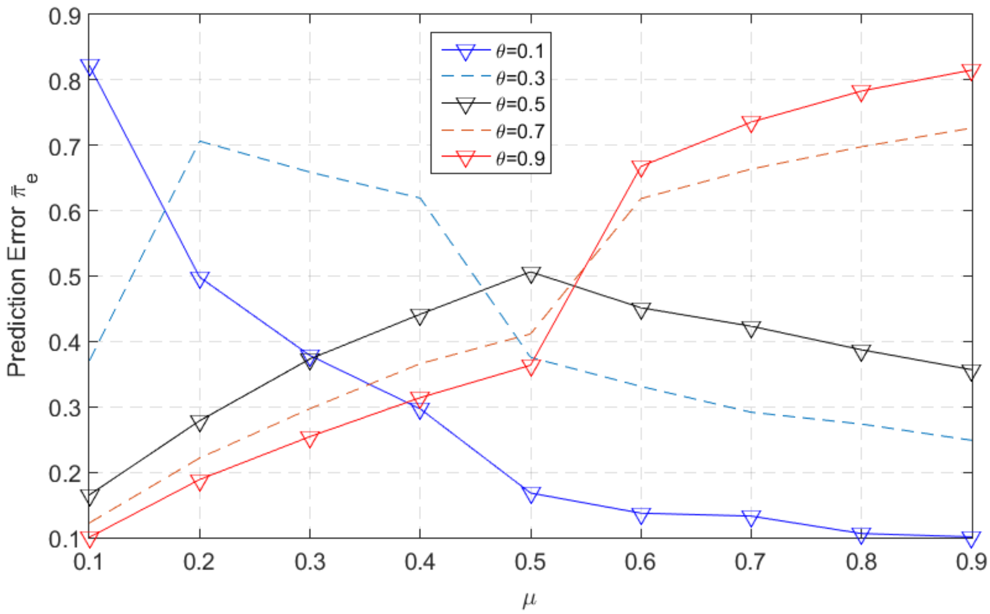


Fig. 3.12 Prediction Error for second Slot (Given error in the first prediction) VS probability to remain in busy state

## 3.5 Summary

In this chapter, we proposed a local prediction algorithm based on hidden Markov model for spectrum occupancy. Then, we presented an analytical approximation of HMM prediction error performance. The results of analytical approximation show that the GBB approximation is no worse than a loose upper bound of HMM stationary predictive posterior probability. The approximation effectively captures with high accuracy the minimum and maximum values of HMM stationary distributions for constant false alarm based predictors. Additionally, we presented a case study of the hidden Markov model based prediction, and simulated the performance bounds of mean prediction error against the model parameters in terms of channel sensing errors, and channel occupancy transitions.

# Chapter 4

# Cooperative Spectrum Prediction based on Hard Fusion

In this chapter, we address hard fusion based cooperative prediction and the expected prediction accuracy gain based on the local prediction model presented in Chapter-3. The chapter provides the analytical approximations of the hard fusion based SOP as consequence of the approximation presented in Subsection-3.3.3 for local spectrum prediction. For cooperative prediction, a new generalized Beta-Binomial **GBB** approximation of HMM decision fusion is proposed. Performance comparison to assess the effectiveness of the GBB approximation are presented in terms of mean prediction error. The results show that the GBB approximation is no worse than a loose upper bound of HMM stationary predictive posterior probability. The approximation effectively captures with high accuracy the minimum and maximum values of HMM stationary distributions for constant false alarm based predictors. This contribution of this chapter is submitted to IEEE communication letters Subsection-1.5.2.

Moreover, the case study analysis contribution of this chapter is published in the conference proceedings of the IEEE International Conference on Communications (**ICC'2016**) workshops [3]. The work in [3] extended the prediction accuracy performance analysis of local HMM SOP in [2] to homogeneous hard fusion based cooperative SOP. The results have shown that hard decision fusion considerably improved the prediction performance compared to local user SOP. Section-4.1 provides background and related work in cooperative work in cooperative predictio, Section-3.2 describes the system model for HMM based prediction, while Section-4.3 elaborates on the generalized Beta-Binomial **GBB** approximation of *majority* based cooperative HMM predictors. Finally, Section-4.4 presents case study analysis of cooperative prediction based on hard fusion.

## 4.1 Background and Related Work

Finite State Markov machines (**FSM**) constitute a major portion of *local* spectrum modelling literature. Spectrum prediction in single secondary user environment has been dubbed *local spectrum prediction* [43]. Consequently, *cooperative spectrum prediction* was proposed to improve the collective accuracy of spectrum occupancy prediction. Cooperative fusion of secondary user's decisions has been studied extensively to address diverse optimization problems, particularly in DSA spectrum sensing. A multitude of decision fusion techniques such as hard fusion (**HF**) and soft combining were used for temporal and spatial fusion of local node decisions [45]. However, literature on cooperative spectrum prediction is limited. The proposals presented an incomplete analysis of cooperative prediction accuracy improvement. In [43] cluster formation, and coalition based game theory were implemented for multi-primary, multi-secondary user environment. The study presented the total HF accuracy improvement, but the results lacked fine details of causes of improvement in terms of primary user activity patterns and secondary user's sensing accuracy.

The work in [115] proposed a cooperative Bayesian non-parametric framework for primary user (**PU**) transmission monitoring. The model tracks PU signal amplitude at each secondary user (**SU**) node based on non-linear particle filter. The non-parametric prior of the cooperative PF mixture model is a Beta distribution. The study presented a quantitative comparison between communication cost and prediction accuracy for single and cooperative. In SOP literature, the work in [43, 152] proposed a coalition game for cooperative multi-primary user activity monitoring based on HMM predictors. However, prediction accuracy results were limited to prediction performance under multiple PU users scenario. Prediction accuracy gains from the fusion technique, or robustness for wireless channel shadowing were not explicitly identifiable. In our previous work presented in chapter-3 [2], we addressed the issue of prediction error performance of HMM, relative to spectrum sensing errors and primary user activity patterns, as well as a generalized Beta-Bernoulli approximation for local prediction error.

In Section-4.3, we present a generalized Beta-Binomial approximation of cooperative hard fusion error performance. In Section-4.4, we further extend the performance analysis of local spectrum prediction, and address possible gains from cooperative spectrum prediction. In particular, we present an analysis of secondary user's mean prediction error in terms of primary user's activity pattern, and spectrum sensing errors. We utilize Bayesian filtering, and known information theory inequalities, to express mean prediction error for single secondary user. Then, we extend the analysis to *Hard Fusion* based cooperative spectrum prediction to present possible improvements of cooperative spectrum prediction.

## 4.2 Cooperative Prediction Model Formulation

In this section, we present a hard fusion based cooperative spectrum prediction. Similar to local prediction in Chapter-3, a DSA time slotted system with one primary user and $R$ secondary users is assumed. Each of the $R$ secondary users opportunistically attempts to utilize spectrum holes on the primary user's channel. The secondary users are assumed to be uniformly distributed around the primary user. The SU's employ *local spectrum sensing* on each time slot, for a duration $T$. The primary user occupancy detection is generated as an (**ON/OFF**) model. The SU ($j \in R$) has it's own pair of ($P_m^j, P_f^j$) values, that represents it's channel sensing error performance for the secondary user. The probabilities ($P_m, P_f$) are defined as the probability of miss-detection, and probability of false alarm, respectively. The values depend on the level of signal to noise ratio **SNR** ($\rho_r$) at each secondary user which is function of distance to the primary user. Following the spectrum sensing, the primary user ON/OFF pattern is passed to a local HMM predictor. The predictor estimates the probability of presence/absence of the PU in the one-step ahead time slot. The predictor models each spectrum sensing outcome as an observation vector of a partially observable Markov chain. HMM model training/parameter estimation is performed on-line using Baum-Welch algorithm. Cooperative spectrum prediction uses hard fusion rule of local spectrum prediction decisions to generate the cooperative decision. HF strategy employs ($m$ out of $R$) rule with state posterior probability based threshold to generate a cooperative spectrum prediction decision (Fig-4.1).

*Homogeneous* cooperative prediction refers to the case when secondary users have identical spectrum sensing error performance in terms of ($P_m, P_f$). While *Heteroge-*

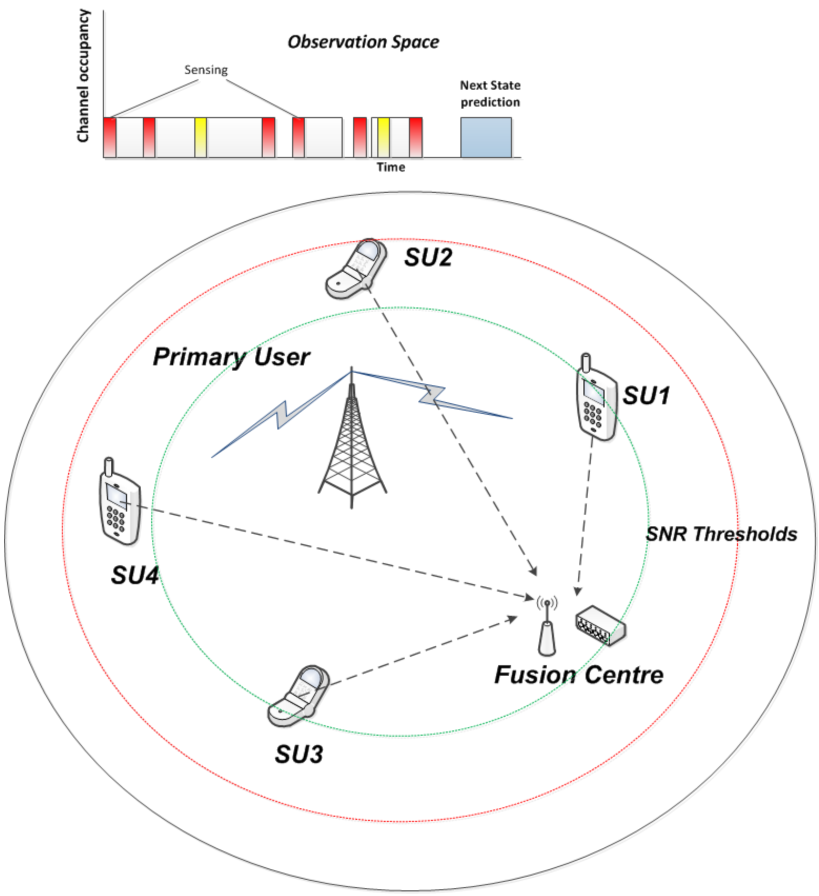


Fig. 4.1 Cooperative Hard Fusion Spectrum Occupancy Prediction

*neous* cooperative prediction refers to the case where secondary users have different sensing error performance. The latter incorporates additional dependency on clustering algorithm and functional dependence on the relative accuracy, while the former depends only on the local spectrum prediction performance. Cooperative spectrum occupancy prediction adopts two stages of decision on the presence or absence of the primary user. The first stage is local spectrum prediction by each secondary user, while the second stage local predictions are fused to generate the cooperative decision to be adopted by all nodes. Each user's local prediction decision is labelled $x_{r,t} \in \{0, 1\}$. The hard fusion strategy based on $m$ out of $R$ the cooperative decision $X_{1:R,t}$ at time instant $t$ can be defined as in (Equation-4.1) [3]:

$$\hat{x}_{1:R,t} = \begin{cases} 1 \text{ (Busy)} & \epsilon_{m,t} \geq m \\ 0 \text{ (Idle)} & \epsilon_{m,t} < m \end{cases} \tag{4.1}$$

$$\epsilon_{m,t} = \sum_{r=1}^{m} 1[x_{r,t}]$$

The three notable rules for $m$ are: $m = 1$ that represents the logical (**OR**), while setting $m = R$ implements the logical (**AND**) rule. Finally setting $m = \frac{R}{2}$ implements (**Majority**) decision rule. The three rules are compared in case study analysis(Section-4.4)in terms of mean prediction error. However, analytical expressions of the Binmoial approximation for homogeneous cooperative prediction are presented for the majority rule in Section-4.3. Consequently, the generalised Beta-Binomial approximation in subsection-4.3.3 also considers majority hard fusion rule.

## 4.3 Cooperative HF Theoretical Analysis

In this section we examine analytical approximations for hard fusion based spectrum prediction. The binomial approximation in Subsection-4.3.2 addresses the majority rule case presented in Section-4.2. The approximation severs as upper bound for hard fusion prediction error performance. On the other hand, generalized Beta-Binomial approximation for HF cooperative prediction is presented in Subsection-4.3.3. The approximation is an extension to local spectrum prediction approximation presented in Subsection-3.3.3. Local prediction, ideal sensing, and binomial approximation provide comparison cases to assess the performance of hard fusion GBB approximation [3]. The scenario models PU spectrum occupancy activity and SU energy detection/sensing/prediction at an equal distance from the PU transmitter [3].

### 4.3.1 Cooperative Prediction Error

The mean prediction error in the perfect sensing represents the lower bound given a specific pair $(\mu, \theta)$ [2, 3]. It is trivial to stipulate that under perfect sensing, all users with identical channel error probabilities will have the same predicted state. Thus, cooperative prediction under perfect sensing will not improve the prediction error below it's lower bound. The most important aspect is that due to the lower bound set by perfect sensing, cooperative prediction error $\bar{\pi}_e$. Thus, all values of $\bar{\pi}_f, \bar{\pi}_m$ that will result in $\bar{\pi}_e$ lower than the bound will be quantized at the bound value in the equations (4.3), (4.4) ,and (4.5), i.e

$$\bar{\pi}_e = \mathcal{H}(x_t|x_{t-1}), \quad : \bar{\pi}_e < \mathcal{H}(x_t|x_{t-1}) \tag{4.2}$$

Prediction accuracy of local as well as cooperative prediction can be measured by the *mean prediction error* ($\bar{\pi}_e$) described in Chapter-3 [2, 3]. The cooperative prediction error based on linear approximations of homogeneous cooperative users (Binomial approximation) is presented in the next subsection. Finally, the miss-detection probability and false alarm probabilities can be calculated as a function of the local prediction counterparts.

### 4.3.2 Binomial approximation for cooperative SOP Error

Linear error approximation assumes a Binomial predictive posterior $\psi_{r,t}$ of local prediction. Numerical averages of the local prediction error are utilized to approximate cooperative prediction error for $m$ out of $R$ secondary users. Using the notion $\bar{\Pi}_e, \bar{\Pi}_m, \bar{\Pi}_f$ to indicate cooperative prediction probability error set, the cooperative error equations can be calculated in (4.3) for homogeneous cooperative case:

$$\begin{aligned}
\bar{\Pi}_m &= \sum_{j=0}^{M-1} \mathbf{C}(j,R)(1-\bar{\pi}_m)^j \bar{\pi}_m^{R-j} \qquad (4.3)\\
\bar{\Pi}_f &= 1 - \sum_{j=0}^{M-1} C(j,R)\bar{\pi}_f^j (1-\bar{\pi}_f)^{R-j}\\
\bar{\Pi}_e &= \bar{\pi}_m.p(x_t=1) + \bar{\pi}_f.p(x_t=0)
\end{aligned}$$

where $C(j,R)$ is a binomial coefficient. For $AND$ and $OR$ fusion rules, the probabilities reduce to the equations (4.4), and (4.5).

$$\begin{aligned}
\bar{\Pi}_m(OR) &= (\bar{\pi}_m)^R \qquad (4.4)\\
\bar{\Pi}_f(OR) &= 1-(1-\bar{\pi}_f)^R\\
\bar{\Pi}_m(AND) &= 1-(1-\bar{\pi}_m)^R \qquad (4.5)\\
\bar{\Pi}_f(AND) &= (\bar{\pi}_f)^R.
\end{aligned}$$

Under **OR** rule, cooperative prediction miss-detection error ($\bar{\Pi}_m$) improves, while cooperative prediction false alarm ($\bar{\Pi}_f$) degrades. On the other hand, **AND** fusion rule improves cooperative prediction false alarm ($\bar{\Pi}_f$), while degrading cooperative prediction miss-detection error($\bar{\Pi}_m$). The total ($\bar{\Pi}_e$) is identical for both *AND/OR* fusion rules. The approximation (quantized by ideal sensing minimum error) serves as a lower bound for cooperative prediction error, and a benchmark for the generalized Beta-Binomial approximation in subsection-3.3.3. The approximation assumes the binomial assumptions hold when HMM predictive posterior probability values are stationary $\psi_{r,t} = \bar{\psi}_r$. However, as previously presented on Beta-Bernoulli approximation in Chapter-3, HMM predictive posterior probability values are non-stationary as presented in Chapter-3. Thus, the generalized Beta-binomial approximation captures time-varying nature of HMM predictive posterior probability values (Subsection-4.3.3).

### 4.3.3 Generalized Beta-Binomial approximation of cooperative Prediction Error

Given the local spectrum predictive posterior probability $\psi_{r,t}$ is approximated by generalized Beta-Bernoulli distribution in Subsection-3.3.3, the approximation for $x_{1:R,t}$ , $\rho_r = \rho$ for $m$ out of $R$ homogeneous users is a Beta-Binomial distributed:

$$\begin{aligned}\hat{p}(x_{1:R,t+1}) &= f(k|R,\alpha,\beta) \sim Binomial(R,\psi_{r,t}) \\ &= \frac{(b-a)\Gamma(N+1)}{\Gamma(k+1)\Gamma(N-k-1)} \frac{\mathbf{B}(k+\alpha, N-k+\beta)}{\mathbf{B}(\alpha,\beta)}\end{aligned}$$

Where $\Gamma(t)$ is the standard Gamma function. The Beta binomial distribution describes the probability of a majority secondary assess the channel busy (or idle) [49, 50]. Hyper-parameters of the Beta-Bernoulli process $\alpha, \beta$ are numerically estimated using maximum likelihood estimation, or the method of moments for Beta distribution based on $\psi_{r,t}$. The samples of $\psi_{r,t}$ can be generated using the recursive equation-3.21 for $R$ homogeneous SU's. Using the method of moments, The values of $\alpha$ , $\beta$ can be calculated as (Fig-4.2):

$$\alpha = (\frac{\bar{\psi}_{r,t} - a}{b-a})(\frac{(\frac{\bar{\psi}_{r,t} - a}{b-a})(1-(\frac{\bar{\psi}_{r,t} - a}{b-a}))}{\sigma^2_{r,t,\psi}} - 1)$$

$$\beta = (1-(\frac{\bar{\psi}_{r,t} - a}{b-a}))(\frac{(\frac{\bar{\psi}_{r,t} - a}{b-a})(1-(\frac{\bar{\psi}_{r,t} - a}{b-a})}{\sigma^2_{r,t,\psi}} - 1)$$

Where $\bar{\psi}_{r,t} = \mathbb{E}(\psi_{r,t})$ is mean value of the posterior probability random variable, and $\sigma^2_{r,t,\psi} = \mathbb{E}((\psi_{r,t} - \bar{\psi}_{r,t})^2)$ is the variance. Finally, if the limiting distribution exists $\lim_{t\to\infty} \hat{p}(x_{r,t+1})$ then the subscript $t$ is dropped e.g. $\psi_r \triangleq \psi_{r,t}, t \to \infty$. Fig-4.2 presents GBB approximation hyper-parameters $\alpha, \beta$ as a function of signal to noise ratio $\rho_r$. Detection accuracy probabilities $(P_f(r), P_d(r))$ equivalent of signal noise ratio for CFAR SU detection are also presented in Fig-4.2.

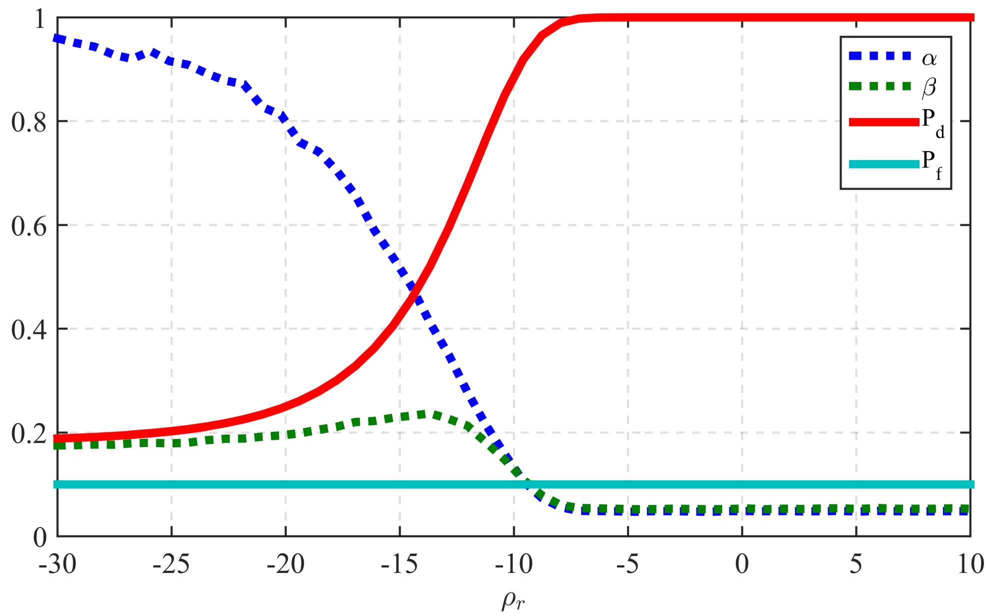


Fig. 4.2 GBB Hyper-parameters

Moreover, GBB Beta-prior is a loose upper bound approximation for the posterior probability $\psi_{r,t}$, while binomial approximation is a lower limit for $-10 > \rho_r > -20$ (Fig-4.3). Beta prior hyper-parameters in GBB approximation captures test statistics

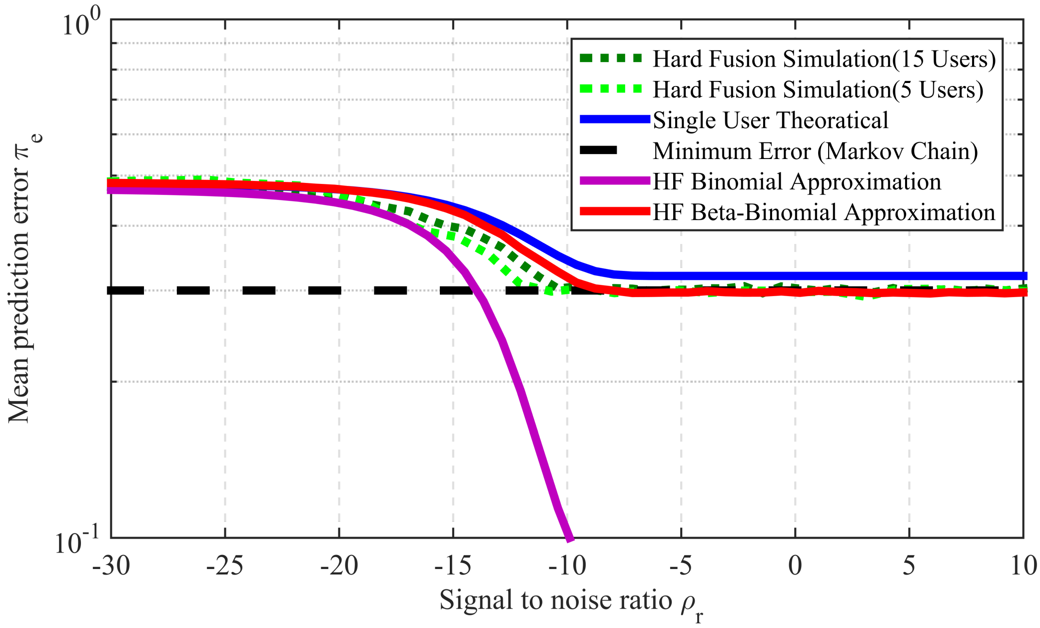


Fig. 4.3 GBB approximation of hard fusion mean prediction error

non-stationary state information variable underestimated by the Binomial approximation. The proposed GBB theoretical approximation matches the simulated example with high accuracy for $\rho_r > -10$ dB or $P_d > 0.9, P_f = 0.1$. The approximation holds for arbitrary number of users with majority hard fusion rule (Fig-4.4).

Finally, GBB approximation is effective for all values of transition probabilities when $\mu = \theta$ for any number of users (Fig-4.5). Both Binomial and GBB approximations match maximum prediction error when $\rho_r < -20$ dB or $P_d = 0.2, P_f = 0.1$ for any arbitrary number of users (Fig-4.3). The next Section provides case study analysis of HF fusion for AND/OR/Majority rules simulation.

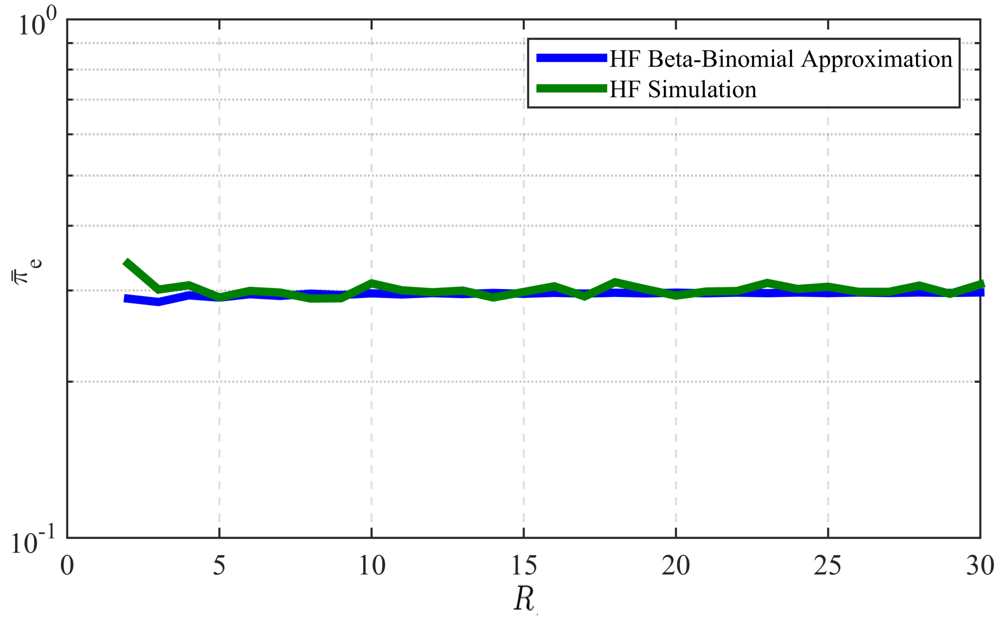


Fig. 4.4 Mean prediction error Vs number of users (15 Users)

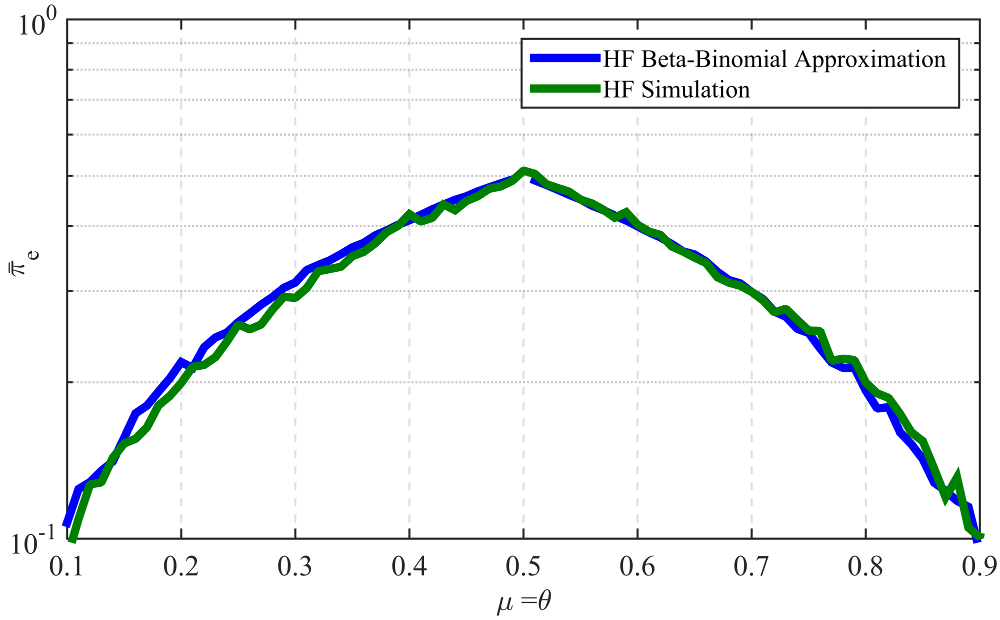


Fig. 4.5 Mean Prediction Error for $\mu = \theta$ (15 users)

## 4.4 Case study analysis

The case study contribution of this chapter published in the conference proceedings of the IEEE International Conference on Communications (**ICC'2016**) workshops

[3]. The work in [3] extended the prediction accuracy performance analysis of local HMM SOP in [2] to homogeneous hard fusion based cooperative SOP. The results have previously shown that hard decision fusion considerably improved the prediction performance compared to local user SOP. The simulation results are presented against the Binomial approximation in Subsection-4.3.2.

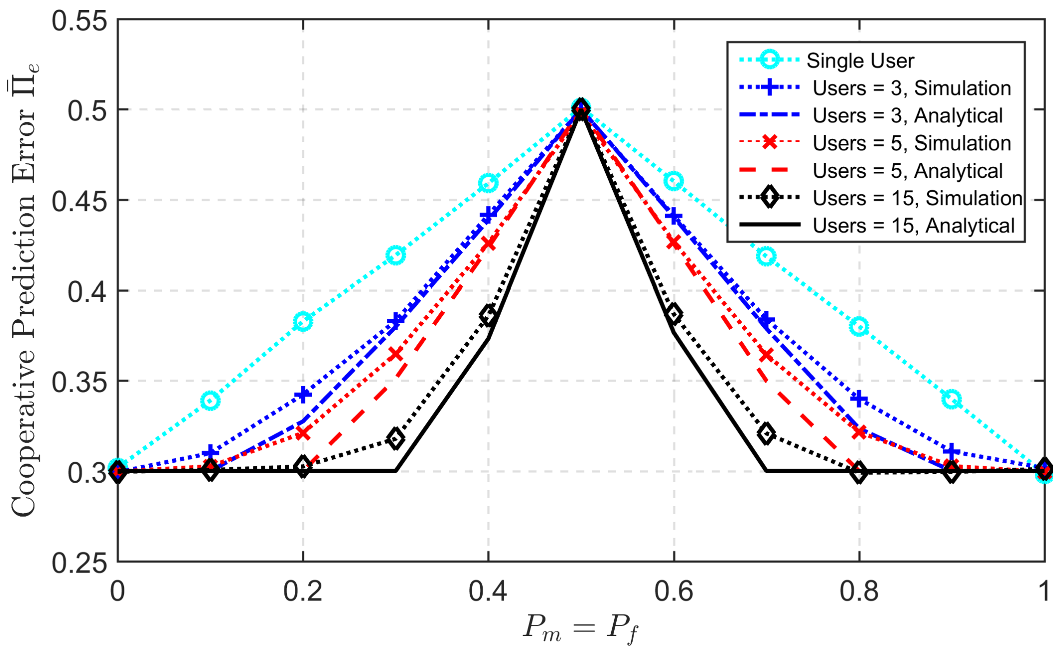


Fig. 4.6 Cooperative Prediction Error for Hard Combining, for 3/5/15 Users, for Majority Rule

Fig.4.6 and Fig.4.7 display mean cooperative prediction error ($\bar{\Pi}_e$) for *Majority*, and *AND/OR* rules respectively. The model values are set at ($\mu = \theta = 0.7$), and only values of ($P_m = P_f$) are displayed. Similarly, in Fig.4.6 for majority rule fusion, the cooperative prediction error improves as the number of users increase. For different pairs of $P_m, P_f$ in both figures, we confirm the lower limit of perfect sensing for cooperative prediction error $\bar{\Pi}_e$. Cooperative prediction error $\bar{\Pi}_e$ improves for *Majority*) / degrades (*AND/OR*) for channel error case $P_m, P_f > 0$, where the level of improvement is a function of the number of users. Similarly, Fig-4.8, and Fig-4.9 present the partial prediction error (in Subsection-4.3.1) values for 3/5/15 users when $P_m = Pf$. Fig-4.8 present the prediction miss-detection error, while Fig-4.9 present prediction false alarm error. Mean prediction error in Fig.4.6 is weighted sum (of stationary distribution) of these partial prediction error. Prediction miss-detection

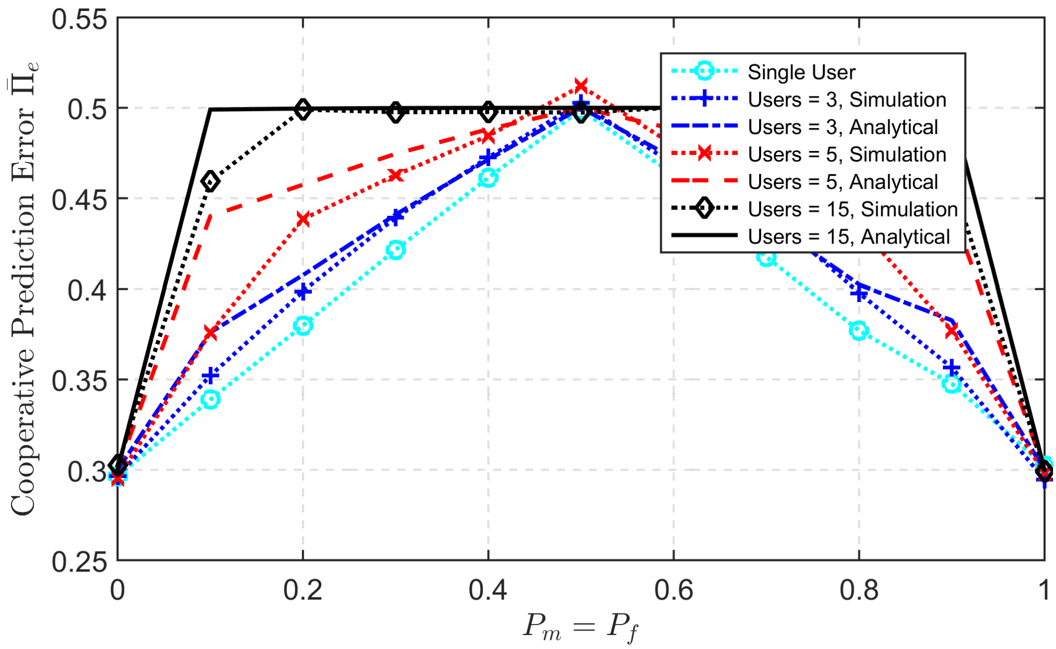


Fig. 4.7 Cooperative Prediction Error, Hard Combining for 3/5/15 Users for OR/AND Rules

and false alarm are only presented for majority rule. The partial prediction error values complement the results in Fig-4.6.

Fig-4.10 for prediction false alarm, and Fig-4.11 prediction miss-detection, present the same scenario for AND/OR/Majority rules for 15 users only. The figure displays binomial *analytical* limits where we can distinguish the lower/upper limit characteristics of the approximation. The figures complement Fig.4.7 since mean prediction error in Fig.4.7 is a weighted sum of of the partial prediction errors.

Finally, Fig-4.12 for prediction miss-detection, and Fig-4.13 for prediction false alarm present different detection false alarm $P_f$ settings for 15 *Majority* rule cooperative users. The figures are the generalized case of partial error Fig-4.8, and Fig-4.9, respectively. The figures highlights how the maximum prediction error shifts as values of $P_m, P_f$ varies, where the curves are shifted version with the maximum error is achieved when $P_f = 1 - P_m$ . In Chapter-3, Section-3.4 case study analysis (Fig.3.8, and Fig-3.9), we highlighted the observation where the mean prediction error($\bar{\pi}_e$) approaches the maximum 0.5 i.e. no information can be deducted from the stochastic source when the values ($P_m + P_f = 1$ or $P_f = P_d$).

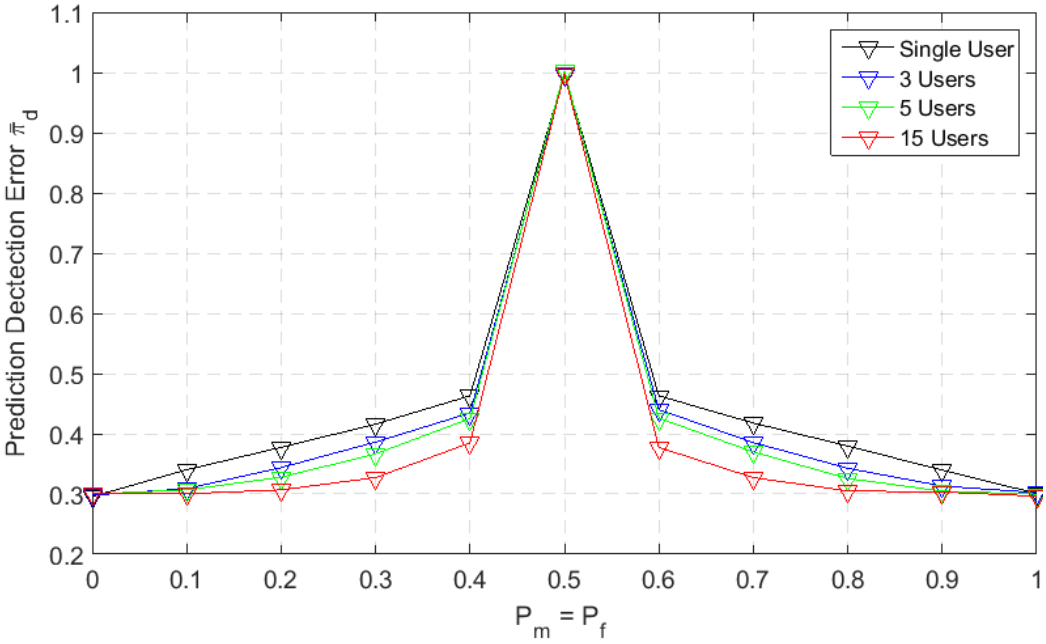


Fig. 4.8 Cooperative prediction miss-detection Error, Hard Combining for 3/5/15 Users for Majority Rule

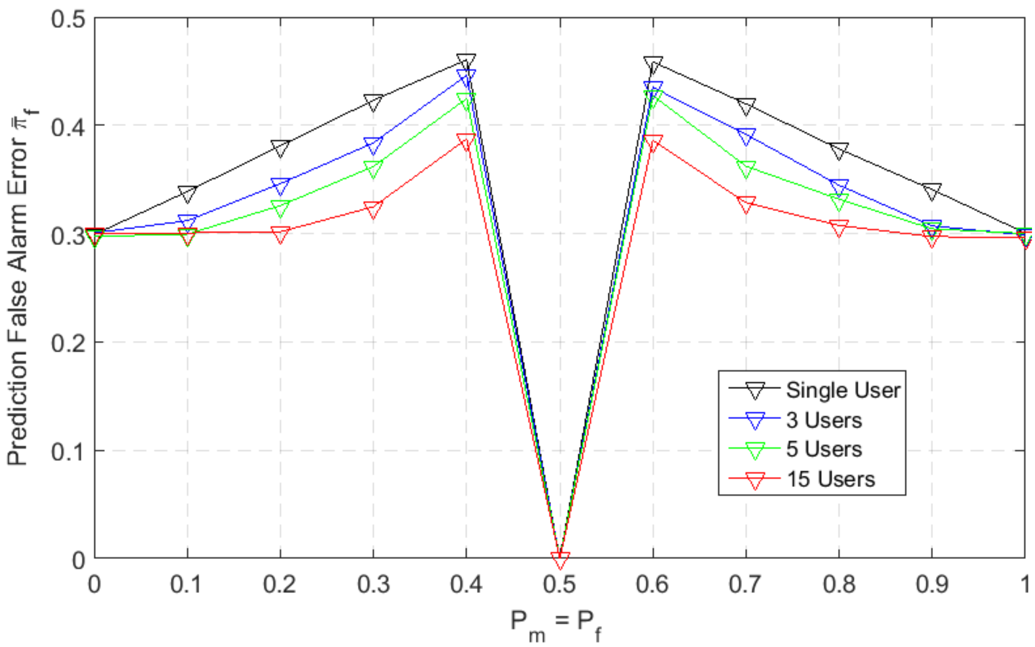


Fig. 4.9 Cooperative prediction false alarm Error, Hard Combining for 3/5/15 Users for Majority Rule

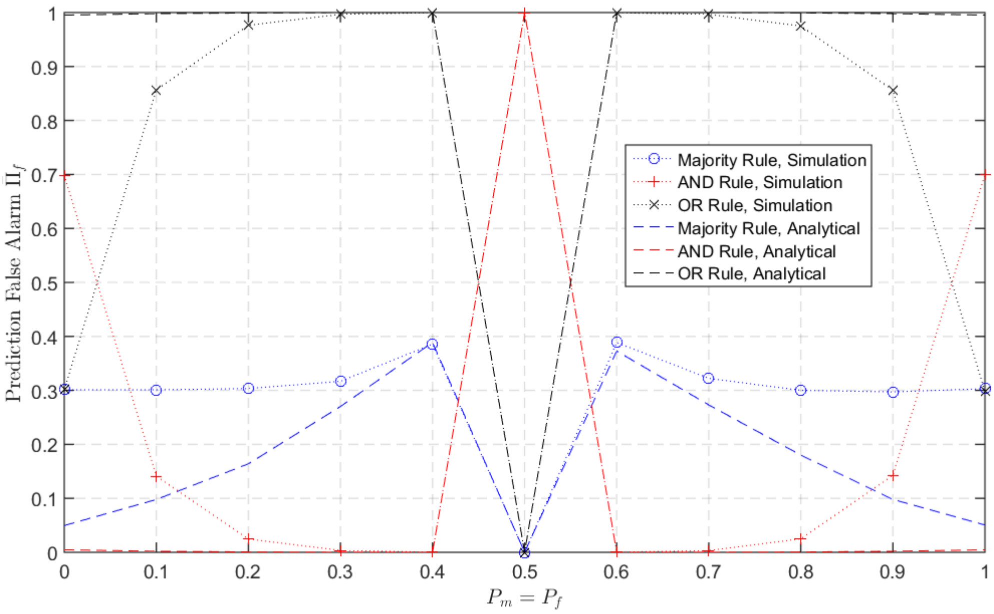


Fig. 4.10 Prediction false alarm Error, Hard Combining for 15 Users for $P_m = P_f$

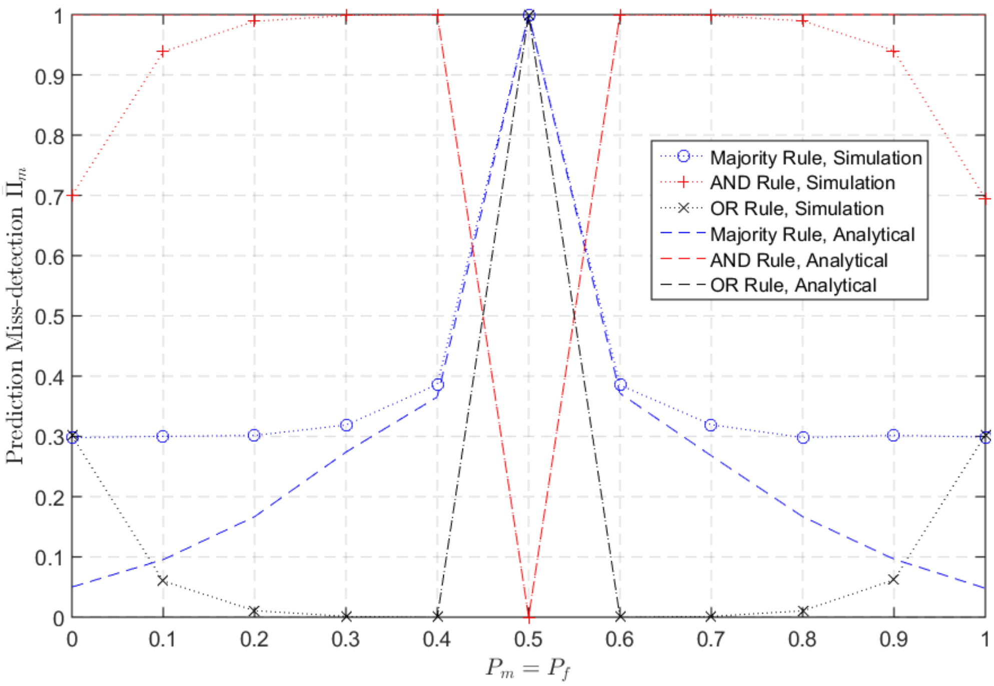


Fig. 4.11 Prediction miss detection Error, Hard Combining for 15 Users for $P_m = P_f$

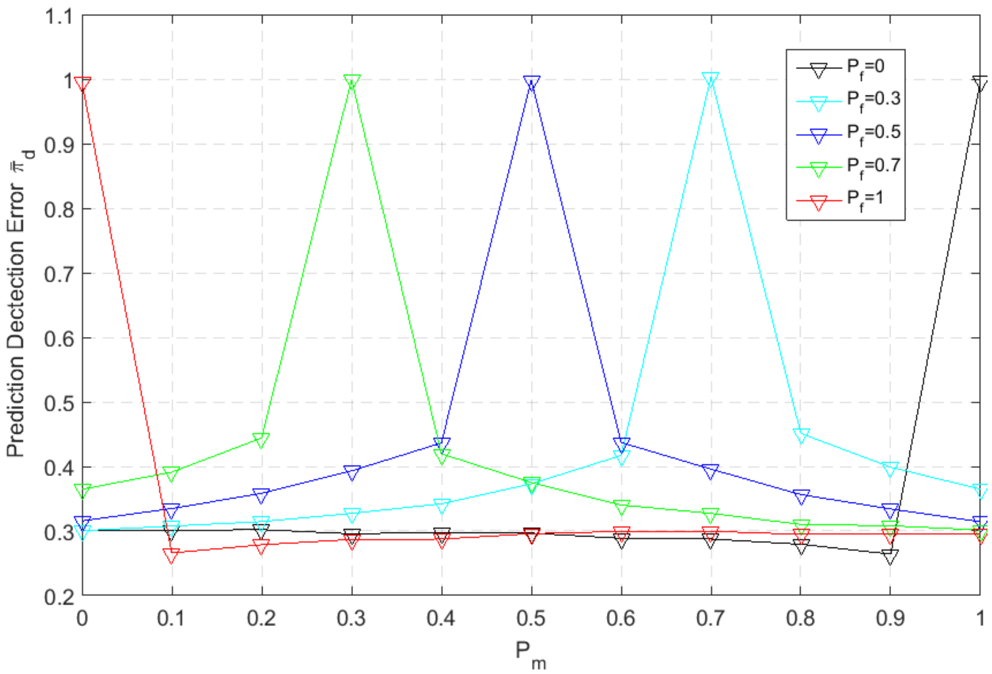


Fig. 4.12 Prediction miss detection Error, Hard Combining for 15 Users

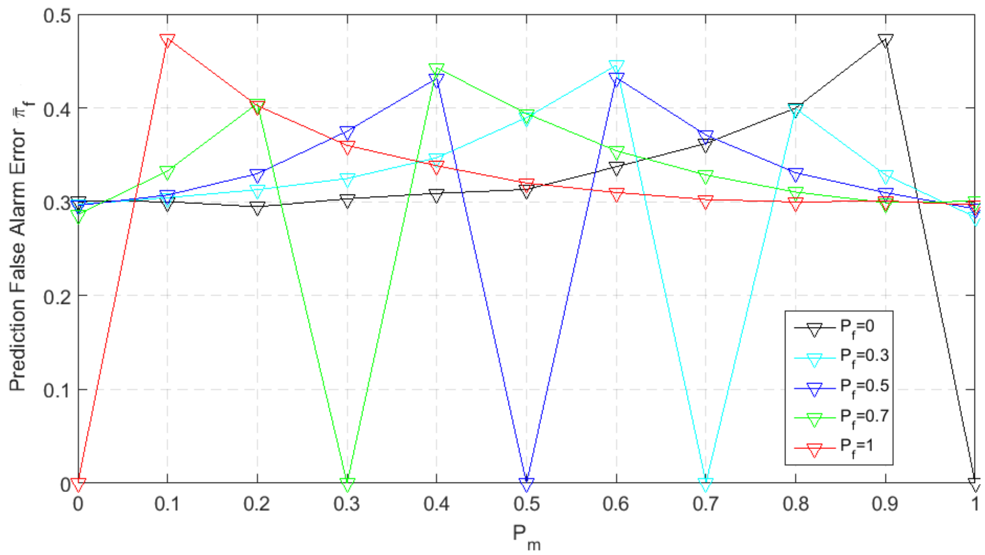


Fig. 4.13 Prediction false alarm Error, Hard Combining for 15 Users

**Results Summary** In conclusion, the choice of the threshold for the $m$ out of $R$ in hard fusion is a function of the threshold adopted by each secondary user. The lowest cooperative prediction error is obtained when hard fusion is identical to the *local prediction* threshold. Thus, $\frac{m}{R} = \zeta$ reflects the condition for the highest level of improvement in prediction accuracy. For example, in Fig.4.6 the majority rule and $\zeta = 0.5$ provides the lowest cooperative prediction error $\bar{\Pi}_e$, while the choice of $\Psi = 0.5$ for both *AND/OR* rules caused degradation in $\bar{\pi}_e$. Secondly, the choice of the hard fusion threshold can be adjusted to optimize partial prediction errors i.e. cooperative prediction miss-detection $\bar{\pi}_m$, or cooperative prediction false alarm $\bar{\pi}_f$. When the design requirements dictates weighted significance of either errors, different threshold for cooperative hard fusion may provide better performance. The performance of cooperative prediction error $\bar{\pi}_e$ of *homogeneous* secondary users, is traceable and identical. However, in the *heterogeneous* case, it has dependency of the choice of the clustering mechanism. Thus, the impact of cooperative hard fusion is dependent upon the variance of the spectrum sensing error probabilities $(P_m^r, P_f^r, \quad r \in 1, .., R)$.

## 4.5 Summary

This chapter presented a new generalized Beta-Bernoulli approximation of the predictive posterior probability for local HMM based SOP. For cooperative prediction, a new generalized Beta-Binomial **GBB** approximation of HMM decision fusion is consequently proposed. The effectiveness of GBB approximation is assessed in terms of mean prediction error. The proposed GBB approximation is no worse than a loose upper bound for HMM stationary posterior distribution.

Additionally, we presented an investigative analysis of the hidden Markov model based prediction, and simulated the performance of mean prediction error against the model parameters in terms of channel sensing errors, and channel occupancy transitions. We presented Hard fusion based cooperative spectrum prediction, and the potential for better accuracy compared to local spectrum prediction. Future research will investigate soft fusion cooperative spectrum prediction.

# Chapter 5

# Cooperative Soft Fusion for HMM based Spectrum Occupancy Prediction

In this chapter, soft decision based fusion for cooperative SOP is presented based on the local prediction model presented in Chapter-3. In this chapter, we propose different test statistics based on the prediction probabilities, signal to noise ratio, and detection probabilities. Additionally, well known data fusion methods for soft test statistics such as the maximal ratio, equal gain, and selection combining are utilized for case study analysis for soft fusion based cooperative prediction. Finally, The soft fusion based techniques are benchmarked based on hard fusion performance presented in Chapter-4.

The case study analysis of soft fusion based SOP performance is analysed in terms of the mean prediction error [2, 3]. Furthermore, soft fusion techniques are presented are compared to local spectrum prediction, and hard fusion cooperative prediction to assess the possible gains of soft fusion based prediction [2, 3]. The results show soft fusion is superior, and more robust compared to hard fusion in terms of prediction accuracy. Section-5.2 describes single SU local prediction model, while Section-5.3 presents cooperative soft fusion prediction techniques. Performance analysis, and concluding remarks are presented in Section-5.4, and Section-5.5, respectively. The contribution of this chapter is submitted to IEEE communication letters.

## 5.1 Introduction

Statistical *spectrum occupancy prediction* (**SOP**) in dynamic spectrum access (**DSA**) systems re-utilises the information obtained by spectrum sensing to predict the underlying spectrum occupancy patterns [8, 45]. Sensing scheduling, channel selection, and proactive hand-off can be optimised using spectrum prediction decisions. SOP models are commonly based on Poisson processes, linear regression, neural networks, and space vector machines [11]. The surveys in [4, 10, 11] provide the taxonomy and framework of SOP techniques.

Finite state Markov machines (**FSM**) constitute a major portion of *local* spectrum modelling literature. FSM based single user local predictors estimate the probability of available time slot at one step-ahead in time [3]. Subsequently, cooperative multi-user scenarios utilise local prediction probability at the secondary user (**SU**) nodes as fusion test statistics. Cooperative prediction exploits the shared test statistics to improve collective prediction accuracy. However, literature on cooperative spectrum prediction is limited, and only a few address cooperative SOP based on soft fusion (**SF**). The work in [115] proposed a cooperative Bayesian non-parametric framework for primary user (**PU**) transmission monitoring. The model tracks PU signal amplitude at each SU node based on non-linear particle filter. The study presented a quantitative comparison between communication cost, and prediction accuracy for single and cooperative tracking scenarios. In SOP literature, the work in [43] proposed a coalition game for cooperative multi-primary user activity monitoring based on HMM predictors. However, prediction accuracy results were limited to cumulative prediction error for multiple PU scenarios. Prediction accuracy gains from the fusion technique, or robustness to wireless channel conditions were not explicitly identifiable.

he work in [3] extended the prediction accuracy performance analysis of local HMM SOP in [2] to homogeneous hard fusion based cooperative SOP. The results have shown that hard decision fusion considerably improved the prediction performance compared to local user SOP. To the authors knowledge, soft fusion of HMM prediction is not addressed in SOP literature. In this letter, soft decision based fusion for cooperative SOP is addressed, and new proposed SF techniques are presented. SF based SOP performance is analysed in terms of the mean prediction error performance. Furthermore, SF techniques are presented and compared to local spectrum prediction, and hard fusion cooperative prediction [2, 3]. The results show SF is superior, and more robust compared to hard fusion in terms of prediction accuracy. Section-5.2 describes single SU local prediction model, while Section-5.3 presents cooperative SF prediction techniques. Performance analysis is then presented in Section-5.4.

## 5.2 System Model

Soft fusion based prediction assumes a discrete slotted time system with $R$ number of SU nodes distributed in space. Spectrum sensing collects spectrum occupancy information at each time slot. The PU spectrum occupancy at time instant $t$ is given by $x_t$ modeled as a first order Markov chain, and the SU observation of the spectrum used by the PU is defined as $y_t$. Thus, the stochastic model that describes the spectrum occupancy/observation of SU/PU is given by a Hidden Markov Model (HMM). Accordingly, at time instant $(t-1)$, using an observation series $y_{1:t-1}$, each SU estimates the probability of the presence or absence of a PU for time slot $t$, and based on this probability when a binary decision is made for the presence/absence of the PU, we refer to this as local SOP. The local observations at $(t-1)$ on the other hand, and any related test statistics, generated by each SU can be shared with a fusion center (FC) to cooperatively predict the PU occupancy, which we refer to as cooperative SOP. When information other than local binary decisions for SOP are shared with the FC then we refer to it as soft decision fusion based cooperative SOP, which is the topic of the work presented here. Local and cooperative prediction system model is presented in Fig-5.1.

### 5.2.1 Primary user activity model

Define $x_t \in \{0,1\}$ as the PU channel occupancy state at time slot $t$. Then, $x_t$ is modelled as an irreducible stationary Markov chain, where $x_t = 1$ and $x_t = 0$ represent an busy and available channel, respectively. First order Markov chain assumes the current state $x_t$ depends only on the previous state $x_{t-1}$. The transition probability matrix $p(x_t|x_{t-1} = i)$ for the two state Markov chain is given by:

$$\mathbf{P} = [p_{ij}] = \begin{bmatrix} \beta & 1-\beta \\ 1-\mu & \mu \end{bmatrix} \tag{5.1}$$

where, $(\mu, \beta)$ represent the probability of remaining in busy and idle states, respectively. Given an initial condition $x_0$, the PU's occupancy can be fully characterized

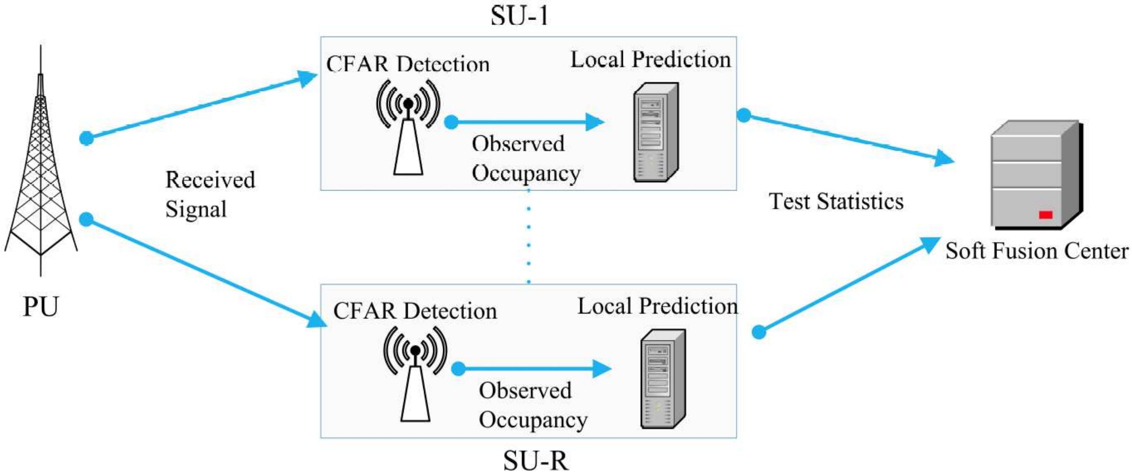


Fig. 5.1 Local and Cooperative Spectrum Prediction Model

by $\{p(x_0), \mathbf{P}\}$.

### 5.2.2 Occupancy Observation Model with Spectrum Sensing

The SU observation of the PU's spectrum occupancy reflects the wireless channel characteristics, and depends on the performance of spectrum sensing procedure. The observation errors due to spectrum sensing are assumed due to log-normal shadowing, and white Gaussian noise. The performance of the energy detector sensing is quantified by the probability of detection $P_d(r)$, and false alarm $P_f(r)$ for the $r^{th}$ SU node, where $r \in \{1, \ldots, R\}$. The detection and false alarm probabilities are approximated using the central limit theorem for a large number of samples $N$ as:[51]

$$P_d(r) \approx Q\left(\frac{\lambda - \sqrt{N}\sigma_w^2(1+\rho_r)}{\sqrt{N}\sigma_w^2(1+\rho_r)}\right)$$
$$P_f(r) \approx Q\left(\frac{\lambda - N\sigma_w^2}{\sqrt{N}\sigma_w^2}\right) \tag{5.2}$$

where, $\lambda$ is the detection threshold, $\sigma_w^2$ is the noise power, and $\rho_r$ is the signal to noise ratio (SNR) at the $r^{th}$ SU node which is Gaussian distributed in dB scale with a standard deviation (Shadowing parameter) of $\sigma_s$(dB). Each SU is assumed to utilize a constant false alarm probability (**CFAR**) strategy [45] for deciding on $\lambda$. HMM process is characterized by the transition, and emission probability matrices, $\mathbf{P}$ and

$\mathbf{E}_r$, respectively, where the emission matrix $\mathbf{E}_r$ for the $r^{th}$ SU is given by:

$$\mathbf{E}_r = \begin{bmatrix} 1 - P_f(r) & P_f(r) \\ 1 - P_d(r) & P_d(r) \end{bmatrix} \tag{5.3}$$

### 5.2.3 Local spectrum prediction model

Given the emission matrix probabilities, the SU calculates the probability of free/occupied $p(\hat{x}_{r,t}|y_{r,1:t-1})$ in a two-step recursive process of *Prediction* (Equation-5.4), and *Update* (Equation-5.5) [126], where the update equation is used to predict the subsequent state of the occupancy $x_{t+1}$ as the process continues iteratively.

$$p(\hat{x}_{r,t}|y_{r,1:t-1}) = \int_{\mathcal{X}} p(x_t|x_{t-1})p(\hat{x}_{r,t-1}|y_{r,1:t-1})dx_{t-1} \tag{5.4}$$

$$p(\hat{x}_{r,t}|y_{r,1:t}) = \frac{p(y_{r,t}|x_{r,t})p(\hat{x}_{r,t}|y_{r,1:t-1})}{p(y_{r,t}|y_{r,1:t-1})} \tag{5.5}$$

$$\text{where,} \quad p(y_{r,t}|y_{r,1:t-1}) = \int_{\mathcal{X}} p(y_{r,t}|x_t)p(\hat{x}_{r,t}|y_{r,1:t-1})dx_t$$

The computations above require the estimates of $\mathbf{P}$ and $\mathbf{E_r}$, that is model parameter estimations. At each SU node $\mathbf{E}_r$ is obtained by computing $P_d(r)$ and $P_f(r)$ using (5.2) by knowing all the related parameters, and the Baum-Welch training [**?** ] is used to iteratively calculate $\mathbf{P}$. The test statistic $p(\hat{x}_{r,t}|y_{r,1:t-1})$ defined as the *predictive posterior probability* is a function of individual SU observation series and thus an analytical expression is hard to obtain in a closed form for arbitrary HMM models. Therefore, numerical recursive calculation based on expectation-maximization algorithms is used [53]. Finally, the local SOP is made by performing a binary decision using a detection threshold $\zeta$, given by,

$$\hat{x}_{r,t} = \begin{cases} 0 & \text{Free} \quad p(\hat{x}_{r,t}|y_{r,1:t-1}) \geq \zeta \\ 1 & \text{Occupied} \quad p(\hat{x}_{r,t}|y_{r,1:t-1}) < \zeta \end{cases} \tag{5.6}$$

To evaluate the local prediction performance, single user mean prediction error can be calculated from the predictive posterior probability [2, 3]. Empirically, under the two state HMM model prediction error is a Bernoulli random variable $E_t = \hat{x}_{r,t} \oplus x_t$, $E_t \in [0, 1]$. Then, $\pi_t$ and $\bar{\pi}_e$ are the instantaneous prediction error at time instant

$t$, and the mean prediction error respectively [2, 3].

$$\begin{aligned} \pi_t &= P(E_t = 1) \\ \bar{\pi}_e &= \lim_{t\to\infty} \mathbf{E}(\pi_t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(\pi_i) \end{aligned} \tag{5.7}$$

The cooperative mixture model can be approximated based on $\psi_{r,t,\theta_r}$.

$$p_w(\bar{x}_{1:Rt}|y_{1:R,1:t}) = \sum_{r=1}^{R} w_r(\theta_r) \times \psi_{r,t,\theta_r} \tag{5.8}$$

However, theoretical approximation is intuitive to calculate for all fusion rules. For equal gain soft fusion, the mean value $\mathbb{E}(\psi_{r,t})$ can be utilised to calculate the error, while the maximum value serves as an indicator for selection combining. Theoretical approximation of soft fusion error is beyond the scope of this work. Instead simulation bassed case study analysis is presented in Section-5.4. Proposed techniques along with those in literature are presented in Section-5.3.

## 5.3 Soft cooperative Prediction Techniques for Spectrum Occupancy

In this section, we present the proposed cooperative prediction methods based on SF of local prediction probabilities. The local test statistics $p(\hat{x}_{r,t}|y_{r,1:t-1})$ for CFAR SU is characterized by $\rho_r$ and $P_d(r)$ [2, 3]. In order to fuse the test statistics for cooperative prediction, three known strategies are considered: (i) maximal-ratio, (ii) equal-gain, and (iii) selection combining methods. It is worth noting that, though the SF strategies are known in literature, the SF test statistics and the concept of soft-cooperative prediction are identified as novel in this work.

### 5.3.1 Maximal-Ratio Based Soft Fusion

Given a non-negative normalised weighting function $w(\theta_r)$ over all the $R$ local test statistics $p(\hat{x}_{r,t}|y_{r,1:t-1})$, the cooperative SF based prediction probability using maximal ratio combining (**MRC**) is defined as follows:

$$p_w(\hat{x}_{1:R,t}|y_{1:R,1:t-1}) = \sum_{r=1}^{R} w_r(\theta_r) p_r(x_{r,t}|y_{r,1:t-1}, \theta_r) \tag{5.9}$$

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

where $\theta_r$ is the test statistic $\theta_r$ at the $r^{th}$ SU node. We propose four techniques based on MRC using three different functional models for the parameter $\theta_r$ as given below:

- **Method MRC-1:** Based on test statistics $\theta_r = p(\hat{x}_{r,t}|y_{1:t-1})$ The observation sequence by $r$th SU is a function of the PU activity, channel propagation and signal detection model i.e. $y_{r,1:t} \sim p(y_{r,1:t}) = f(t, \rho_r|P_d, P_f, \mu, \theta_r)$. Thus, the predictive posterior probability measure the confidence of each secondary user that the spectrum opportunity is available/occupied (Section-3.2). Accordingly, the technique to exploit diversity in secondary user's prediction decision, without additional side information exchanged with fusion centre.

- **Method MRC-2:** Based on the probability of detection $\theta_r = P_d(r)$ The probability of successful detection is a function of signal detection model (Subsection-1.4.2). Accordingly, the technique to reflect the heterogeneous uncertainty in prediction decision as a function of detection accuracy. However compared the MRC-1, it ignores the uncertainty in the prediction decision induced by the observation series, and HMM model parameters. Practically, each SU would transmit the value along with fusion test statistics to the soft fusion centre.

- **Method MRC-3:** Based on the *log* of probability of detection$\theta_r = \log(P_d(r))$ This technique is a log transformed version of MRC-2 above, and is aimed to add robustness to the soft fusion decision. The log transformation is used to reduce skewness in the probability of detection values. the skewness is introduced by the wireless channel impairments and spatial distribution could result in $P_d$ values with high leverage and influence on the cooperative decision.

- **Method MRC-4:** Based on signal to noise ratio $\theta_r = \rho_r$
  This diversity combining techniques adjust the gain from each secondary user proportional to the signal, and inversely proportional to the channel noise. The

well known soft fusion technique [45] is the baseline of maximal ratio combining methods MRC-1,MRC-2, and MRC-3.

The variations in the weighting functions given is expected influence the performance of the cooperative mean prediction error as further analysed in the simulation section. At an equal distance from the PU, SNR values $\rho_r$ are identical for all SU's. Then, MRC approach is expected to perform similar to equal gain approach described below.

### 5.3.2 Equal-Gain Based Soft Fusion

Equal gain (**EG**) combining assumes all secondary users have an equal "weight" i.e. $w_r = \frac{1}{R}$. The fusion strategy ignores the heterogeneous nature of SU's detection and prediction performance. However, it does not require the additional computational operations compared to MRC.

$$p_w(\hat{x}_{1:R,t}|y_{1:R,1:t-1}) = \frac{1}{R}\sum_{r=1}^{R} p_r(x_{r,t}|y_{r,1:t-1}) \tag{5.10}$$

### 5.3.3 Selection Combining Based Soft Fusion

In the selection combining (**SC**) based approach the best test statistic from set In the selection combining (SC) based approach the test statistic with the maximum SNR out of all the SU nodes is selected to perform the prediction at the fusion centre, as given below.

$$p_w(\hat{x}_{1:R,t}|y_{1:R,1:t-1}) = \max_{\rho_r} p(\hat{x}_{r,t}|y_{r,1:t-1}, \rho_r) \tag{5.11}$$

Note here that the shadowing in the wireless channel will have a significant influence in the fusion process for SC. At an equal distance from the primary user, signal to noise ratio values $\rho_r$ are identical for all SU's. Then, soft fusion weighting function $(w_{\theta_r})$ is similar to equal case approach.

## 5.4 Performance analysis

Prediction accuracy of local and cooperative prediction can be quantified by the *mean prediction error* $(\bar{\pi}_e)$ [2, 3], where $E_t = \hat{x}_{1:R,t} \oplus x_t$ in (5.7) for the cooperative prediction case. The error performance of soft cooperative SOP is compared against the local prediction $(\hat{x}_{r,t})$, and also against ideal sensing $(P_D(r) = 1, P_F(r) = 0)$ cases. Moreover, hard fusion (**HF**) based cooperative SOP [3] is also compared to SF techniques to highlight the improvement in the SF approach. This analysis consider the *homogeneous* users scenario where all users are at equal distance from the primary user [3], noting that the treatment of random SU positions requires the inclusion of stochastic geometry based modelling for accurately studying the prediction performance and is not in the scope of this paper. The values of the test statistics $p(\hat{x}_{r,t}|y_{r,1:t-1}, \rho_r)$ are stochastic in nature based on the system model described in Sec-II. Table-5.1 presents the parameter values for local SOP model used in the simulations, where we assume $\zeta, P_f(r)$ and $\sigma_w^2$ are the same for all the SUs. In which case our performance analysis reflects the improvement considering the use the spatial diversity for SOP.

Table 5.1 Simulation Parameters

| Variable | Definition | Value |
|---|---|---|
| $\mu$, $\beta$ | Transition matrix parameters | $0.7, 0.7$ |
| $\zeta$ | decision threshold | 0.5 |
| $\sigma_w^2$ | | -95 dBW |
| $R$ | Number of Users | 5, 15 |
| $P_f$ | Probability of false alarm | 0.1 |
| | Simulation Runs per sample | $40,000$ |

Fig-5.2 presents the mean prediction error curves for all the techniques described in this work and are compared with other techniques. The performance curves for the MRC and EG techniques overlay as seen in Fig-5.2 due to the homogeneous simulation environment. Minimum prediction error for local prediction error is achieved under high detection accuracy (at $\rho_r > -10$ dB or $P_d > 0.9, P_f = 0.1$). For lower signal to noise ratio values, HMM local prediction error performance degrades as a function of signal to noise ratio $\rho_r$. Consequently, cooperative prediction is expected to enable predictors to reduce prediction error, and maintain minimum error at worse channel conditions ($\rho_r$) compared to local prediction. It was previously shown that hard fusion of local prediction decision converges faster toward minimum error at lower detection accuracy values compared to local prediction (Fig-5.3)[3]. HF prediction error binomial approximation, as well as simulated cooperative prediction error are presented in [3]. The binomial approximation error is lower bounded by a minimum prediction error defined under ideal sensing conditions [2]. Comparatively,

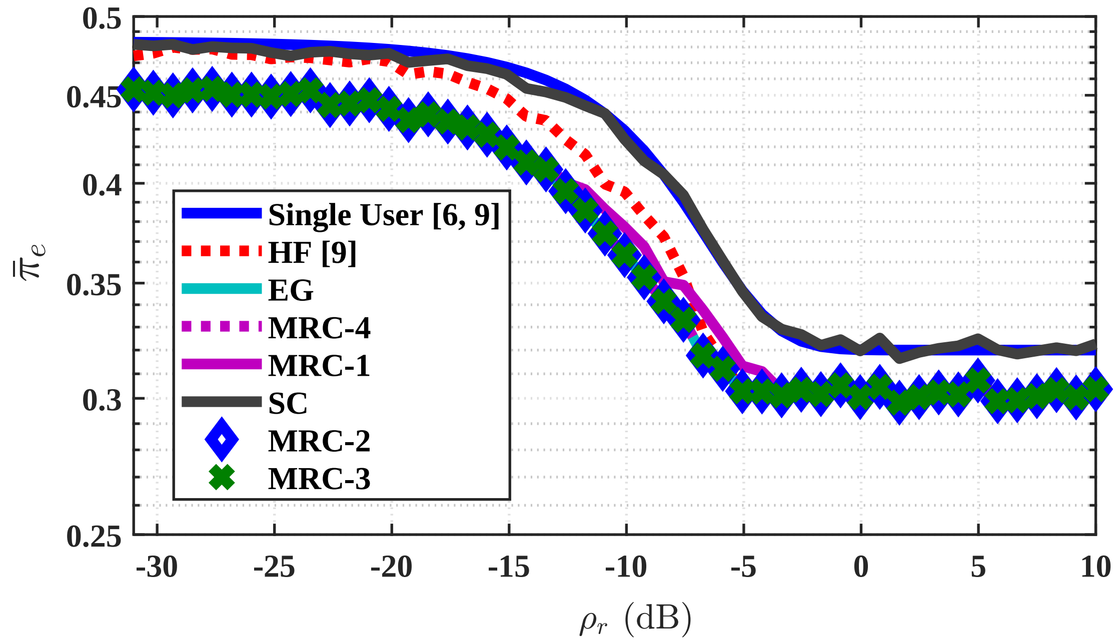


Fig. 5.2 Comparing the soft fusion mean prediction error for 15 SUs

SF techniques maintain minimum prediction error at even lower signal to noise ratio values. Moreover, SF decreases the maximum prediction error as a function of the number of cooperative users a feat not achieved by HF prediction (Fig-5.2 and Fig-5.3). Additionally, mean prediction error for SF techniques (EG/MRC) reduce more rapidly as a function of the number of cooperative user compared to HF techniques (Fig-5.3).

SF techniques add robustness to cooperative prediction error performance under log-normal shadow fading (Fig-5.4). Firstly, SC error performance is more robust compared to single user case as the fusion centre prioritizes the best user in terms of $\rho_r$. Secondly, SF techniques (EG and MRC) are more robust compared to HF techniques under poor channel condition $\rho_r < -15$ dB. Finally, SF techniques (EG and MRC) are robust for different values of false alarm i.e. for different settings of CFAR detection strategy (Fig-5.5). In contrast, hard fusion performance degrades under the same settings toward local prediction values. In Summary, SF techniques provide faster convergence toward minimum prediction error as a function of the number of cooperative users. Robustness under shadowing, and detection performance are superior to hard fusion techniques in terms of prediction accuracy.

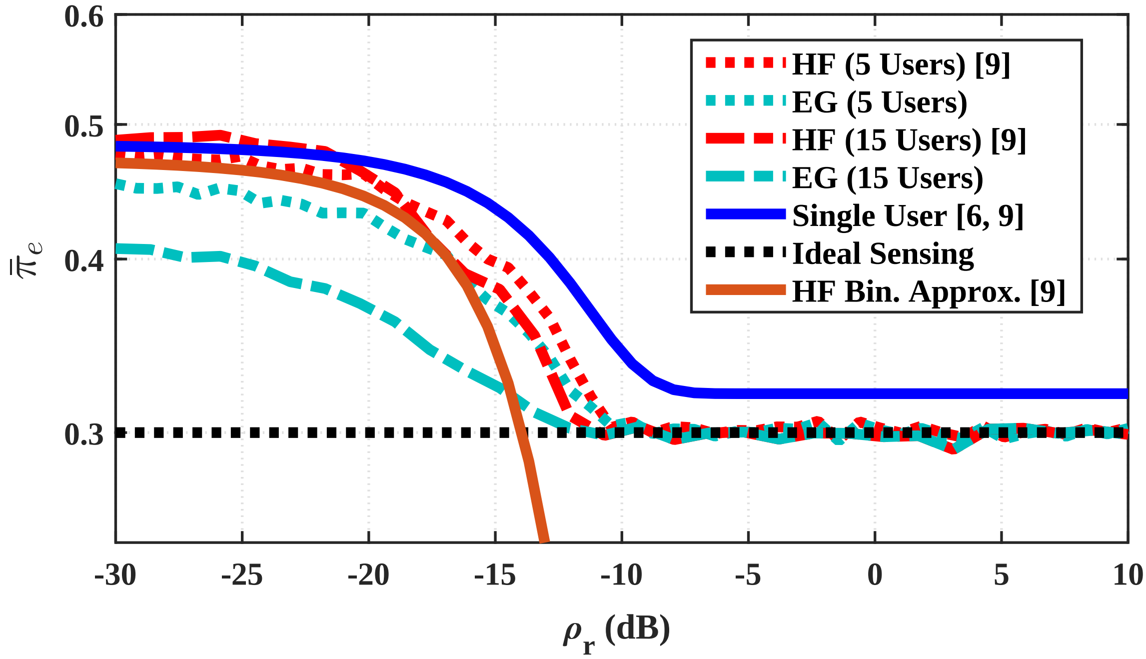


Fig. 5.3 Comparing the soft fusion mean prediction error for different number of SUs

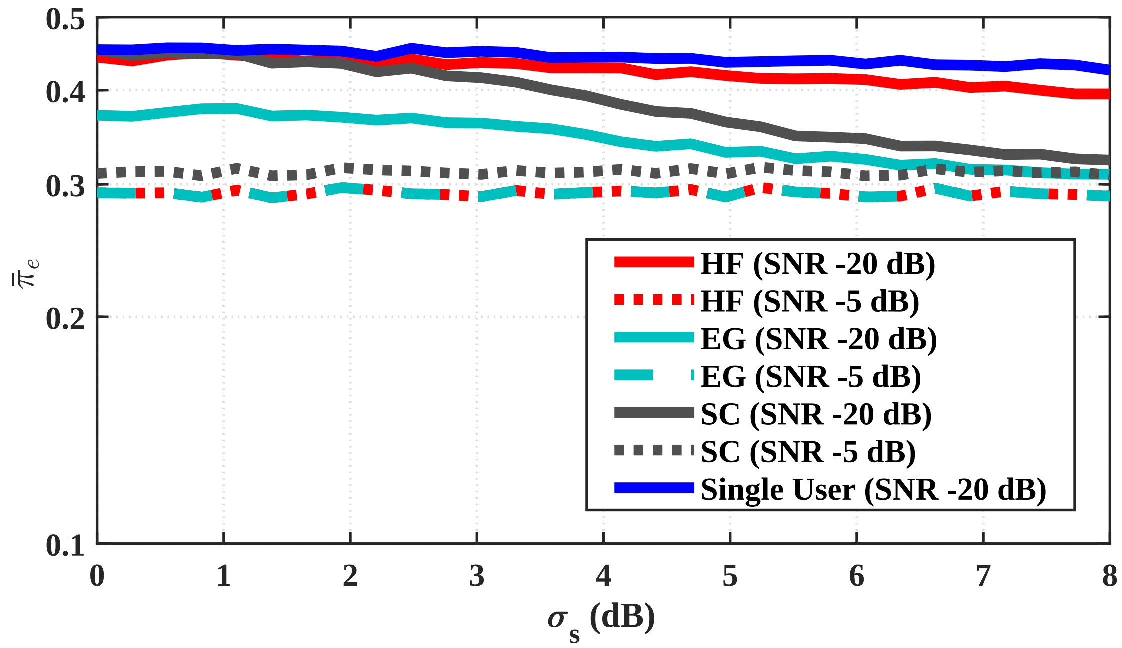


Fig. 5.4 Soft fusion mean prediction error under different shadowing levels

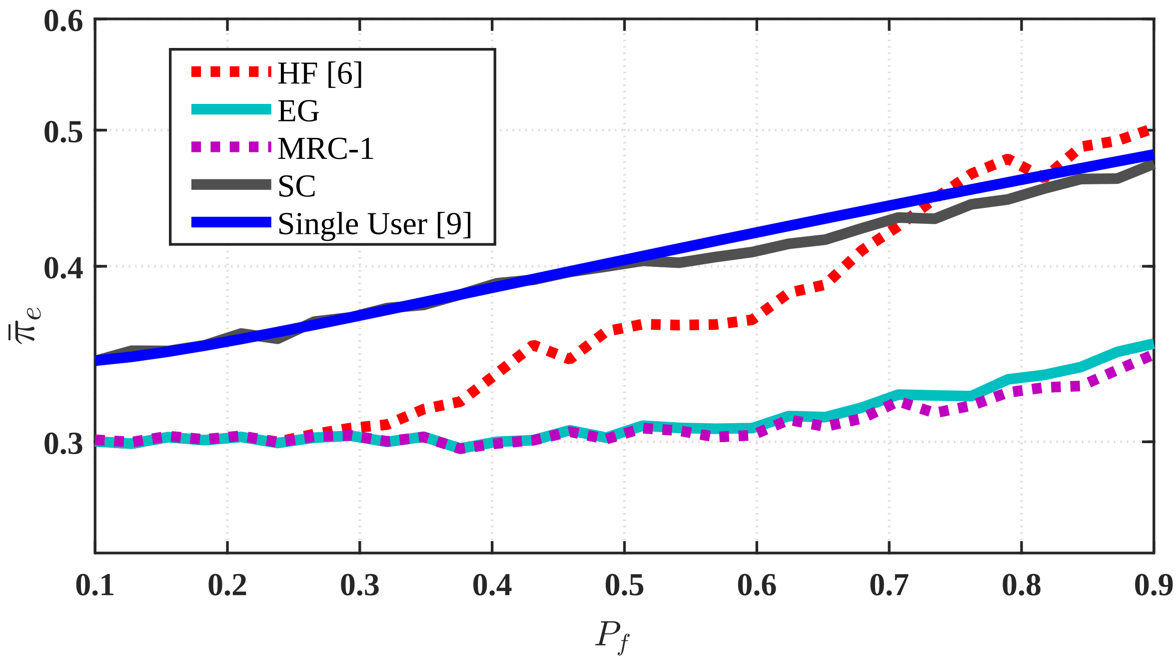


Fig. 5.5 soft fusion mean prediction error for different probability of false alarm values

Minimum prediction error for local prediction error is achieved under high detection accuracy (at $\rho_r > -10$ dB or $P_d > 0.9, P_f = 0.1$). For lower signal to noise ratio values, HMM local prediction prediction error performance diverge from the minimum prediction error as a function of signal to noise ratio $\rho_r$. Cooperative prediction is expected to enable predictors to reduce maximum prediction error, and maintain minimum error at worse channel conditions ($\rho_r$) compared to local HMM prediction. It was previously shown that hard fusion of local prediction decision converges toward minimum error at lower detection accuracy values (Fig-5.3). HF prediction error Binomial approximation, as well as simulated error cooperative prediction are presented in our work [3]. The binomial approximation error is lower bounded by minimum prediction error defined under ideal sensing conditions [2]. Comparatively, soft fusion techniques maintain minimum prediction error at even lower signal to noise ratio values. Moreover, soft fusion decreases the maximum prediction error as a function of the number of cooperative users a feat not achieved by HF prediction (Fig-5.2 and Fig-5.3). Additionally, mean prediction error for soft fusion techniques (EG/MRC) reduce more rapidly as a function of the number of cooperative user compared to HF techniques (Fig-5.3).

Soft fusion techniques add robustness to cooperative prediction error performance under log-normal shadow fading (Fig-5.4). Firstly, SC error performance is more robust compared to single user performance since the fusion centre prioritize the best user in terms of $\rho_r$. Secondly, soft fusion techniques (EG and MRC) are more robust compare to HF techniques under poor channel condition $\rho_r < -15$dB.Finally, soft fusion techniques (EG and MRC) are robust for different values of false alarm i.e. for different settings of CFAR detection strategy (Fig-**??**). In contrast, hard fusion performance degrades under the same settings toward local prediction values.

**Results Summary** Soft fusion techniques provide faster convergence toward minimum prediction error as a function of the number of cooperative users. Robustness under shadowing, and detection performance are superior to hard fusion techniques diversity gains in terms of prediction accuracy. Finally, alternative maximal ratio combining based techniques conceptually requires less common control channel capacity with identical performance to legacy techniques.

## 5.5 Summary

This chapter demonstrated soft fusion based techniques to improve prediction error, by using spatial diversity, compared to single user error performance. Soft fusion techniques are robust under shadowing, and for different false alarm probability settings. Soft fusion is superior to hard fusion techniques in terms of robustness and prediction accuracy. Future work includes *heterogeneous* cooperative scenario, and secondary user clustering techniques with the use of stochastic geometry tool for analysis.

# Chapter 6

# Conclusion and Future Work

In this thesis, the primary focus is the key enabler for cognitive radios which is spectrum prediction that plays an important role in optimizing cognitive cycle and improving the spectral efficiency. In particular, this research is devoted to explore Bayesian based techniques (mainly HMM) analytically, and through Monte-Carlo simulations. Understanding performance bounds of HMM based spectrum prediction allows us to understand the performance limitations, which is one of the key contributions in the thesis. Throughout our research work, case study analysis is presented to evaluate the performance of the proposed approaches. All the work in this research has been peer reviewed and published, or submitted for publication. In summary, the findings of this thesis focuses on two major folds; firstly, single user (local) occupancy prediction where prediction model selection and performance are addressed. Secondly, cooperative spectrum prediction performance of is studied for both decision (hard), as well as data (soft) fusion. The main contributions of our work are presented in Chapter-2 - Chapter-5.

In the first contribution, a survey in current models proposed in literature as well as a consolidated framework based on sequential prediction theory are presented. It is identified that Prediction model selection is not instantly clear in SOP literature based on an in-depth review of sequential prediction. 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. We extended mixture model formulation to cooperative spectrum occupancy prediction using decision (Hard), and data (Soft) fusion techniques. Finally, theoretical and practical challenges of sequential spectrum occupancy prediction implementation are elaborated.

In the second contribution, analytical approximation of HMM based local occupancy prediction as well as using Monte-Carlo simulation techniques are presented and compared. In a case study analysis, the prediction error of one step-ahead (single time slot) prediction is presented against the channel detection errors, as well as primary user's state transition probability. Prediction error is also investigated against the observation sequence length to examine the temporal dependency between prediction accuracy, and the observation length. Consequently, a new recursive equation to estimate HMM prediction performance si proposed as a function of channel detection errors based on HMM posterior probability distribution. Finally, a new generalized Beta-Bernoulli approximation of the predictive posterior probability for local HMM based SOP is presented which provides a tractable expression of prediction performance. The approximation captured HMM prediction error with high accuracy.

The third contribution of our work put forth performance analysis of cooperative hard fusion based spectrum prediction. We further extended the performance analysis of local spectrum prediction using Monte-Carlo simulation techniques to hard fusion cooperative prediction. The contribution presented an analysis of secondary user's mean prediction error in terms of primary user's activity pattern, and spectrum sensing errors. We utilized Bayesian filtering, and known information theory inequalities, to express cooperative prediction error bounds. Finally, a new generalized Beta-Binomial approximation of the predictive posterior probability for cooperative hard fusion based SOP is presented which provides a tractable expression of prediction performance.

Finally, the fourth contribution addressed soft decision based fusion for cooperative SOP. Soft fusion techniques are compared to local spectrum prediction, as well as benchmarked against hard fusion techniques. In particular, soft fusion superiority in terms of robustness as well as prediction accuracy is identified. Accordingly, an alternative soft fusion techniques is proposed based on local prediction model parameters. The alternative techniques conceptually attempt to avoid common control channel requirements, while providing identical performance to known soft fusion techniques.

Remarkably, the case study analysis confirmed that the statistical approximation is able to predict the performance of local and hard fusion cooperative prediction accurately, capturing all the essential aspects of signal detection performance, temporal dependency of primary user activity as well as the finite nature of the network. It is worth noting that the scope of this work was limited to a single PU activity patterns, as well as identical channel conditions for secondary users (homogeneous users). A possible extension would be to a consider heterogeneous SU cooperative prediction to develop novel algorithms or analytical framework for such scenario. Additionally, our analytical approximation for hard fusion error is based on equally likely stationary distribution of PU activity, it will be an interesting future direction to analyse the performance of spectrum prediction under different PU stationary distribution assumptions. Finally, another possible future avenue is to derive an analytical approximation for soft fusion spectrum prediction. The direction would incorporate different prior assumptions i.e. clustering assumption of cooperative spectrum prediction. It will also be interesting to investigate spectrum prediction with reinforcement learning where spectrum access decisions are incorporated in spectrum prediction along with sensing results e.g. Markov decision process.

# References


[1] A. Al-Hourani, V. Trajkovic, S. Chandrasekharan, and S. Kandeepan, "Spectrum occupancy measurements for different urban environments," *2015 European Conference on Networks and Communications (EuCNC)*, 2015.

[2] H. Eltom, S. Kandeepan, B. Moran, and R. J. Evans, "Spectrum occupancy prediction using a hidden Markov model," *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pp. 1–8, dec 2015.

[3] H. Eltom, S. Kandeepan, Y. C. Liang, B. Moran, and R. J. Evans, "HMM based cooperative spectrum occupancy prediction using hard fusion," *2016 IEEE International Conference on Communications Workshops (ICC)*, pp. 669–675, May 2016.

[4] H. Eltom, S. Kandeepan, R. J. Evans, Y. C. Liang, and B. Ristic, "Statistical spectrum occupancy prediction for dynamic spectrum access: a classification," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 29, Feb 2018. [Online]. Available: https://doi.org/10.1186/s13638-017-1019-8

[5] A. Al-Hourani, R. J. Evans, S. Kandeepan, B. Moran, and H. Eltom, "Stochastic geometry methods for modeling automotive radar interference," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–12, 2017.

[6] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[7] C. Baylis, M. Fellows, L. Cohen, and R. J. M. II, "Solving the spectrum crisis: Intelligent, reconfigurable microwave transmitter amplifiers for cognitive radar," *IEEE Microwave Magazine*, vol. 15, no. 5, pp. 94–107, 2014.

[8] Y. C. Liang, K. C. Chen, G. Y. Li, and P. Mahonen, "Cognitive radio networking and communications: An overview," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 3386–3407, 2011.

[9] I. Akyildiz, W.-Y. Lee, M. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 40–48, 2008.

[10] X. Xing, T. Jing, W. Cheng, Y. Huo, and X. Cheng, "Spectrum prediction in cognitive radio networks," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 90–96, 2013.

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

[12] C. Stevenson, G. Chouinard, L. Zhongding, H. Wendong, S. J. Shellhammer, and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 130–138, 2009.

[13] F. Smits, A. Huizing, W. van Rossum, and P. Hiemstra, "A cognitive radar network: Architecture and application to multiplatform radar management," in *European Radar Conference, EuRAD*, 2008, Conference Proceedings, pp. 312–315.

[14] S. Haykin, "Cognitive radar: a way of the future," *IEEE Signal Processing Magazine*, vol. 23, no. 1, pp. 30–40, 2006.

[15] R. J. Evans, P. M. Farrell, G. Felic, H. T. Duong, H. V. Le, J. Li, M. Li, W. Moran, and E. Skafidas, "Consumer radar: Opportunities and challenges," in *2014 $11^{th}$ European Radar Conference*. IEEE, 2014, pp. 5–8.

[16] R. f. p. g. Spectrum planning team, "A review of automotive radar systems-devices and regulation frameworks."

[17] H. Bolfarine and S. Zacks, *Prediction Theory for Finite Populations*, ser. Springer Series in Statistics. New York, USA: Springer, 1992.

[18] Y. Chen and H.-S. Oh, "A survey of measurement-based spectrum occupancy modeling for cognitive radios," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 848–859, 2016.

[19] C. An, H. Ji, P. Si, and X. Mao, "Spectrum sensing sequence prediction in cognitive radio networks," *High Technology Letters*, vol. 17, no. 4, pp. 371–376, 2011.

[20] Z. Chen, N. Guo, Z. Hu, and R. C. Qiu, "Experimental validation of channel state prediction considering delays in practical cognitive radio," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1314–1325, 2011.

[21] H. Li, X. Xu, B. Wu, and X. Chen, "Multilayer feedforward neural network based efficient spectrum occupancy prediction scheme for cognitive radio system," *Journal of Computational Information Systems*, vol. 10, no. 10, pp. 4017–4028, 2014.

[22] J. Lee and H. K. Park, "Channel prediction-based channel aladdress scheme for multichannel cognitive radio networks," *Journal of Communications and Networks*, vol. 16, no. 2, pp. 209–216, 2014.

[23] W. Pu and I. F. Akyildiz, "Asymptotic queuing analysis for dynamic spectrum access networks in the presence of heavy tails," *IEEE journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 514–522, 2013.

[24] S. J. Kim and G. B. Giannakis, "Dynamic learning for cognitive radio sensing," in *2013 5$^{th}$ IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2013, pp. 388–391.

[25] X. Li and S. A. Zekavat, "Cognitive radio based spectrum sharing: Evaluating channel availability via traffic pattern prediction," *Journal of Communications and Networks*, vol. 11, no. 2, pp. 104–114, 2009.

[26] V. K. Tumuluru, P. Wang, and D. Niyato, "Channel status prediction for cognitive radio networks," *Wireless Communications & Mobile Computing*, vol. 12, no. 10, pp. 862–874, 2012.

[27] S. Chen and L. Tong, "Maximum throughput region of multiuser cognitive access of continuous time Markovian channels," *IEEE journal on Selected Areas in Communications*, vol. 29, no. 10, pp. 1959–1969, 2011.

[28] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1136–1159, 2013.

[29] A. He, K. K. Bae, T. R. Newman, J. Gaeddert, K. Kim, R. Menon, L. Morales-Tirado, J. J. Neel, Y. Zhao, J. H. Reed, and W. H. Tranter, "A survey of artificial intelligence for cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1578–1592, 2010.

[30] N. Cesa-Bianchi, G. Lugosi, and L. Prediction, "Learning,prediction, and games."

[31] M. Hutter, "Convergence and loss bounds for bayesian sequence prediction," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 2061–2067, 2003.

[32] P. Algoet, "The strong law of large numbers for sequential decisions under uncertainty," *IEEE Transactions on Information Theory*, vol. 40, no. 3, pp. 609–633, 1994.

[33] N. Merhav and M. Feder, "A strong version of the redundancy-capacity theorem of universal coding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 714–722, may 1995.

[34] W. Kotłowski and P. Grünwald, "Sequential normalized maximum likelihood in log-loss prediction," in *IEEE Information Theory Workshop (ITW), 2012*, 2012, pp. 547–551.

[35] N. Merhav and M. Feder, "Universal prediction," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2124–2147, 1998.

[36] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Transactions on Information Theory*, vol. 30, no. 4, pp. 629–636, 1984.

[37] A. Goldsmith and P. Varaiya, “Capacity, mutual information, and coding for finite-state Markov channels,” *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 868–886, 1996.

[38] J. A. T. Thomas M. Cover, *Elements of Information Theory*. New York, USA: Wiley John + Sons, sep 2006.

[39] R. D. Smallwood and E. J. Sondik, “The optimal control of partially observable Markov processes over a finite horizon,” *Operations Research*, vol. 21, no. 5, pp. 1071–1088, 1973.

[40] D. Blackwell, “The entropy of functions of finite-state Markov chains,” in *Transactions of the first Prague conference on information theory, Statistical decision functions, random processes held at Liblice near Prague from November*, vol. 28, 1957, pp. 13–20.

[41] T. Kaijser, “A limit theorem for partially observed Markov chains,” *The Annals of Probability*, vol. 3, no. 4, pp. 677–696, 1975.

[42] J. Marroquin, S. Mitter, and T. Poggio, “Probabilistic solution of ill-posed problems in computational vision,” *Journal of the American statistical association*, vol. 82, no. 397, pp. 76–89, 1987.

[43] X. Xing, T. Jing, W. Cheng, Y. Huo, X. Cheng, and T. Znati, “Cooperative spectrum prediction in multi-PU multi-SU cognitive radio networks,” *Mobile Networks and Applications*, vol. 19, no. 4, pp. 502–511, 2014.

[44] C. Xianfu, Z. Honggang, A. B. Mackenzie, and M. Matinmikko, “Predicting spectrum occupancies using a non-stationary hidden Markov model,” *IEEE Wireless Communications Letters*, vol. 3, no. 4, pp. 333–336, 2014.

[45] K. Sithamparanathan and A. Giorgetti, *Cognitive Radio Techniques: Spectrum Sensing, Interference Mitigation, and Localization*, ser. Artech House mobile communications library. Boston: Artech House, 2012.

[46] S. S. Dias and M. G. S. Bruno, “Cooperative target tracking using decentralized particle filtering and RSS sensors,” *IEEE Transactions on Signal Processing*, vol. 61, no. 14, pp. 3632–3646, 2013.

[47] L. R. Welch, “Hidden Markov models and the baum-welch algorithm,” *IEEE Information Theory Society Newsletter*, vol. 53, no. 4, pp. 10–13, 2003.

[48] H. Kobayashi, B. L. Mark, and W. Turin, *Probability, Random Processes, and Statistical Analysis: Applications to Communications, Signal Processing, Queueing Theory and Mathematical Finance*. New York, NY, USA: Cambridge University Press, 2011.

[49] B. P. Carlin and T. A. Louis, *Bayes and empirical Bayes methods for data analysis*. Chapman & Hall/CRC Boca Raton, FL, 2000, vol. 17.

[50] R. C. Tripathi, R. C. Gupta, and J. Gurland, "Estimation of parameters in the beta binomial model," *Annals of the Institute of Statistical Mathematics*, vol. 46, no. 2, pp. 317–331, 1994.

[51] S. Atapattu, C. Tellambura, and H. Jiang, *Energy Detection for Spectrum Sensing in Cognitive Radio*. Springer Publishing Company, Incorporated, 2014.

[52] L. R. Rabiner, "Readings in speech recognition," A. Waibel and K.-F. Lee, Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990, ch. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, pp. 267–296. [Online]. Available: http://dl.acm.org/citation.cfm?id=108235.108253

[53] T. Holliday, A. Goldsmith, and P. Glynn, "Capacity of finite state channels based on lyapunov exponents of random matrices," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3509–3532, 2006.

[54] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. New York, NY, USA: Cambridge University Press, 2006.

[55] C. E. Shannon, "Prediction and entropy of printed english," *Bell System Technical journal*, vol. 30, no. 1, pp. 50–64, 1951.

[56] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 337–343, 1977.

[57] J. L. KELLY, "A new interpretation of information rate," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 25–34, 2011.

[58] G. Shafer and V. Vovk, *Probability and Finance: It's Only a Game!*, ser. Wiley Series in Probability and Statistics. New York, USA: Wiley, 2005.

[59] P. D. Grnwald, I. J. Myung, and M. A. Pitt, *Advances in Minimum Description Length: Theory and Applications (Neural Information Processing)*. Cambridge, MA: The MIT Press, 2005.

[60] N. N. Cencov, *Statistical Decision Rules and Optimal Inference (Translations of Mathematical Monographs)*. Providence, RI USA: American Mathematical Society, 2000, no. 53.

[61] P. P. Vaidyanathan, "The theory of linear prediction," *Synthesis Lectures on Signal Processing*, vol. 2, no. 1, pp. 1–184, 2007.

[62] J. Rissanen, "Strong optimality of the normalized ML models as universal codes and information in data," *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 1712–1717, 2001.

[63] E. A. Wan and R. V. D. Merwe, "The unscented kalman filter for nonlinear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, 2000, pp. 153–158.

[64] B. Ristic, S. Arulampalam, and N. J. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* London,UK: Artech house, 2004.

[65] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 249–265, 2000.

[66] M. Dudík, S. J. Phillips, and R. E. Schapire, "Performance guarantees for regularized maximum entropy density estimation," in *Learning Theory.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 472–486.

[67] Y. W. Teh, *Dirichlet Process.* Boston, MA: Springer US, 2010, pp. 280–287.

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

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

[70] 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," *2008 3$^{rd}$ International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom 2008)*, pp. 1–7, 2008.

[71] W. Tang, J. Zhou, H. Yu, and S. Li, "A fair scheduling scheme based on collision statistics for cognitive radio networks," *Concurrency and Computation: Practice and Experience*, vol. 25, no. 9, pp. 1091–1100, 2012.

[72] C. Xu and H. Jianwei, "Evolutionarily stable spectrum access," *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, pp. 1281–1293, 2013.

[73] P. De and Y.-C. Liang, "Blind spectrum sensing algorithms for cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 5, pp. 2834–2842, 2008.

[74] P. Huang, C.-J. Liu, L. Xiao, and J. Chen, "Wireless spectrum occupancy prediction based on partial periodic pattern mining," *2012 IEEE 20$^{th}$ International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, vol. 25, no. 7, pp. 1925–1934, 2012.

[75] I. A. Akbar and W. H. Tranter, "Dynamic spectrum aladdress in cognitive radio using hidden Markov models: Poisson distributed case," in *Proceedings of 2007 IEEE SoutheastCon*, 2007, pp. 196–201.

[76] S. D. Barnes and B. T. Maharaj, "Prediction based channel aladdress performance for cognitive radio," *AEU - International journal of Electronics and Communications*, vol. 68, no. 4, pp. 336–345, 2014.

[77] Z. Chen, N. Guo, Z. Hu, and R. C. Qiu, "Channel state prediction in cognitive radio, part II: Single-user prediction," in *Proceedings of IEEE Southeastcon*. IEEE, 2011, pp. 50–54.

[78] D. C. Karia, B. Lande, and R. Daruwala, "Performance analysis of HMM- and ANN-based spectrum vacancy predictor behaviour for cognitive radios," *International Journal of Ad Hoc and Ubiquitous Computing*, vol. 11, no. 4, pp. 206–213, 2012.

[79] L. Melián-Gutiérrez, S. Zazo, J. L. Blanco-Murillo, I. Pérez-Álvarez, A. García-Rodríguez, and B. Pérez-Díaz, "HF spectrum activity prediction model based on HMM for cognitive radio applications," *Physical Communication*, vol. 9, pp. 199–211, 2013.

[80] C. H. Park, S. W. Kim, S. M. Lim, and M. S. Song, "HMM based channel status predictor for cognitive radio," *2007 Asia Pacific Microwave Conference, Vols 1-5*, pp. 675–678, 2007.

[81] S. H. Sohn, "HMM-based adaptive frequency-hopping cognitive radio system to reduce interference time and to improve throughput," *KSII Transactions on Internet and Information Systems*, vol. 4, no. 4, pp. 475–490, 2010.

[82] T. Nguyen, B. L. Mark, and Y. Ephraim, "Spectrum sensing using a hidden bivariate Markov model," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4582–4591, 2013.

[83] C. Xianfu, Z. Zhifeng, and Z. Honggang, "Stochastic power adaptation with multiagent reinforcement learning for cognitive wireless mesh networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 11, pp. 2155–2166, 2013.

[84] K. W. Sung, S.-L. Kim, and J. Zander, "Temporal spectrum sharing based on primary user activity prediction," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3848–3855, 2010.

[85] B. G. Najashi and W. Feng, "Cooperative spectrum occupancy based spectrum prediction modeling," *Journal of Computational Information Systems*, vol. 10, no. 10, pp. 4093–4100, 2014.

[86] Y. Wang, Z. Zhang, L. Ma, and J. Chen, "SVM-based spectrum mobility prediction scheme in mobile cognitive radio networks," *The Scientific World journal*, vol. 2014, pp. 1–11, 2014.

[87] H. Pei, L. Chin-Jung, Y. Xi, X. Li, and C. Jin, "Wireless spectrum occupancy prediction based on partial periodic pattern mining," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 7, pp. 1925–1934, 2014.

[88] S. Arunthavanathan, S. Kandeepan, and R. J. Evans, "Reinforcement learning based secondary user transmissions in cognitive radio networks," *2013 IEEE Globecom Workshops (GC)*, pp. 374–379, 2013.

[89] A. Banaei and C. N. Georghiades, “Throughput analysis of a randomized sensing scheme in cell-based ad-hoc cognitive networks,” *2009 IEEE International Conference on Communications*, pp. 1–6, 2009.

[90] J. Gambini, O. Simeone, U. Spagnolini, Y. Bar-Ness, and Y. Kim, “Cognitive radio with secondary packet-by-packet vertical handover,” *2008 IEEE International Conference on Communications*, pp. 1050–1054, 2008.

[91] M. Derakhshani and T. Le-Ngoc, “Learning-based opportunistic spectrum access with adaptive hopping transmission strategy,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3957–3967, 2012.

[92] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, “Primary user behavior in cellular networks and implications for dynamic spectrum access,” *IEEE Communications Magazine*, vol. 47, no. 3, pp. 88–95, 2009.

[93] J. Zhang, G. Ding, Y. Xu, and F. Song, “On the usefulness of spectrum prediction for dynamic spectrum access,” in *IEEE $8^{th}$ International Conference on Wireless Communications & Signal Processing (WCSP)*, 2016, pp. 1–4.

[94] J. Yang and H. Zhao, “Enhanced throughput of cognitive radio networks by imperfect spectrum prediction,” *IEEE Communications Letters*, vol. 19, no. 10, pp. 1738–1741, 2015.

[95] M. H. Rehmani, A. C. Viana, H. Khalife, and S. Fdida, “SURF: A distributed channel selection strategy for data dissemination in multi-hop cognitive radio networks,” *Computer Communications*, vol. 36, no. 10, pp. 1172–1185, 2013.

[96] S. Bayhan and F. Alagöz, “Distributed channel selection in CRAHNs: A non-selfish scheme for mitigating spectrum fragmentation,” *Ad Hoc Networks*, vol. 10, no. 5, pp. 774–788, 2012, special Issue on Cognitive Radio Ad Hoc Networks.

[97] M. Khabazian, S. Aissa, and N. Tadayon, “Performance modeling of a two-tier primary-secondary network operated with IEEE 802.11 DCF mechanism,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 9, pp. 3047–3057, 2012.

[98] Z. Wang and S. Salous, “Spectrum occupancy statistics and time series models for cognitive radio,” *Journal of Signal Processing Systems*, vol. 62, no. 2, pp. 145–155, 2011.

[99] P. Thakur, A. Kumar, S. Pandit, G. Singh, and S. N. Satashia, “Performance improvement of cognitive radio network using spectrum prediction and monitoring techniques for spectrum mobility,” in *IEEE Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2016, pp. 679–684.

[100] F. Hou, X. Chen, H. Huang, and X. Jing, “Throughput performance improvement in cognitive radio networks based on spectrum prediction,” *2016 $16^{th}$ International Symposium on Communications and Information Technologies (ISCIT)*, pp. 655–658, 2016.

[101] S. Joshi, P. Pawelczak, D. Cabric, and J. Villasenor, “When channel bonding is beneficial for opportunistic spectrum access networks,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3942–3956, 2012.

[102] W. Wang, T. Lv, T. Wang, and X. Yu, “Primary user activity based channel aladdress in cognitive radio networks,” in *IEEE 72$^{nd}$ Conference on Vehicular Technology (VTC 2010-Fall), 2010.* IEEE, 2010, pp. 1–5.

[103] L. Csurgai-Horváth and J. Bito, “Primary and secondary user activity models for cognitive wireless network,” in *Proceedings of the 2011 11$^{th}$ International Conference on Telecommunications (ConTEL)*, pp. 301–306.

[104] S. Bayhan and F. Alagöz, “A Markovian approach for best-fit channel selection in cognitive radio networks,” *Ad Hoc Networks*, vol. 12, pp. 165–177, 2014.

[105] A. W. Min and K. G. Shin, “Exploiting multi-channel diversity in spectrum-agile networks,” *IEEE Conference on Computer Communications*, 2008.

[106] Q. Zhao, L. Tong, A. Swami, and Y. Chen, “Decentralized cognitive mac for opportunistic spectrum access in ad-hoc networks: A pomdp framework,” *IEEE journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, 2007.

[107] Y. Li, Y.-n. Dong, H. Zhang, H.-t. Zhao, H.-x. Shi, and X.-x. Zhao, “Spectrum usage prediction based on high-order Markov model for cognitive radio networks,” in *IEEE 10$^{th}$ International Conference on Computer and Information Technology (CIT)*, 2010, pp. 2784–2788.

[108] J. Riihijärvi, J. Nasreddine, and P. Mähönen, “Impact of primary user activity patterns on spatial spectrum reuse opportunities,” in *European Wireless Conference (EW)*, 2010, pp. 962–968.

[109] M. Wellens, J. Riihijarvi, and P. Mahonen, “Modelling primary system activity in dynamic spectrum access networks by aggregated on/off-processes,” in *IEEE Annual Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops*, jun 2009, pp. 1–6.

[110] S. Wang, J. Zhang, and L. Tong, “A characterization of delay performance of cognitive medium access,” *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 800–809, 2012.

[111] L. Jiao, E. Song, V. Pla, and F. Y. Li, “Capacity upper bound of channel assembling in cognitive radio networks with quasistationary primary user activities,” *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1849–1855, 2013.

[112] C.-H. Liu and D. Cabric, “Prediction of Erlang-2 distributed primary user traffic for dynamic spectrum access,” *IEEE Wireless Communications Letters*, vol. 4, no. 5, pp. 481–484, 2015.

[113] A. Saad, B. Staehle, and R. Knorr, “Spectrum prediction using hidden Markov models for industrial cognitive radio,” in *IEEE 12$^{th}$ International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2016, pp. 1–7.

[114] Y. Zhao, Z. Hong, G. Wang, and J. Huang, "High-order hidden bivariate Markov model: A novel approach on spectrum prediction," in *IEEE 25$^{th}$ International Conference on Computer Communication and Networks (ICCCN)*, 2016, pp. 1–7.

[115] W. Saad, Z. Han, H. V. Poor, T. Basar, and J. B. Song, "A cooperative bayesian nonparametric framework for primary user activity monitoring in cognitive radio networks," *IEEE journal on Selected Areas in Communications*, vol. 30, no. 9, pp. 1815–1822, 2012.

[116] H. Li and R. C. Qiu, "A graphical framework for spectrum modeling and decision making in cognitive radio networks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2010, pp. 1–6.

[117] D. Dash and A. Sabharwal, "Paranoid secondary: Waterfilling in a cognitive interference channel with partial knowledge," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 1045–1055, 2012.

[118] B. Canberk, I. F. Akyildiz, and S. Oktug, "Primary user activity modeling using first-difference filter clustering and correlation in cognitive radio networks," *IEEE/ACM Transactions on Networking*, vol. 19, no. 1, pp. 170–183, 2011.

[119] Z. Wen, T. Luo, W. Xiang, S. Majhi, and Y. Ma, "Autoregressive spectrum hole prediction model for cognitive radio systems," *IEEE International Conference on Communications*, pp. 154–157, 2008.

[120] A. Eltholth, "Forward Backward autoregressive spectrum prediction scheme in cognitive radio systems," in *2015 9$^{th}$ International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2015, pp. 1–5.

[121] S.-S. Gu and S.-N. Yu, "A chaotic neural network-based algorithm for relational structure matching," *IEEE 2004 International Conference on Machine Learning and Cybernetics*, vol. 6, pp. 3328–3333, aug 2004.

[122] J. Yang, H. Zhao, and X. Chen, "Genetic algorithm optimized training for neural network spectrum prediction," in *IEEE 2$^{nd}$ International Conference on Computer and Communications (ICCC)*, 2016, pp. 2949–2954.

[123] S. Ni, X. Bai, Z. Wang, and B. Guo, "A new method of cognitive radio spectrum prediction research," in *IEEE International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2016, pp. 982–986.

[124] A. Agarwal, S. Dubey, M. A. Khan, R. Gangopadhyay, and S. Debnath, "Learning based primary user activity prediction in cognitive radio networks for efficient dynamic spectrum access," in *2016 International Conference on Signal Processing and Communications (SPCOM)*, 2016, pp. 1–5.

[125] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, ser. Athena Scientific books. Belmont, MA, USA: Athena Scientific, 2002.

[126] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[127] D. Haussler, A. Barron, and S. C. C. R. L. University of California, *How well do Bayes methods work for on-line prediction of [[+ or - ]1] values?*, ser. Technical reports. California, USA: University of California, Santa Cruz, Computer Research Laboratory, 1992.

[128] D. Barber, *Bayesian Time Series Models*. New York, NY: Cambridge University Press, 2011.

[129] C. Clancy, J. Hecker, E. Stuntebeck, and T. O'Shea, "Applications of machine learning to cognitive radio networks," *IEEE Wireless Communications*, vol. 14, no. 4, pp. 47–52, 2007.

[130] L. Gavrilovska, V. Atanasovski, I. Macaluso, and L. A. DaSilva, "Learning and reasoning in cognitive radio networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1761–1777, 2013.

[131] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York, NY: Springer, 1993.

[132] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *journal of artificial intelligence research*, pp. 237–285.

[133] T. Jiang, P. Mitchell, and D. Grace, "Efficient exploration in reinforcement learning-based cognitive radio spectrum sharing," *IET Communications*, vol. 5, no. 10, pp. 1309–1317, 2011.

[134] F. Fu and M. van der Schaar, "Learning to compete for resources in wireless stochastic games," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1904–1919, 2009.

[135] S. K. Jayaweera, M. Bkassiny, and K. A. Avery, "Asymmetric cooperative communications based spectrum leasing via auctions in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2716–2724, 2011.

[136] S. Filippi, O. Cappe, and A. Garivier, "Optimally sensing a single channel without prior information: The tiling algorithm and regret bounds," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 68–76, 2011.

[137] J. Lunden, S. R. Kulkarni, V. Koivunen, and H. V. Poor, "Multiagent reinforcement learning based spectrum sensing policies for cognitive radio networks," *IEEE journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 858–868, 2013.

[138] T. Jiang, D. Grace, and Y. Liu, "Two-stage reinforcement-learning-based cognitive radio with exploration control," *IET Communications*, vol. 5, no. 5, pp. 644–651, 2011.

[139] Z. Qu, R. Cui, Q. Song, and S. Yin, "Predictive spectrum sensing strategy based on reinforcement learning," *Communications, China*, vol. 11, no. 10, pp. 117–125, 2014.

[140] W. Yeqing, H. Fei, S. Kumar, Z. Yingying, A. Talari, N. Rahnavard, and J. D. Matyjas, "A learning-based qoe-driven spectrum handoff scheme for multimedia transmissions over cognitive radio networks," *IEEE journal on Selected Areas in Communications*, vol. 32, no. 11, pp. 2134–2148, 2014.

[141] Y. SiXing, C. Dawei, Z. Qian, and L. ShuFang, "Prediction-based throughput optimization for dynamic spectrum access," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 3, pp. 1284–1289, 2011.

[142] M. Lopez-Benitez and F. Casadevall, "Time-dimension models of spectrum usage for the analysis, design, and simulation of cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2091–2104, 2013.

[143] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Probability & Its Applications*, vol. 14, no. 1, pp. 153–158, 1969.

[144] I. Macaluso, D. Finn, B. Ozgul, and L. A. DaSilva, "Complexity of spectrum activity and benefits of reinforcement learning for dynamic channel selection," *IEEE journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2237–2248, 2013.

[145] M. Mushkin and I. Bar-David, "Capacity and coding for the gilbert-elliott channels," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 1277–1290, 1989.

[146] Y. Ephraim and N. Merhav, "Hidden Markov processes," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1518–1569, 2002.

[147] M. Hutter, "On universal prediction and bayesian confirmation," *Theoretical Computer Science*, vol. 384, no. 1, pp. 33–48, 2007.

[148] A. Kavcic, "On the capacity of Markov sources over noisy channels," in *Global Telecommunications Conference*, vol. 5, 2001.

[149] J.-P. Shen and J. Gill, "Analysis on a hidden Markov channel model," in *Global Telecommunications Conference.*, vol. 1A. IEEE, 1999, pp. 437–441.

[150] M. Rezaeian, "Computation of capacity for gilbert-elliott channels, using a statistical method," in *2005 Australian Communications Theory Workshop*, 2005, pp. 56–61.

[151] R. Ganti and M. Haenggi, "Spatial and temporal correlation of the interference in ALOHA ad hoc networks," *IEEE Communications Letters*, vol. 13, no. 9, pp. 631–633, 2009.

[152] T. Jing, X. Xing, W. Cheng, Y. Huo, and T. Znati, "Cooperative spectrum prediction in multi-PU multi-SU cognitive radio networks," in *8th International Conference on Cognitive Radio Oriented Wireless Networks.* IEEE, jul 2013, pp. 25–30.