GENERALISED BAYESIAN MODEL SELECTION USING REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in Electronics and Communication Engineering

by
Oktay KARAKUŞ

November 2017
İZMİR

We approve the thesis of **Oktay KARAKUŞ**

Examining Committee Members:

_____

Assoc. Prof. Dr. Mustafa A. ALTINKAYA
Department of Electrical-Electronics Engineering, İzmir Institute of Technology

_____

Prof. Dr. Bilge KARAÇALI
Department of Electrical-Electronics Engineering, İzmir Institute of Technology

_____

Assoc. Prof. Dr. Olcay AKAY
Department of Electrical-Electronics Engineering, Dokuz Eylül University

_____

Prof. Dr. Aydın AKAN
Department of Biomedical Engineering, Katip Çelebi University

_____

Assoc. Prof. Dr. Ali Taylan CEMGİL
Department of Computer Engineering, Boğaziçi University

24 November 2017

_____   _____

Assoc. Prof. Dr. Mustafa A. ALTINKAYA   **Assoc. Prof. Dr. Ercan E. KURUOĞLU**
Supervisor   Co-Supervisor
Department of Electrical-Electronics Engineering   ISTI-CNR
İzmir Institute of Technology   Pisa, Italy

_____   _____

Prof. Dr. Enver TATLICIOĞLU   **Prof. Dr. Aysun SOFUOĞLU**
Head of the Department of   Dean of the Graduate School of
Electrical-Electronics Engineering   Engineering and Sciences

ACKNOWLEDGMENTS

ABSTRACT

GENERALISED BAYESIAN MODEL SELECTION USING REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO

The main objective of this thesis is to suggest a general Bayesian framework for model selection based on *reversible jump Markov chain Monte Carlo* (RJMCMC) algorithm. In particular, we aim to reveal the undiscovered potentials of RJMCMC in model selection applications by exploiting the original formulation to explore spaces of different classes or structures and thus, to show that RJMCMC offers a wider interpretation than just being a trans-dimensional model selection algorithm.

The general practice is to use RJMCMC in a trans-dimensional framework e.g. in model estimation studies of linear time series, such as AR and ARMA and mixture processes, etc. In this thesis, we propose a new interpretation on RJMCMC which reveals the undiscovered potentials of the algorithm. This new interpretation, firstly, extends the classical trans-dimensional approach to a much wider meaning by exploring the spaces of linear and nonlinear models in terms of the nonlinear (polynomial) time series models. Polynomial process modelling is followed by the definition of a new type of RJM-CMC move that performs transitions between various generic model spaces irrespective of model sizes. Then, we apply this new framework to the identification of Volterra systems with an application of nonlinear channel estimation of an OFDM communication system. The proposed RJMCMC move has been adjusted to explore the spaces of different distribution families by matching the common properties of the model spaces such as norm, and this leads us to perform a distribution estimation study of the observed real-life data sets including, impulsive noise in power-line communications, seismic acceleration time series, remote sensing images, etc.

Simulation results demonstrate the remarkable performance of the proposed method in nonlinearity degree estimation and in transitions between different classes of models. The proposed method uses RJMCMC in an unorthodox way and reveals its potential to be a general estimation method by performing the reversible jump mechanism between spaces of different model classes.

ÖZET

TERSİNE ATLAMALI MARKOV ZİNCİRİ MONTE CARLO KULLANARAK GENELLEŞTİRİLMİŞ BAYESÇİ MODEL SEÇİMİ

Bu tezin temel amacı, *tersine atlamalı Markov zinciri Monte Carlo* (RJMCMC) algoritmasına dayanan model seçimi için genel Bayesçi bir çerçeve önermektir. Özellikle, farklı sınıfların veya yapıların uzaylarını keşfetmek şeklindeki orijinal formülasyonundan istifade ederek RJMCMC'nin keşfedilmemiş potansiyellerini model seçim uygulamalarında ortaya koymayı ve böylece RJMCMC'nin sadece boyutlar arası bir model seçim algoritması olmaktan daha geniş bir yorum sunduğunu göstermeyi amaçlıyoruz.

Genel uygulama, RJMCMC'yi boyutlar arası çerçevede, örneğin; AR ve ARMA ve karışım süreçleri gibi doğrusal zaman serilerinin model kestirimi çalışmalarında kullanmak yönündedir. Biz bu tezde, RJMCMC üzerine, algoritmanın keşfedilmemiş potansiyelini ortaya koyan yeni bir yorum öneriyoruz. Bu yeni yorum, önce doğrusal ve doğrusal olmayan modellerin uzaylarını, doğrusal olmayan (polinom) zaman serisi modelleri ile araştırarak, klasik boyutlar arası yaklaşımdan çok daha derin bir anlama genişletmektedir. Polinom süreç modelleme çalışmasını takiben, model boyutlarına bakılmaksızın çeşitli genel model uzayları arasında geçişler gerçekleştiren yeni tip bir RJMCMC geçişinin tanımı yapılmıştır. Ardından, bu yeni yaklaşımı bir OFDM iletişim sisteminin doğrusal olmayan kanal kestirimi uygulaması ile Volterra sistemlerinin tanılanmasında kullanılmıştır. Önerilen RJMCMC geçişi, model uzaylarının ortak özelliklerini (norm gibi) eşleştirerek farklı dağılım ailelerinin uzaylarını keşfetmek üzere ayarlanmış ve bu, bize, gözlemlenen gerçek hayat veri setleri için, örneğin; güç hattı iletişim sistemlerindeki dürtüsel gürültü, sismik ivme zaman serileri, uzaktan algılama imgeleri vb., bir dağılım kestirimi çalışması yapmaya yöneltmiştir.

Benzetim sonuçları, doğrusal olmayan derece tahmininde ve farklı model sınıfları arasındaki geçişlerde önerilen yöntemin olağanüstü performansını ortaya koymaktadır. Önerilen yöntem, RJMCMC'yi alışılagelmemiş bir şekilde kullanmakta ve farklı model sınıflarının uzayları arasında tersine atlama mekanizmasını gerçekleştirerek genel bir kestirim yöntemi olma potansiyelini ortaya koymaktadır.

Kızıma
ve
Eşime

To my daughter Yağmur,
for giving me strength, patience
and countless reasons to smile
when the life is tough.

To my beloved wife Duygu,
for her never-ending love
and understanding even when
I am such a tough man.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Motivation

Model Selection can be defined as the task of selecting the best (or the most suitable) statistical model from a set of candidate models, given data and over the years, several approaches have been proposed for this purpose. Approaches such as Akaike Information Criterion (AIC) , Bayesian Information Criterion (BIC) , Mallow's C_p, Bayes factors, etc. use a decision criterion by evaluating different properties of the models in the model space (Kadane and Lazar, 2004). A problem arises when the number of candidate models in the model space is high. Thus, it is difficult to decide which model is the best or which criterion you are going to deal with to select the most useful model for the associated problem.

Approaches stated above can be classified as either frequentist or Bayesian. Frequentist approaches use data dependent procedures and in decision procedure, never take the prior and the probability of the hypothesis into account. On the other hand, if the selection criterion is based on prior information, the model selection is named to be *Bayesian Model Selection*. Suppose we have a set of k models such as $\mathcal{M} = \{M_1, \ldots, M_k\}$ all of which have parameters $\Theta = \{\theta_1, \ldots, \theta_k\}$ under data Y. Data Y has density $f(Y|\theta_k, M_k)$ where θ_k is a vector of unknown variables for the model M_k. The Bayesian formula for the posterior density of model M_k given data Y can be written as (Chipman et al., 2001)

$$f(M_k|Y) = \frac{f(Y|M_k)f(M_k)}{\sum_k f(Y|M_k)f(M_k)} \tag{1.1}$$

where the posterior model probability $f(M_k|Y)$ is obtained by integrating out $f(M_k, \theta_k|Y)$ over Θ. Additionally, $f(M_k)$ is the model prior probability and the marginal likelihood

$f(Y|M_k)$ can be obtained as

$$f(Y|M_k) = \int f(Y|\theta_k, M_k) f(\theta_k|M_k) d\theta_k. \qquad (1.2)$$

Modelling posterior distributions $f(M_1|Y), \ldots, f(M_k|Y)$ is the main object of interest for the Bayesian model selection. By treating $f(M_k|Y)$ as a measure of the truth of model M_k, a natural and simple strategy for model selection is to choose the most probable M_k, the one for which $f(M_k|Y)$ is the largest. This strategy is named as the maximum posterior estimation (Chipman et al., 2001).

Bayesian methodology (1.1) involves calculation of summations (or integrations) such as the normalization term, $\sum_k f(Y|M_k) f(M_k)$, the denominator of (1.1), which may be analytically intractable. Numerical and Monte Carlo methods provide a solution to this problem. Monte Carlo simulation is a method that draws n i.i.d samples from a target distribution $f(\cdot)$. These n samples are used to approximate these integrals or sums that converge to the real value when n tends to infinity. When the target distribution does not have a standard form, it is hard to apply direct sampling or inverse sampling methods and more sophisticated sampling techniques are needed such as; rejection sampling, importance sampling and Markov Chain Monte Carlo (MCMC) (Andrieu et al., 2003).

MCMC uses certain properties of Markov chain theory such as irreducibility and aperiodicity that ensures convergence to a stationary distribution to construct the correct posterior using both prior information and the information from observations. There are three main methods that implement the properties of MCMC algorithm: The Metropolis (Metropolis et al., 1953) algorithm, The Metropolis-Hastings (MH) algorithm (Hastings, 1970) and The Gibbs Sampler (Geman and Geman, 1984).

Reversible Jump Markov Chain Monte Carlo (RJMCMC) is first introduced by (Green, 1995) as a model identification tool. RJMCMC can also be defined as a generalization and an extended version of the classical MH method which allows the sampler to jump between spaces of different dimensionality. In particular, in cases where the parameter dimension is not known, it provides means to sample and hence estimate the dimension of the parameter. This is achieved by doing Markov chain sampling in spaces of varying dimensions. RJMCMC achieves this by treating the dimension as a new parameter in the model parameter vector.

In Figure 1.1, a comparison between MCMC and RJMCMC is depicted. In MCMC, there is only a single parameter space θ with dimension k, whereas RJMCMC performs a search in multiple parameter subspaces, θ_t, each of which has dimension k_t.
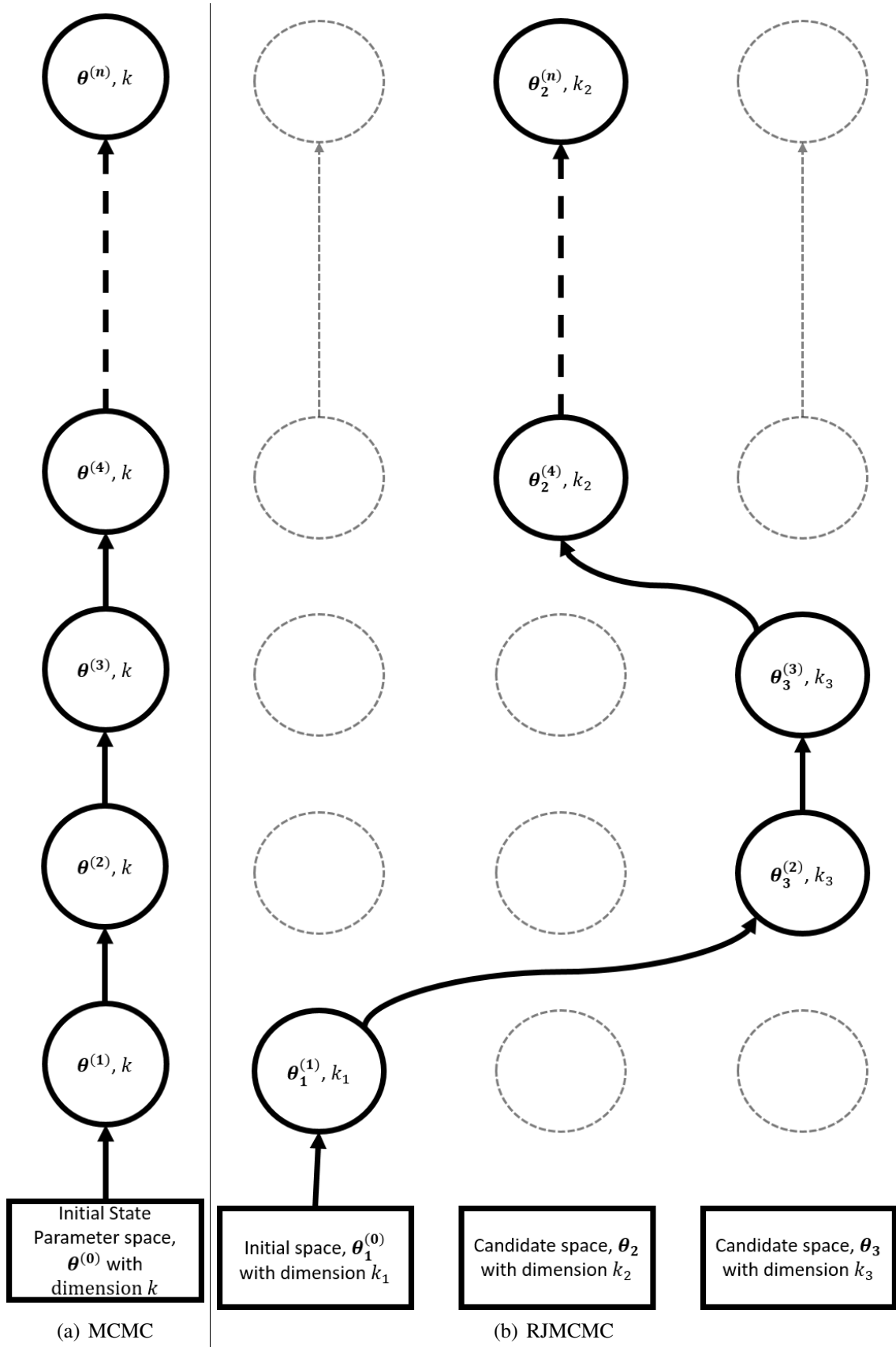
Figure 1.1. Illustration of (a) MCMC and (b) RJMCMC.

As stated above, RJMCMC is able to change the dimension of the state space in a move. This move is called a reversible jump. As it can be understood from the name of the algorithm, any chosen move which is able to change the dimension should be reversible so that the algorithm can switch back to the recent space with a later move (Smith, 2007). Any move can be defined in RJMCMC as far as it is reversible. This gives the algorithm a powerful flexibility and makes the acceptance ratio analytically tractable.

One key advantage of RJMCMC is to determine the "best" model by searching the space through an intelligent random walk that avoids doing an exhaustive search and hence it achieves computational gain compared to other classical methods performing an exhaustive search. At this point, one might argue that training separate MCMC samplers for each of the candidate model spaces and comparing them afterwards would be computationally more advantageous. However, it can clearly be stated that using a single Markov chain with RJMCMC could be simpler when there are unknown (or dramatically large) number of candidate models. Furthermore, in cases where the number of models is small, by efficiently choosing the proposal distributions, one can provide better mixing between candidate model spaces (i.e. faster convergence to the stationary distribution) and incorporating the reversible jump mechanism would be advantageous.

In the literature, the general practice is to use RJMCMC in linear model estimation problems. In (Troughton and Godsill, 1998), model uncertainty problem for autoregressive (AR) models has been studied by exploring the spaces for different AR orders. They have developed their approach by proposing two types of proposals which are partially and fully conditional. Apart from AR models, RJMCMC has also been used in autoregressive integrated moving average (ARIMA) models (Ehlers and Brooks, 2004) and in fractional ARIMA (f-ARIMA) models (Eğri et al., 2010).

In addition to linear model estimation studies, RJMCMC has also been applied to the problems of estimating the nonlinear threshold MA (TMA) models (Xia et al., 2010) and in threshold autoregressive moving average (TARMA) models (Liang et al., 2017). Furthermore, (Troughton and Godsill, 2001) employed RJMCMC in restoring nonlinearly distorted AR signals. For financial time series models, a threshold nonlinearity test in terms of GARCH and threshold GARCH (TGARCH) models have been discussed in (So et al., 2005). RJMCMC has also been used to track a variable number of targets with particle filters which are known to be successful in handling complicated and nonlinear measurement models (Khan et al., 2005).

Another popular application of RJMCMC is to analyze mixtures of distributions. In (Richardson and Green, 1997), RJMCMC has been employed to estimate the number

of Gaussians in a mixture process, where (Viallefont et al., 2002; Salas-Gonzalez et al., 2010) have performed similar estimations for the cases of Poisson and symmetric α-stable distributions, respectively.

Apart from the popular applications above, RJMCMC has been used in other various applications such as detection of clusters in disease maps (Knorr-Held and Raßer, 2000), graphical models based variable selection and automatic curve fitting (Lunn et al., 2009), log-linear model selection (Dellaportas and Forster, 1999), non-parametric drift estimation (Van Der Meulen et al., 2014), delimiting species using multilocus sequence data (Rannala and Yang, 2013), random effect models (Oedekoven et al., 2016), and generation of lane-accurate road network maps from vehicle trajectory data (Roeth et al., 2017).

Despite having various application areas, all these studies have used RJMCMC in a limited perspective, particularly, within the same classes of models and trans-dimensional cases. However, the original formulation of Green lends itself to a much wider interpretation than just exploring spaces ("*jumping*" in RJMCMC jargon) of different dimensions. A very interesting and important example that we study extensively in this thesis: the same formulation can be used to explore spaces of different generic models such as linear and nonlinear variable spaces. This is more than just exploring spaces of different sizes corresponding to the dimension of the parameter vector.

1.2. Objectives of the Thesis

In this thesis, our most important aim is to reveal unknown potentials of RJMCMC algorithm as a general Bayesian model selection method beyond trans-dimensionality. Particularly, in order to demonstrate the capability of RJMCMC in exploring not only model spaces of different dimensionality but also different generic models, we theoretically aim

- to propose new types of RJMCMC moves which extend the classical trans-dimensional approaches to a wider meaning with applications to a system identification study for Volterra system models,

- to perform model estimation studies on nonlinear time series models, firstly, by proposing a perspective for classical trans-dimensional approaches in transitions between linear and nonlinear model spaces.

- to generalize the proposed approach and to utilize RJMCMC in exploring the spaces of different generic models with applications in exploring different distribution families.

- to show that RJMCMC can be used to test the nonlinearity of any observed data set by estimating the polynomial nonlinearity degree of the time series.

- and finally, to highlight future research directions on Bayesian signal processing with RJMCMC algorithm.

These theoretical aims have the following application areas also studied in this thesis:

- Estimation of nonlinear time series models with polynomial nonlinearity which are polynomial autoregressive (PAR), polynomial moving average (PMA) and polynomial autoregressive moving average (PARMA) models

- Volterra system identification, the estimation of nonlinear communication channels expressed in terms of discrete time baseband Volterra model

- Modelling distributions of various real life data, in particular distribution of impulsive noise

- Modelling the distribution of wind speed measurements of various places in the world

1.3. Organization of the Thesis

We start our discussion in Chapter 2 with the general introduction of Bayesian model selection methods. The main focus is the Metropolis-Hastings algorithm and its extension to a reversible jump perspective. The new interpretation of RJMCMC will also be addressed in this chapter.

Chapter 3 introduces the first application for the new perspective of RJMCMC. In order to show the use of the classical trans-dimensional RJMCMC in transitions between linear and nonlinear models, we have firstly performed nonlinear model estimation studies of PAR, PMA and PARMA. In addition to the system memory and coefficients, estimating the nonlinearity degree of the models is an important contribution of the proposed method.

Chapter 4 extends the model estimation studies of the previous chapter to a system identification study. In this chapter, we study the identification of Volterra systems

with RJMCMC. We propose a new type of move which performs transitions between spaces of different structural models irrespective of model dimensions. This move is called the switch move. As a case study, we investigate the performance of the proposed method on synthetically generated data including orthogonal frequency division multiplexing (OFDM) communications over a nonlinear channel.

Chapter 5 includes distribution modelling of a given noise/data with RJMCMC. Firstly, synthetically generated data sets are modeled with RJMCMC. We use the proposed method to choose among various impulsive distribution families, such as symmetric α-Stable, generalized Gaussian and Student's t to model both synthetically generated noise processes and real life measurements on powerline communications (PLC) impulsive noises, 2-D DWT coefficients and seismic acceleration time series. Moreover, another simulation scenario is implemented in order to show the modelling performance of the algorithm when the distribution in question is not symmetrical and positive semi-infinite (envelope distributions). To achieve this, we create a distribution family space with Nakagami, κ, Gamma, Weibull and Generalized Rayleigh distribution families. The algorithm is tested to model distribution of real life measurements of wind speed from various locations around the world.

We summarize and conclude the study presented in this thesis in Chapter 6. In the end, current and future research directions with RJMCMC are given.

CHAPTER 2

METHODOLOGY: BAYESIAN MODEL SELECTION

In this chapter, general information about Bayesian model selection will be presented in a perspective extending from Bayes Theorem to RJMCMC. In addition, the linear AR model estimation study, which has a large preliminary in Bayesian model uncertainty applications in the literature, will be presented as a toy example. This toy example is crucial to see how the classical trans-dimensional RJMCMC approach works in solving the model uncertainty applications and in estimating the models in question. The following chapters will make references to this development and show how it is extended for more complicated problems. The chapter will be concluded with a brief description of the proposed method, the trans-space RJMCMC.

2.1. Bayes Theorem

Suppose we are given observed data y with parameter θ. The joint probability distribution for θ and y can be expressed in terms of a product of two densities as (Gelman et al., 2003)

$$f(\theta, y) = f(\theta)f(y|\theta) \tag{2.1}$$

where $f(\theta)$ refers to the prior distribution and $p(y|\theta)$ is the data distribution. Then, by using the basic property of the conditional probability, Bayes' rule, yields the posterior density (Gelman et al., 2003)

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} \tag{2.2}$$

where $f(y)$ is the total probability and can be expressed as $\int f(y|\theta)f(\theta)d\theta$. The total probability, $f(y)$, does not depend on the parameter θ and is considered to be constant for a fixed y. Thus, another representation of Bayes's rule which omits the denominator of

the equation (2.2), yields the unnormalized posterior density (Gelman et al., 2003)

$$f(\theta|y) \propto f(y|\theta)f(\theta). \tag{2.3}$$

2.2. Bayes Factors

Let us suppose that we have two models, M_1 and M_2, with parameter vectors θ_1 and θ_2. The aim is to select the best model among these two models given the observed data $\mathbf{y}$. Then, by using the Bayes theorem in (2.2), the posterior probability of the model M_1 can be easily obtained as

$$f(M_1|\mathbf{y}) = \frac{f(\mathbf{y}|M_1)f(M_1)}{f(\mathbf{y}|M_1)f(M_1) + f(\mathbf{y}|M_2)f(M_2)} \tag{2.4}$$

where $f(\mathbf{y}|M_1)$ refers to the marginal likelihood distribution for model M_1. Similarly, the same can be obtained for model M_2.

In order to create a decision mechanism for questioning whether the observed data support M_2 over M_1 or not, posterior odds which are the ratio of their posteriors are measured. From (2.4), we can easily write

$$\frac{f(M_2|\mathbf{y})}{f(M_1|\mathbf{y})} = \left[\frac{f(\mathbf{y}|M_2)}{f(\mathbf{y}|M_1)}\right]\left[\frac{f(M_2)}{f(M_1)}\right] \tag{2.5}$$

where the term on the left-hand side is the *posterior odds*, the first term on the right-hand side is named as *Bayes factor* for M_2 against M_1, denoted by B_{21} and is also the ratio of the marginal likelihoods of the models. Moreover, the last term is the *prior odds* and thus, (2.5) can be rewritten as

$$\text{Posterior Odds} = \text{Bayes Factor} \times \text{Prior Odds}. \tag{2.6}$$

Posterior odds are equal to Bayes factor when the models are equally likely (prior odds are equal to 1). Usage and evaluation of the Bayes factors are studied first by Jeffreys

who published some rules to decide between models (Jeffreys, 1998). According to these rules, when $B_{ij} > 1$ data are more likely to be modeled with M_i, whereas for $B_{ij} < 1$ model M_j is favourable. Table 2.1 shows the Jeffreys' rules for evidence of Bayes factors which is an example of subjective, qualitative partitioning of decision regions of Bayes factor values (Raftery, 1995).

In some elementary cases, the marginal likelihood or the posterior distributions can be easily evaluated analytically. However, in many cases, these distributions require to solve intractable integrals and evaluating Bayes factors become impossible. Using methods which provide approximations to these intractable integrals, makes it possible to evaluate Bayes factors to decide between models. To name a few, *Laplace's Method* performs an asymptotic approximation by assuming that the posterior is highly peaky at its maximum while *Schwarz criterion* (or namely *BIC*) performs a rough approximation to the logarithm of the Bayes factors by avoiding using priors in the calculation of the marginal likelihoods (Kass and Raftery, 1995).

Table 2.1. Evidence for Bayes factors

Bayes Factor B_{ij}	Evidence for M_i
$1 \leq B_{ij} \leq 3$	not worth more than a bare mention
$3 \leq B_{ij} \leq 10$	positive
$10 \leq B_{ij} \leq 100$	strong
$B_{ij} > 100$	decisive

Another important approach is to estimate (or to approximate) the target distribution of interest (likelihood, posterior, etc.) via direct or inverse sampling by performing Monte Carlo simulations. The problem occurs when it is not possible to sample directly from target distribution of interest. Instead, other distinguished methods such as importance sampling, MCMC methods, etc., which enable to sample from a distribution which becomes closer and closer to the target distribution at each iteration, can be used in these cases.

2.3. Markov Chain Monte Carlo Methods

The key property of MCMC methods is to create a Markov process the stationary distribution π of which is equal to the target distribution of interest, even when we

have non-informative priors and likelihoods with systems that cannot be observed directly or with missing data (Besag et al., 1995). If we run the simulation long enough, the distribution of our samples is close to this stationary distribution. This makes MCMC fundamentally more outstanding than the other sampling algorithms such as importance sampling and the like (Gelman et al., 2003).

We give an explanation about the notation. A probability distribution is used sometimes for the probability density function (pdf) but generally to deal with a more complete installation of the probabilities to all measurable subsets of outcomes. In the rest of the dissertation, π (or f) will be used to denote both probability distribution and density in order to obtain a clear and simpler notation.

Firstly, suppose that a probability distribution π has a density with respect to a measure μ, and let π be of the form (Tierney, 1998)

$$\pi(dx) = \pi(x)\mu(dx) \qquad (2.7)$$

where $\pi(x)$ represents the probability density.

In order to use the key properties of MCMC, we need to create a Markov chain which is *irreducible*, *aperiodic* and *time reversible*. In particular,

- a Markov chain is *irreducible* with stationary distribution π, if it has a nonzero probability to enter any states which π assigns a positive probability for any initial states.

- a state of Markov chain has period d if any return to this state is the multiple of d time steps. If this multiple d is equal to 1, the chain is *aperiodic*.

If a Markov chain with stationary distribution π, is irreducible and aperiodic, then π is the unique stationary distribution and π is also the equilibrium distribution (Tierney, 1994; Andrieu et al., 2003). One can easily design MCMC samplers providing the Markov chain satisfies the detailed balance (or reversibility) equation (sufficient but not necessary) which is (Tierney, 1998)

$$\pi(dx)P(x, dx') = \pi(dx')P(x', dx) \qquad (2.8)$$

where $\pi(dx)$ and $\pi(dx')$ are the stationary distributions for states x and x', respectively,

and $P(x, dx')$ is the Markov transition kernel[1] from state x to x'.

The three main methods which implement the properties above will be examined in this sub-section. First, MCMC methodology will be constructed by jointly providing details of the Metropolis-Hastings (MH) algorithm (Hastings, 1970) which is the generalization of the Metropolis (Metropolis et al., 1953) algorithm. Then, the Gibbs Sampler (Geman and Geman, 1984) which is a special case of The MH algorithm will be examined.

2.3.1. The Metropolis-Hastings Algorithm

The MH algorithm accommodates many alternative transition kernels, P, in order to have a Markov chain which has the properties above. The construction, at a given state x is as follows. A candidate state x' is proposed from a probability distribution $q(dx'|x)$.

Let us assume that the transition kernel which satisfies the properties above has been constructed in two steps; firstly proposing a new candidate space x' and then accepting this state transition with a probability of $\alpha(x \rightarrow x')$. Thus, the transition kernel is

$$P(x, dx') = q(dx'|x)\alpha(x \rightarrow x') \tag{2.10}$$

where $q(dx'|x)$ is the proposal distribution which has a density with respect to a measure in the same form that is shown in (2.8), and the Markov chain remains at the same state with probability $1 - \alpha(x \rightarrow x')$.

In the MH algorithm, the main objective is to derive an expression for *the acceptance ratio*, $\alpha(x \rightarrow x')$, which achieves the stated aim of providing the detailed balance. The transition kernel P satisfies the time reversibility condition in (2.8). So, substituting

---

[1]**Definition:** (Tierney, 1994) A Markov transition kernel, $P(X_n, A)$, is a conditional probability representation when defining a transition from the state X_n to the next state X_{n+1} which is a member of A. The transition kernel is defined by

$$P(X_n, A) = P\{X_{n+1} \in A | X_0, \ldots, X_n\} \tag{2.9}$$

for all measurable sets A.

the expression in (2.10) into (2.8) gives

$$\pi(dx)q(dx'|x)\alpha(x \to x') = \pi(dx')q(dx|x')\alpha(x' \to x). \qquad (2.11)$$

The time reversibility equality in (2.11) can be written as

$$\alpha(x \to x') = \alpha(x' \to x)\frac{\pi(dx')q(dx|x')}{\pi(dx)q(dx'|x)} \qquad (2.12)$$

$$= \alpha(x' \to x)r(x' \to x). \qquad (2.13)$$

As (Peskun, 1973) proved for the finite state space case, to make the acceptance ratio as large as possible, $\alpha(x \to x')$ is selected as

$$\alpha(x \to x') = \min\{1, r(x' \to x)\} = \min\left\{1, \frac{\pi(dx')q(dx|x')}{\pi(dx)q(dx'|x)}\right\} = \alpha_{\mathrm{MH}}(x \to x'). \qquad (2.14)$$

(Tierney, 1998) extends the results of (Peskun, 1973) from finite state space case to the general state space case. (Tierney, 1998) states that the acceptance ratio in (2.14) is optimal in the sense that it minimizes the asymptotic variance of the sample path averages among all acceptance rates satisfying (2.11). This acceptance ratio is the maximal Metropolis-Hastings kernel for P.

Corollary 2.1 (Sawyer (2006), Corollary 2.4.1, pg. 9) *The function $\alpha(x \to x')$ in (2.14) is the pointwise maximum value of all functions with $0 \le \alpha(x \to x') \le 1$ that satisfy the reversibility condition in (2.8) for $P(x, dx')$ in (2.10).*

Proof   Let Q be the set of all functions $\beta(x \to x')$ such that $0 \le \beta(x \to x') \le 1$ and

$$\pi(dx)q(dx'|x)\beta(x \to x') = \pi(dx')q(dx|x')\beta(x' \to x) \qquad (2.15)$$

for all $x, x' \in X$. Note that if $\beta_1(x \to x')$ and $\beta_2(x \to x')$ both satisfy (2.15), then so does $\beta_3(x \to x') = \max\{\beta_1(x \to x'), \beta_2(x \to x')\}$, and similarly so does $\alpha(x \to x') = \max_{\beta \in Q} \beta(x \to x')$.

(i) If $\pi(dx)q(dx'|x) = 0$ then $\beta(x \to x')$ can take any value due to (2.15) is 0. $\alpha(x \to x')$ must be selected as 1, in order to provide the fact $\alpha(x \to x') = \max_{\beta \in Q} \beta(x \to x')$.

13

(ii) If $0 \leq \pi(dx')q(dx|x') \leq \pi(dx)q(dx'|x)$ then $r(x' \rightarrow x) \leq 1$. From the definition $\beta(x \rightarrow x') = \beta(x' \rightarrow x)r(x' \rightarrow x)$, $\alpha(x' \rightarrow x)$ must be 1 in order to provide the fact $\alpha(x \rightarrow x') = \max_{\beta \in Q} \beta(x \rightarrow x')$. Thus,

$$\alpha(x \rightarrow x') = \frac{\pi(dx')q(dx|x')}{\pi(dx)q(dx'|x)} \leq 1. \tag{2.16}$$

(iii) If $0 \leq \pi(dx)q(dx'|x) \leq \pi(dx')q(dx|x')$ then $r(x' \rightarrow x) \geq 1$. From the definition $\beta(x \rightarrow x') = \beta(x' \rightarrow x)r(x' \rightarrow x)$, $\alpha(x' \rightarrow x)$ must be $\dfrac{1}{r(x' \rightarrow x)}$ in order to provide the facts $\alpha(x \rightarrow x') = \max_{\beta \in Q} \beta(x \rightarrow x')$ and $0 \leq \beta(x \rightarrow x') \leq 1$. Hence,

$$\alpha(x \rightarrow x') = 1. \tag{2.17}$$

By using these three cases, acceptance ratio $\alpha(x \rightarrow x')$ can be written clearly as

$$\alpha(x \rightarrow x') = \min\left\{1, \frac{\pi(dx')q(dx|x')}{\pi(dx)q(dx'|x)}\right\}. \tag{2.18}$$

This completes the proof of Corollary 2.1.■   □

Corollary 2.1 states that, the form of $\alpha(x \rightarrow x')$ in (2.14), including all acceptance ratio expressions satisfying (2.11), has the smallest probability of remaining at the same state x for all $x \in X$.

Acceptance ratio can be defined with respect to probability densities by using the property in (2.7)

$$\alpha_{\mathrm{MH}}(x \rightarrow x') = \min\left\{1, \frac{\pi(x')\mu(dx')q(x|x')\mu(dx)}{\pi(x)\mu(dx)q(x'|x)\mu(dx')}\right\} \tag{2.19}$$

$$= \min\left\{1, \frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)}\right\}. \tag{2.20}$$

Substituting the target distribution for parameters $\boldsymbol{\theta}$ given data $\mathbf{y}$ with equilibrium distri-

bution gives

$$\alpha_{\text{MH}}(\boldsymbol{\theta} \to \boldsymbol{\theta}^*) = \min\left\{1, \frac{f(\boldsymbol{\theta}^*|\mathbf{y})q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{f(\boldsymbol{\theta}|\mathbf{y})q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}\right\} \tag{2.21}$$

where $\boldsymbol{\theta}^*$ is the proposed parameters from distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$. In Figure 2.1 an example of Bayesian inference in terms of the MH algorithm is given for estimating a mixture of two univariate Gaussian distributions.
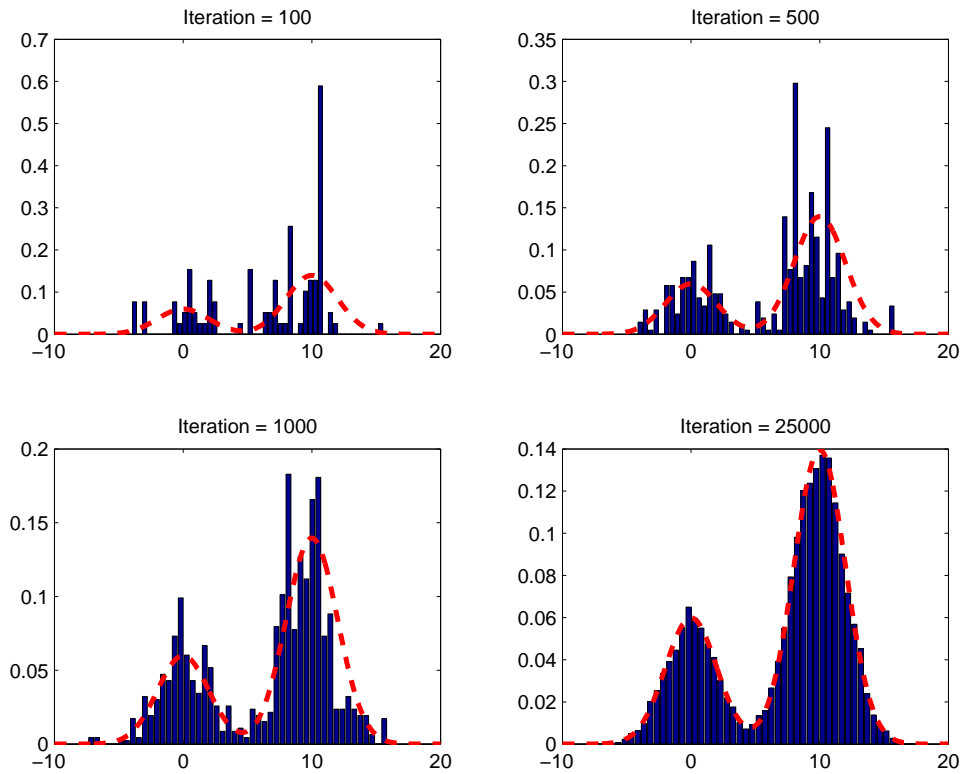
Figure 2.1. The Metropolis-Hastings Algorithm. Dotted line represents target distribution and bars refer to the estimated distribution of Metropolis-Hastings algorithm (Source: (Andrieu et al., 2003)).

2.3.2. The Metropolis Algorithm

The Metropolis Algorithm is constructed by (Metropolis et al., 1953) before Hastings generalized the method in 1970. This method uses the facts in Subsection 2.3.1 in a more specific manner. At every iteration, we sample $\boldsymbol{\theta}^*$ from a *proposal distribution*

$q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$. This proposal distribution which is used in the Metropolis algorithm should be symmetric. This means that; $q(\boldsymbol{\theta}_a|\boldsymbol{\theta}_b) = q(\boldsymbol{\theta}_b|\boldsymbol{\theta}_a)$. The ratio which is required to reach the acceptance ratio is given by (Gelman et al., 2003) as

$$\alpha_{\mathrm{Metropolis}}(\boldsymbol{\theta} \rightarrow \boldsymbol{\theta}^*) = \min\left\{1, \frac{f(\boldsymbol{\theta}^*|\mathbf{y})}{f(\boldsymbol{\theta}|\mathbf{y})}\right\}. \qquad (2.22)$$

In Figure 2.2 an example of Metropolis algorithm is shown for estimating the distribution of a univariate Gaussian distribution problem.
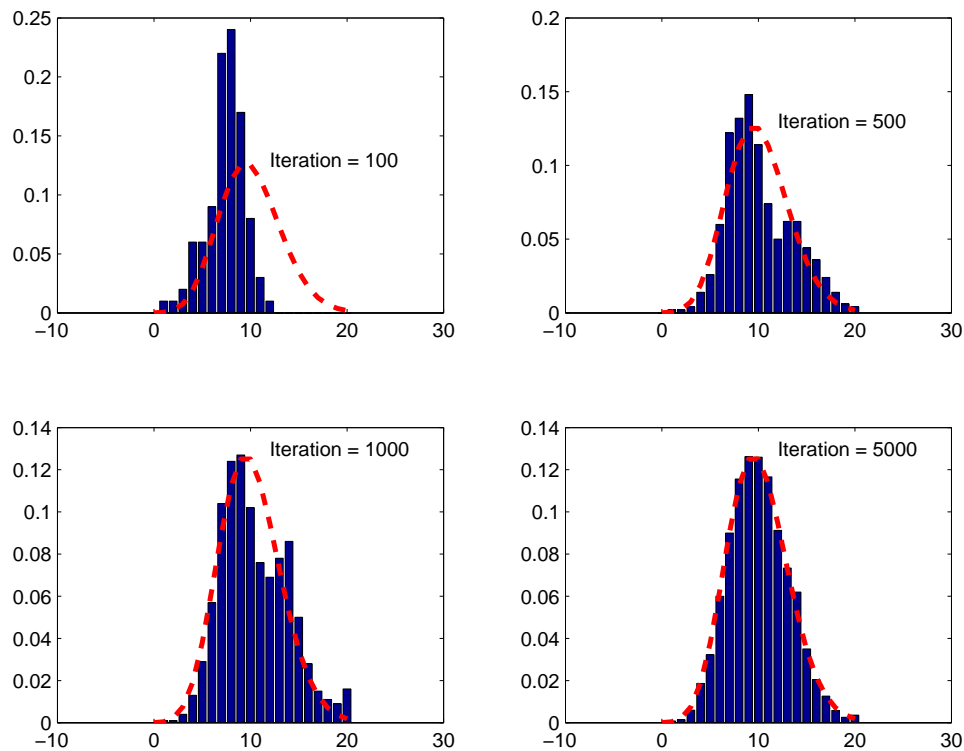
Figure 2.2. The Metropolis Algorithm. Dotted line represents target distribution and bars refer to the estimated distribution of Metropolis algorithm.

2.3.3. The Gibbs Sampler

Gibbs Sampler is an important Markov chain algorithm which is very useful in many multidimensional problems. The parameter vector of Gibbs Sampler, $\boldsymbol{\theta}$, is divided into d subvectors, $\boldsymbol{\theta} = \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_d$. Gibbs sampler needs d conditional distributions for each

subvector and at each iteration t, each $\boldsymbol{\theta}_j$ is sampled from these conditional distributions given all other subvectors of $\boldsymbol{\theta}$ (Gelman et al., 2003)

$$f(\boldsymbol{\theta}_j|\boldsymbol{\theta}_{-j}, \mathbf{y}) \tag{2.23}$$

where $\boldsymbol{\theta}_{-j}$ represents all the components of $\boldsymbol{\theta}$, except for $\boldsymbol{\theta}_j$, at their current values

$$\boldsymbol{\theta}_{-j} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{j-1}, \boldsymbol{\theta}_{j+1}, ..., \boldsymbol{\theta}_d). \tag{2.24}$$

Gibbs Sampler may be viewed as a special case of the MH algorithm where acceptance ratio is always equal to 1 leading to the result that every jump is accepted by choosing the proposal distribution in (2.14) as

$$q(x'|x) = \pi(x') \Rightarrow \alpha_{\text{Gibbs}}(x \rightarrow x') = 1. \tag{2.25}$$

In Figure 2.3 an example of Gibbs Sampler is shown for estimating the distribution of a bivariate Gaussian distribution.

2.4. RJMCMC

Up to this point, we have seen in general what the Bayesian model selection is and what type of Bayesian methods will ultimately achieve this. The MH algorithm has been widely used for model estimation studies within a single model space. However, in cases where the model space contains a huge number of models, estimating each model space separately and deciding the best model can be computationally prohibitive. At this point, Peter J. Green has extended the classical MH algorithm by treating the model space dimension as a new parameter and hence provided a model determination algorithm achieving transitions between model spaces of differing dimensions in (Green, 1995). The most important feature of this method is that by locating model spaces of different dimensionality in a general model space, it is possible to perform trans-dimensional transitions between these spaces and decide the best model without performing an exhaustive search.

Following (Green, 1995), when the current state is x, we propose a move type m
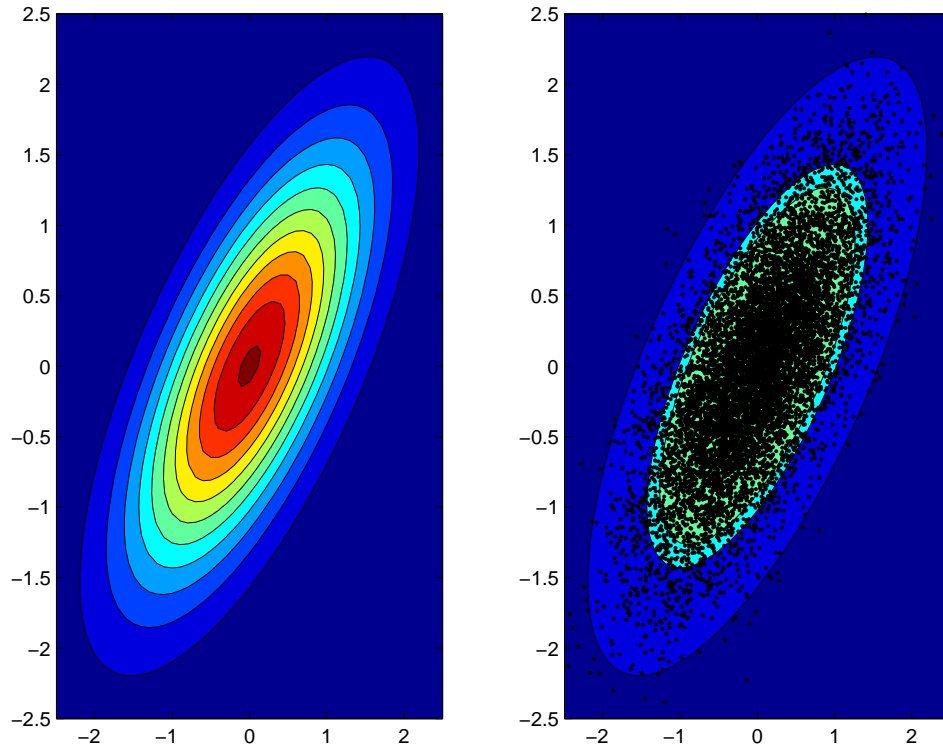
Figure 2.3. Contour plots for Gibbs Sampler. Figure on the left represents target distribution; figure on the right refers to the estimated distribution with dots.

with probability $Pr(x \to x')$, which changes dimension, and takes the state to x'. The algorithm follows the same procedure like MH algorithm and as usual, the proposal is not automatically accepted. The acceptance probability which is denoted by $\alpha(x \to x')$ needs to be calculated.

The assumption of (Green, 1995) in pg.715 states that, if we suppose that $\pi(dx)q(dx'|x)$ has a finite (or discrete) density f_m with respect to a symmetric measure ξ_m on $\mathcal{C} \times \mathcal{C}$,

$$f_m(dx, dx') = f_m(x, x')\xi_m(dx, dx') \qquad (2.26)$$

which can be written by using the property in (2.7), where $\mathcal{C}$ is the parameter space. Then,

for Borel sets[2] A and B,

$$\int_A \pi(dx) \int_B q(dx'|x)\alpha(x \to x') = \int_A \int_B \xi_m(dx, dx') f_m(x, x')\alpha(x \to x') \qquad (2.27)$$

$$= \int_B \int_A \xi_m(dx', dx) f_m(x', x)\alpha(x' \to x) \qquad (2.28)$$

$$= \int_B \pi(dx') \int_A q(dx|x')\alpha(x' \to x). \qquad (2.29)$$

This equality holds by the assumed symmetry of ξ_m, provided that

$$f_m(x, x')\alpha(x \to x') = f_m(x', x)\alpha(x' \to x). \qquad (2.30)$$

Equation (2.30) shows that detailed balance equation in (2.8) is satisfied under the assumption of (Green, 1995) and the acceptance ratio can be constructed by using the same procedure in MH part given in Corollary 2.1. Acceptance ratio can be written as

$$\alpha(x \to x') = \min\left\{1, \frac{f_m(dx', dx)}{f_m(dx, dx')}\right\} \qquad (2.31)$$

$$= \min\left\{1, \frac{f_m(x', x)\xi_m(dx', dx)}{f_m(x, x')\xi_m(dx, dx')}\right\} \qquad (2.32)$$

$$= \min\left\{1, \frac{f_m(x', x)}{f_m(x, x')}\right\} \qquad (2.33)$$

where density, f_m is

$$f_m(x, x') = \pi(x)q(x'|x). \qquad (2.34)$$

We firstly consider a set-up before implementing the dimension matching requirement. A reversible move is a projection of state x to state x' via a deterministic function w. This deterministic function w must be reversible to apply a move from one space to another. As stated above, RJMCMC provides jumping between spaces with different dimensions. As a result of this, dimension difference between the two states occurs and

---

[2]**Definition:** (Leon-Garcia, 2008) A Borel field $\mathcal{B}$, is the σ-field generated by countable unions, countable intersections and complements of events. Thus, we can select our event field $\mathcal{F}$ as a Borel field in probabilistic manner because $\mathcal{B}$ contains all events for discrete and countable sample spaces of real line.

dimension matching is required. For this reason, auxiliary variables $\mathbf{u}_1$ with length m_1, and its reverse move counterpart $\mathbf{u}_2$ with length m_2, are proposed from distributions $q_1(\mathbf{u}_1)$ and $q_2(\mathbf{u}_2)$, respectively. Resulting dimension matching is (Green, 1995)

$$dim_{x'} + m_2 = dim_x + m_1. \qquad (2.35)$$

For the moves that involve dimension change, we should also define a move that switches back to the previous state. These moves can be named as *reversible move pairs*. For example, a move which reduces the dimension of the state by 1 is a reversible move pair with a move that increases the dimension by 1 (Smith, 2007).

Suppose we are at state $x \in C_1$ with parameters $\boldsymbol{\theta}$ and we propose a move to another state $x' \in C_2$ with parameters $\boldsymbol{\theta}^*$ where C_1 and C_2 are subspaces of the main parameter space C, such as $C_1 \subset C$ and $C_2 \subset C$. Also suppose that $f(\boldsymbol{\theta}|x)$ and $f(\boldsymbol{\theta}^*|x')$ are proper densities in $\mathcal{R}^{n_1}$ and $\mathcal{R}^{n_2}$. In applying this move, typically we propose a vector of variables $\mathbf{u}_1$, independent of $\boldsymbol{\theta}$ as stated above and set $\boldsymbol{\theta}^*$ in terms of $\mathbf{u}_1$ and $\boldsymbol{\theta}$. A similar process should be applied in order to turn back to the current state. Specifically, we need to propose a vector of variables $\mathbf{u}_2$ independent of $\boldsymbol{\theta}^*$, and set $\boldsymbol{\theta}$ in terms of $\mathbf{u}_2$ and $\boldsymbol{\theta}^*$. There must be a bijection[3] between parameter spaces $(\boldsymbol{\theta}, \mathbf{u}_1)$ and $(\boldsymbol{\theta}^*, \mathbf{u}_2)$ in order to provide the dimension matching. For this bijection, the deterministic function w will be used. At the end, all of the parameters must satisfy the rule in (2.35). The proposal distribution $q(x'|x)$ can be defined by the distributions q_1 and q_2. These distributions are required to be proper densities[4] with respect to Lebesgue measures in $\mathcal{R}^{m_1}$ and $\mathcal{R}^{m_2}$, respectively (Green, 1995).

This set-up shows us that now our proposal distribution should include both probabilities of proposing a move from state x to x' and proposing the new parameters (Vermaak et al., 2004). By using this fact, it is more appropriate to write proposal distribution $q(x'|x)$ in the form of $q(x', \boldsymbol{\theta}^*|x, \boldsymbol{\theta})$. However, the proposal distribution includes both the move probability and the probability of proposing the new parameters, which two probabilities are independent. In order to provide this conditional independence, proposal

---

[3]**Definition:** (Wolf, 1997) If g is a one-to-one function from A onto B, g is called a *bijection* or *one-to-one correspondence* between A and B. The Notation is: $f : A \xrightarrow{bij} B$

[4]**Definition:** (Stark and Woods, 2002) A proper density, $g(x)$, is a probability density function which satisfies the properties above

(i) $g(x) \geq 0$,

(ii) $\int_{-\infty}^{\infty} g(x)dx = 1$.

distribution $q(x', \boldsymbol{\theta}^*|x, \boldsymbol{\theta})$ can be written as a multiplication of two different densities such as

$$q(x', \boldsymbol{\theta}^*|x, \boldsymbol{\theta}) = Pr(x \to x')f(\boldsymbol{\theta}^*|\boldsymbol{\theta}) \qquad (2.36)$$

where $Pr(x \to x')$ represents the probability of choosing a move m from the state x to x'. Since the only proposed parameter value is the auxiliary variable $\mathbf{u}_1$ in order to provide the dimension change for move m, $f(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ will be the probability of the auxiliary variable $\mathbf{u}_1$ sampled from the distribution q_1. The resulting equation for $q(x', \boldsymbol{\theta}^*|x, \boldsymbol{\theta})$ is

$$q(x', \boldsymbol{\theta}^*|x, \boldsymbol{\theta}) = Pr(x \to x')q_1(\mathbf{u}_1). \qquad (2.37)$$

The same procedure is appropriate for the case of reverse move version of the distribution, which is

$$q(x, \boldsymbol{\theta}|x', \boldsymbol{\theta}^*) = Pr(x' \to x)f(\boldsymbol{\theta}|\boldsymbol{\theta}^*) \qquad (2.38)$$

where $Pr(x' \to x)$ is the probability of the reverse move that is from x' to x and $f(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ is the proposal distribution when the recent state is x' with parameters $(\boldsymbol{\theta}^*, \mathbf{u}_2)$. This parameter space is a mapping of $(\boldsymbol{\theta}, \mathbf{u}_1)$ to $(\boldsymbol{\theta}^*, \mathbf{u}_2)$ via a deterministic function w. This operation needs a calculation in order to satisfy the following theorem.

Theorem 2.1 ((Vomisescu, 2003), Change of variables on Lebesgue measures) *Let $\Omega \subset \mathcal{R}^n$ be an open set and L be a Lebesgue measure on Ω. Let $T(x) = (y_1(x), \ldots, y_n(x))^T$ and $x = (x_1, \ldots, x_n)^T$ be a homeomorphism[5] $T : \Omega \to \mathcal{R}^n$ with continuous derivatives and*

---

[5]***Definition:*** *(Hubbard and West, 1995) X and Y are the subsets of $\mathcal{R}^n$ and a mapping $g : X \to Y$ is a homeomorphism if*

- *g is a continuous bijection,*
- *g is one-to-one and onto, so inverse function g^{-1} exists,*
- *g^{-1} is continuous.*

$J(T, x)$ is the Jacobian matrix for all $x \in \Omega$. Then, for any function f, we have

$$\int_{T\Omega} g(y) dy = \int_{\Omega} g(Tx) |J(T, x)| \, dx. \tag{2.39}$$

Hence, $f(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ can be written as

$$f(\boldsymbol{\theta}|\boldsymbol{\theta}^*) = q_2(\mathbf{u}_2) \left| \frac{\partial(\boldsymbol{\theta}^*, \mathbf{u}_2)}{\partial(\boldsymbol{\theta}, \mathbf{u}_1)} \right| \tag{2.40}$$

and the resulting equation for $q(x, \boldsymbol{\theta}|x', \boldsymbol{\theta}^*)$ is

$$q(x, \boldsymbol{\theta}|x', \boldsymbol{\theta}^*) = Pr(x' \to x) q_2(\mathbf{u}_2) \left| \frac{\partial(\boldsymbol{\theta}^*, \mathbf{u}_2)}{\partial(\boldsymbol{\theta}, \mathbf{u}_1)} \right|. \tag{2.41}$$

Invoking the assumption of (Green, 1995) shown in (2.27), the finite densities $f_m(x, x')$ and $f_m(x', x)$ are selected as

$$f_m(x, x') = \pi(x) q(x', \boldsymbol{\theta}^*|x, \boldsymbol{\theta}) = f(\boldsymbol{\theta}|y) Pr(x \to x') q_1(\mathbf{u}_1) \tag{2.42}$$

$$f_m(x', x) = \pi(x') q(x, \boldsymbol{\theta}|x', \boldsymbol{\theta}^*) = f(\boldsymbol{\theta}^*|y) Pr(x' \to x) q_2(\mathbf{u}_2) \left| \frac{\partial(\boldsymbol{\theta}^*, \mathbf{u}_2)}{\partial(\boldsymbol{\theta}, \mathbf{u}_1)} \right|. \tag{2.43}$$

Combining (2.42) and (2.43) in (2.31), the general expression for the acceptance ratio defining a move with a transition function w from state x to x' is given as

$$\alpha_{\text{RJMCMC}}(x \to x') = \min \left\{ 1, \frac{f(\boldsymbol{\theta}^*|\mathbf{y})}{f(\boldsymbol{\theta}|\mathbf{y})} \times \frac{Pr(x' \to x)}{Pr(x \to x')} \times \frac{q_2(\mathbf{u}_2)}{q_1(\mathbf{u}_1)} \times \left| \frac{\partial(\boldsymbol{\theta}^*, \mathbf{u}_2)}{\partial(\boldsymbol{\theta}, \mathbf{u}_1)} \right| \right\} \tag{2.44}$$

where $f(.|\mathbf{y})$ is target distribution of interest, $Pr(x \to x')$ and $Pr(x' \to x)$ represent the probabilities for move m and its reverse move, $q_1(\mathbf{u}_1)$ is the proposal distribution for auxiliary variable $\mathbf{u}_1$ for move m, $q_2(\mathbf{u}_2)$ is the proposal distribution for auxiliary variable $\mathbf{u}_2$ for reverse move and $\left| \frac{\partial(\boldsymbol{\theta}^*, \mathbf{u}_2)}{\partial(\boldsymbol{\theta}, \mathbf{u}_1)} \right|$ is the magnitude of the Jacobian determinant. This form of acceptance ratio is also the one which is defined in (Green, 1995) as Equation (7).

In practice, generally, there is no need to generate an auxiliary variable u_i for one of the reversible move pairs. Here $\mathbf{u}_2$, which is the reverse move counterpart of $\mathbf{u}_1$, will not be generated; so, the acceptance ratio is simplified. This simplified form of acceptance

ratio is given below and it forms as it's defined in (Green, 1995) as Equation (8)

$$\alpha_{\text{RJMCMC}}(x \to x') = \min \left\{ 1, \frac{f(\boldsymbol{\theta}^*|\mathbf{y})Pr(x' \to x)}{f(\boldsymbol{\theta}|\mathbf{y})Pr(x \to x')q_1(\mathbf{u}_1)} \left| \frac{\partial \boldsymbol{\theta}^*}{\partial(\boldsymbol{\theta}, \mathbf{u}_1)} \right| \right\}. \qquad (2.45)$$

In this derivation, parameter $\mathbf{u}_1$ is generated independently from the parameter vector $\boldsymbol{\theta}$. In real applications, $\mathbf{u}_1$ may be generated as depending on $\boldsymbol{\theta}$. This requires a modification to be made. In these cases, $q_1(\mathbf{u}_1)$ in (2.45) is replaced by the conditional density $q_1(\mathbf{u}_1|\boldsymbol{\theta})$ (Green, 1995).

If we assume that target distributions are equal to likelihoods times priors the general form of RJMCMC acceptance ratio is $\alpha_{\text{RJMCMC}}(x \to x') = \min\{1, r\}$ and where r is (Green, 1995)

$$r = \left\{ \begin{array}{c} \text{Likelihood} \\ \text{Ratio} \end{array} \right\} \times \left\{ \begin{array}{c} \text{Prior} \\ \text{Ratio} \end{array} \right\} \times \left\{ \begin{array}{c} \text{Proposal} \\ \text{Ratio} \end{array} \right\} \times \{Jacobian\}. \qquad (2.46)$$

2.5. On Convergence and Complexity of (RJ)MCMC Algorithms

As stated in sections above, the central objective of (RJ)MCMC sampling algorithms is to provide an approximation for the distribution of the samples to the target distribution or the posterior. As long as we run the simulation long enough, the distribution of the samples converges to this stationary distribution. Estimation statistics such as the mean and the autocorrelation may be used to monitor the convergence of (RJ)MCMC methods, in the absence of techniques to select a suitable run length a priori. Although there are diagnostic convergence tests which provide results about convergence, it is very difficult to be sure that the convergence of a multivariate distribution has been adequately monitored. The main reason is the high number of parameter dimensions (Hastie and Green, 2012).

Consequently, selection of an optimal run length a priori or monitoring convergence is an open problem. Despite all these ambiguities about the convergence, in the literature, there are studies (Gelman and Rubin, 1992a,b; Brooks and Giudici, 2000; Rosenthal, 1995) proposing methods to give an idea of the convergence. Particularly, in (Gelman and Rubin, 1992a,b) authors proposed a method which tries to decide whether

the algorithm converges to a stationary distribution by replicating multiple chains. This method has been generalized as a two-way *analysis of variance* (ANOVA) based method for univariate cases in (Brooks and Giudici, 2000). The multivariate version of Brooks and Guidici's method has been proposed in (Castelloe and Zimmerman, 2002). Sisson and Fan proposed a specific distance-based diagnostic which is designed for trans-dimensional cases and covers the modelling scenarios like finite-mixture problems and change point analyses in (Sisson and Fan, 2007). In all the applications in this thesis, the convergence of the algorithm was analyzed with some pre-tests, and the length of the algorithm was determined according to these tests. Therefore, the selected Markov chain was assumed to be long enough and it was assumed to converge correctly with a Markov chain in this length.

Apart from the convergence of RJMCMC, the computational complexity is also a key issue and known to be high as in the other sampling algorithms. It is also an open problem as the convergence and directly related to the convergence of the algorithm. Notwithstanding this, the complexity has also been studied in the literature. In (Belloni and Chernozhukov, 2009), it has been shown that MCMC algorithms have polynomial complexity in terms of computational time and computational complexity of MCMC methods is lower than the classical maximum likelihood methods as long as the log-likelihood are nonconcave or nonsmooth.

2.6. Toy Example: Autoregressive Model Estimation using RJMCMC

A classical application of RJMCMC method in model estimation studies can be basically discussed on AR model uncertainty. For some detailed studies on AR model estimation, interested readers may see (Troughton and Godsill, 1998), (Vermaak et al., 2004) and (Ehlers and Brooks, 2003). We provide a development of the RJMCMC approach for this previously studied problem since we will use this terminology on more complicated problems in the following chapters.

The Autoregressive (AR) data model is given below

$$y(l) = \sum_{i=1}^{k} a_i y(l - i) + \epsilon(l) \tag{2.47}$$

where $\epsilon(l)$ is an excitation sequence the distribution of which is a zero mean Gaussian distribution with a variance of σ_e^2, which can be shown as $\mathcal{N}(0, \sigma_e^2)$. Moreover, a_i is the AR coefficient for lag i and k is the model order.

This can be written in matrix-vector form as

$$\mathbf{y} = \mathbf{Y}\mathbf{a}^{(k)} + \boldsymbol{\epsilon} \tag{2.48}$$

where $\mathbf{y}$ is data vector with length n, $\mathbf{Y}$ is $n \times k$ matrix of past samples of data, $\mathbf{a}^{(k)}$ is coefficient vector of AR(k) model and $\boldsymbol{\epsilon}$ is vector of excitation sequence whose forms are given below

$$\mathbf{y} = [y(1), y(2), \ldots, y(n)]^T \tag{2.49}$$

$$\mathbf{Y} = \begin{bmatrix} y(0) & y(-1) & \ldots & y(1-k) \\ y(1) & y(0) & \ldots & y(2-k) \\ \vdots & \vdots & \ddots & \vdots \\ y(n-1) & y(n-2) & \ldots & y(n-k) \end{bmatrix}, \tag{2.50}$$

$$\mathbf{a}^{(k)} = [a_1, a_2, \ldots, a_k]^T \tag{2.51}$$

$$\boldsymbol{\epsilon} = [\epsilon(1), \epsilon(2), \ldots, \epsilon(n)]^T. \tag{2.52}$$

Since the excitation sequence is Gaussian, the approximate likelihood is (Troughton and Godsill, 1998)

$$f(\mathbf{y}|k, \mathbf{a}^{(k)}, \sigma_e^2) \approx \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \sigma_e^2 \mathbf{I}_n) \tag{2.53}$$

where $\boldsymbol{\epsilon}$ is the excitation sequence vector.

2.6.1. Hierarchial Model and Priors

Target distribution $f(\boldsymbol{\theta}|\mathbf{y})$ can be defined as shown below for AR model order selection problem

$$f(\boldsymbol{\theta}|\mathbf{y}) = f(k, \mathbf{a}^{(k)}, \sigma_e^2, \sigma_a^2|\mathbf{y}) \qquad (2.54)$$

where σ_a^2 is the variance for parameter vector $\mathbf{a}^{(k)}$. Hence, this posterior density can be written from Bayes Theorem as

$$f(k, \mathbf{a}^{(k)}, \sigma_e^2, \sigma_a^2|\mathbf{y}) \propto f(\mathbf{y}|k, \mathbf{a}^{(k)}, \sigma_e^2) f(\mathbf{a}^{(k)}|k) f(\sigma_a^2) f(\sigma_e^2) f(k). \qquad (2.55)$$

Determining the prior distributions of the parameters is effective on the convergence speed of the RJMCMC algorithm. Firstly, we choose a uniform prior for model order with maximum value $k_{\max}$

$$f(k) = \mathcal{U}(1, k_{\max}). \qquad (2.56)$$

We also choose conjugate priors for other parameters (Gelman et al., 2003)

$$f(\mathbf{a}^{(k)}|k) = \mathcal{N}(\mathbf{a}^{(k)}|\mathbf{0}, \sigma_a^2 \mathbf{I_k}) \qquad (2.57)$$

$$f(\sigma_a^2) = \mathcal{IG}(\sigma_a^2|\alpha_a, \beta_a) \qquad (2.58)$$

$$f(\sigma_e^2) = \mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e) \qquad (2.59)$$

where $\mathcal{IG}(\alpha, \beta)$ is the Inverse Gamma distribution with shape parameter α and scale parameter β.

2.6.2.   RJMCMC Transitions

The general expression for acceptance ratio is given in (2.45). This equation will be the main expression for this problem and by adding the probabilistic information about the likelihood and the priors to this expression we are going to obtain a specific acceptance ratio expression.

Sampling model order value k requires changing dimension in the algorithm. Firstly, we define the move pairs for this problem. There are 4 types of moves for this problem which are:

(a) birth of a new parameter,

(b) death of an existing parameter,

(c) life move, updating the AR coefficients via MH algorithm,

(d) updating σ_e^2 via Gibbs sampling.

Birth and death moves change dimension by 1 up and down, respectively and for this birth-death-move, reversible jump mechanism is needed. We design these two moves in tandem so that they are a reversible pair.

Life move does not require changing the dimension of the space, it updates the AR coefficients for each move and applies MH algorithm to accept the proposals. Move (d) is a Gibbs sampling application and it is repeated at each iteration according to the values that are proposed with respect to the first 3 moves. Life move and move (d) are designed in a condition so that the reversible pair is also move's itself (Smith, 2007).

Each move has probabilities P_{birth} for birth move, P_{death} for death move and P_{life} for life move depending only on the current value of model order parameter k and satisfying $P_{\text{birth}} + P_{\text{death}} + P_{\text{life}} = 1$. There is no need to assign a probability for move (d) because it is repeated at each iteration independently from the selected move. Naturally for $k = k_{\max}$ $P_{\text{birth}} = 0$ and $P_{\text{life}} = 1 - P_{\text{death}}$, for $k = 1$, $P_{\text{death}} = 0$ and $P_{\text{life}} = 1 - P_{\text{birth}}$. Except these constraints, these probabilities are chosen in a way that (Green, 1995)

$$P_{\text{birth}} f(k) = P_{\text{death}} f(k + 1). \tag{2.60}$$

Choosing birth and death move probabilities according to this condition would guarantee the certain acceptance in the corresponding MH sampler for the model order parameter.

2.6.2.1. Birth-Death Moves

Firstly, we propose a move m which gives a transition from a space with model order parameter k to a space with higher dimension $k' = k + 1$. Due to the increasing order, we should propose new parameters to provide the dimension matching. This move is called birth move. By using the proposed model order value k' and satisfying the dimension matching, we apply RJMCMC for birth move according to Algorithm (1).

Acceptance ratio is defined as $\alpha_{\text{birth}}(k \rightarrow k') = \min\{1, r_{\text{birth}}\}$. According to this, the corresponding value of r_{birth} which is based on the expression in (2.45), is given below

$$r_{\text{birth}} = \frac{f(\mathbf{y}|k', \mathbf{a}^{(\mathbf{k}')}, \sigma_e^2) f(\mathbf{a}^{(\mathbf{k}')}|k') f(\sigma_a^2) f(\sigma_e^2) f(k) P_{\text{death}}}{f(\mathbf{y}|k_{t-1}, \mathbf{a}^{(\mathbf{k_{t-1}})}, \sigma_e^2) f(\mathbf{a}^{(\mathbf{k_{t-1}})}|k_{t-1}) f(\sigma_a^2) f(\sigma_e^2) f(k) P_{\text{birth}} q(u)} \left| \frac{\partial \mathbf{a}^{(\mathbf{k}')}}{\partial (\mathbf{a}^{(\mathbf{k_{t-1}})}, u)} \right| \quad (2.61)$$

where u is the auxiliary variable with $n_u = 1$ to match the dimension between two spaces and $q(u)$ is the distribution that we sample u from. Dimension change occurs in autoregressive parameter space. Thus, Jacobian calculation depends on these values. If we simplify the equation, it attains the form given in (2.46)

$$r_{\text{birth}} = \frac{f(\mathbf{y}|k', \mathbf{a}^{(\mathbf{k}')}, \sigma_e^2)}{f(\mathbf{y}|k_{t-1}, \mathbf{a}^{(\mathbf{k_{t-1}})}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(\mathbf{k}')}|k')}{f(\mathbf{a}^{(\mathbf{k_{t-1}})}|k_{t-1})} \times \frac{P_{\text{death}}}{P_{\text{birth}} q(u)} \times \left| \frac{\partial \mathbf{a}^{(\mathbf{k}')}}{\partial (\mathbf{a}^{(\mathbf{k_{t-1}})}, u)} \right|. \quad (2.62)$$

As we define above, birth and death moves are reversible pairs and so, acceptance ratio for death move can be clearly determined. We propose a move m which changes states from k to $k' = k - 1$. Here no new parameters are proposed. We truncate $\mathbf{a}^{(\mathbf{k_{t-1}})}$ at the k'th element. The resulting algorithm for death move is given in Algorithm 2.

For $k > k'$ then $\alpha_{\text{death}}(k \rightarrow k') = \min\{1, r_{\text{death}}\}$ and the corresponding value for acceptance ratio of death move is calculated directly from the expression below

$$\alpha_{\text{death}}(k \rightarrow k') = \min\{1, 1/r_{\text{birth}}\} \quad (2.63)$$

where r_{birth} is calculated from

$$\alpha_{\text{birth}}(k' \rightarrow k) = \min\{1, r_{\text{birth}}\}. \quad (2.64)$$

Algorithm 1 Birth Move

1: **procedure** BIRTH MOVE FOR AR MODEL ORDER SELECTION WITH RJMCMC
2:     Given data $\mathbf{y}$
3:     $k_t, N_{iter}, \mathbf{a}^{(k_0)}, \alpha_e, \beta_e$ values are initialized at $t = 0$
4:     RJMCMC will be performed for $t = 1, \ldots, N_{iter}$
5:     Assume a birth move proposed at iteration t
6:     $k' \leftarrow k_{t-1} + 1$
7:     Propose $u \sim q(u)$
8:     Propose $\mathbf{a}^{(\mathbf{k}')}$ according to the Scenario
9:     Calculate $\alpha_{\text{birth}}(k_{t-1} \rightarrow k')$
10:    Propose $num \sim Unif(0, 1)$
11:    **if** $num \leq \alpha_{\text{birth}}(k_{t-1} \rightarrow k')$ **then**
12:       $k_t = k'$
13:       $\mathbf{a}^{(k_t)} = \mathbf{a}^{(k')}$
14:    **else**
15:       $k_t = k_{t-1}$
16:       $\mathbf{a}^{(k_t)} = \mathbf{a}^{(k_{t-1})}$
17:    **end if**
18:    Update $\sigma_e^2 \sim \mathcal{IG}(\alpha_n, \beta_n)$ via Gibbs sampling
19: **end procedure**

Algorithm 2 Death Move

1: **procedure** DEATH MOVE FOR AR MODEL ORDER SELECTION WITH RJMCMC
2:     Given data $\mathbf{y}$
3:     $k_t, N_{iter}, \mathbf{a}^{(k_0)}, \alpha_e, \beta_e$ values are initialized at $t = 0$
4:     RJMCMC will be performed for $t = 1, \ldots, N_{iter}$
5:     Assume a death move proposed at iteration t
6:     $k' \leftarrow k_{t-1} - 1$ Truncate a to the first k_(t-1)-1 elements as a^(k')
7:     Truncate $\mathbf{a}^{(\mathbf{k_{t-1}})}$ to the first k' elements as $\mathbf{a}^{(\mathbf{k}')}$
8:     Calculate $\alpha_{\text{death}}(k_{t-1} \rightarrow k')$
9:     Propose $num \sim Unif(0, 1)$
10:    **if** $num \leq \alpha_{\text{death}}(k_{t-1} \rightarrow k')$ **then**
11:       $k_t = k'$
12:       $\mathbf{a}^{(k_t)} = \mathbf{a}^{(k')}$
13:    **else**
14:       $k_t = k_{t-1}$
15:       $\mathbf{a}^{(k_t)} = \mathbf{a}^{(k_{t-1})}$
16:    **end if**
17:    Update $\sigma_e^2 \sim \mathcal{IG}(\alpha_n, \beta_n)$ via Gibbs sampling
18: **end procedure**

Algorithm 3 Life Move

 1: **procedure** Life Move for AR Model Order Selection with RJMCMC
 2:     Given data $\mathbf{y}$
 3:     $k_t, N_{iter}, a^{(k_0)}, \alpha_e, \beta_e$ values are initialized at $t = 0$
 4:     RJMCMC will be performed for $t = 1, \ldots, N_{iter}$
 5:     Assume a life move proposed at iteration t
 6:     $k' \leftarrow k_{t-1}$
 7:     Propose $\mathbf{a}^{(k')}$ from $q(\mathbf{a}^{(k')}|k, \mathbf{a}^{(k_{t-1})})$
 8:     Calculate α_{life}
 9:     Propose $num \sim Unif(0, 1)$
10:     **if** $num \leq \alpha_{\text{life}}$ **then**
11:         $k_t = k'$
12:         $\mathbf{a}^{(k_t)} = \mathbf{a}^{(k')}$
13:     **else**
14:         $k_t = k_{t-1}$
15:         $\mathbf{a}^{(k_t)} = \mathbf{a}^{(k_{t-1})}$
16:     **end if**
17:     Update $\sigma_e^2 \sim \mathcal{IG}(\alpha_n, \beta_n)$ via Gibbs sampling
18: **end procedure**

2.6.2.2. Life Move

Life move uses an MH application to update the AR coefficients with probability P_{life}. We apply MH algorithm because there is no dimension change for life move. Life move is shown in Algorithm 3 below.

Acceptance ratio is defined as $\alpha_{\text{life}} = \min\{1, r\}$. The corresponding value of r is given below

$$r = \frac{f(\mathbf{y}|k', \mathbf{a}^{(\mathbf{k}')}, \sigma_e^2)}{f(\mathbf{y}|k_{t-1}, \mathbf{a}^{(\mathbf{k_{t-1}})}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(k')}|k')}{f(\mathbf{a}^{(k_{t-1})}|k_{t-1})} \times \frac{q(\mathbf{a}^{(k_{t-1})}|k', \mathbf{a}^{(k')})}{q(\mathbf{a}^{(k')}|k, \mathbf{a}^{(k_{t-1})})} \tag{2.65}$$

where $f(\mathbf{y}|k', \mathbf{a}^{(k')}, \sigma_e^2)$ and $f(\mathbf{y}|k_{t-1}, \mathbf{a}^{(k_{t-1})}, \sigma_e^2)$ are likelihood distributions and $f(\mathbf{a}^{(k')}|k')$ and $f(\mathbf{a}^{(k_{t-1})}|k_{t-1})$ are prior distributions for parameter vector. Updated values of parameters are proposed from the distribution $q(\mathbf{a}^{(k')}|k, \mathbf{a}^{(k_{t-1})})$ which is defined below

$$\mathbf{a}^{(k')} \sim q(\mathbf{a}^{(k')}|k, \mathbf{a}^{(k_{t-1})}) = \mathcal{N}(\mathbf{a}^{(k')}|\boldsymbol{\mu}_n, \sigma_e^2 \boldsymbol{\Sigma}_n^{-1}) \tag{2.66}$$

where $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ are

$$\boldsymbol{\mu}_n = \mathbf{a}^{(k_{t-1})} \qquad \text{and} \qquad \boldsymbol{\Sigma}_n = \mathbf{Y}^T \mathbf{Y} + \frac{1}{\sigma_a^2}\mathbf{I}_{k'} \tag{2.67}$$

where $\mathbf{Y}$ represents the matrix consisting of the past values of data which is formed as in (2.50) (Troughton and Godsill, 1998).

2.6.2.3. Updating Excitation variance σ_e^2

Independently from defined reversible moves, excitation variance σ_e^2 is updated at each iteration t by using all other parameters which are sampled at iteration $t - 1$. The updating mechanism follows Gibbs sampling methodology. The posterior distribution for σ_e^2 is calculated as

$$f(\sigma_e^2|\mathbf{y}, k, \mathbf{a}^{(k)}, \sigma_a^2) \propto f(\mathbf{y}|k, \mathbf{a}^{(k)}, \sigma_a^2, \sigma_e^2)f(\sigma_e^2). \tag{2.68}$$

From (2.53) we can replace approximate likelihood in a way that

$$f(\sigma_e^2|\mathbf{y}, k, \mathbf{a}^{(k)}, \sigma_a^2) \approx \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \sigma_e^2\mathbf{I}_n)\mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e) \tag{2.69}$$

$$\propto \frac{1}{\sqrt{(2\pi\sigma_e^2)^n}} \exp\left(-\frac{1}{\sigma_e^2}\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}\right)\sigma_e^{-2(\alpha_e+1)} \exp\left(-\frac{\beta_e}{\sigma_e^2}\right) \tag{2.70}$$

$$= \mathcal{IG}(\sigma_e^2|\alpha_n, \beta_n) \tag{2.71}$$

where α_n and β_n are

$$\alpha_n = \alpha_e + \frac{1}{2}n \qquad \text{and} \qquad \beta_n = \beta_e + \frac{1}{2}\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}. \tag{2.72}$$

Excitation variance σ_e^2 will be sampled directly at each iteration by using the Inverse Gamma distribution as in (2.71).

2.6.3. Defining the Proposals

As stated above, (Green, 1995) realized that to provide the dimension matching we should introduce $dim(k') - dim(k)$ auxiliary variable u. After proposing these variables we should define a deterministic function w that maps state x with dimension k to state x' with dimension k'.

Suppose that we propose a birth move from dimension k to $k' = k + 1$. Because we have one difference between two spaces, a scalar u is proposed from distributions $q(u)$ which is defined as

$$u \sim q(u) = \mathcal{N}(0, \sigma_a^2) \tag{2.73}$$

where σ_a^2 is known at the beginning of the algorithm. We choose this proposal ratio in order to obtain a scenario at which every birth move creates a new AR coefficient which is independent of the other coefficients. This gives us the opportunity that Jacobian is equal to unity. According to the Bayesian data analysis strategy, the posterior distribution is Gaussian as far as the conjugate prior is Gaussian given the data (Gelman et al., 2003). We select a Gaussian distribution for proposal distribution of newly born parameters in order to satisfy this fact.

In light of this information about proposal ratio, each element of the coefficient vector is defined as shown below

$$\mathbf{a}^{(k)} = [a_1, a_2, \ldots, a_k]_{1 \times k} \tag{2.74}$$

$$\mathbf{a}^{(k')} = [a'_1, a'_2, \ldots, a'_{k'}]_{1 \times (k+1)}. \tag{2.75}$$

Using the parameters $a^{(k)}$ and u we define the mapping function w as

$$\mathbf{a}^{(k')} = w(\mathbf{a}^{(k)}, u) \tag{2.76}$$

$$a'_1 = a_1, \quad a'_2 = a_2, \quad \ldots \quad , a'_{k'-1} = a_k, \quad a'_{k'} = u. \tag{2.77}$$

Computing the Jacobian value by using the expressions in (2.77) gives the result below

$$
|\mathcal{J}| = \begin{vmatrix} \dfrac{\partial a'_1}{\partial a_1} & \dfrac{\partial a'_2}{\partial a_1} & \cdots & \dfrac{\partial a'_{k'}}{\partial a_1} \\[2mm] \dfrac{\partial a'_1}{\partial a_2} & \dfrac{\partial a'_2}{\partial a_2} & \cdots & \dfrac{\partial a'_{k'}}{\partial a_2} \\[2mm] \vdots & \vdots & \ddots & \vdots \\[2mm] \dfrac{\partial a'_1}{\partial u} & \dfrac{\partial a'_2}{\partial u} & \cdots & \dfrac{\partial a'_{k'}}{\partial u} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{vmatrix} = det(\mathbf{I}_{k'}) = 1 \qquad (2.78)
$$

where $\mathbf{I}_{k'}$ is the identity matrix with size $k' \times k'$.

2.6.4. Simulation Setup

Initial values for excitation variance prior distribution is selected as $\alpha_e = 1$ and $\beta_e = 2$. The initial value for model order, $k^{(0)}$ is set to 1. $\mathbf{a}^{k^{(0)}}$ is sampled from the prior distribution which is defined in (2.57). Auxiliary variable u is sampled according to the procedure in Section 2.6.3.

Move probabilities are selected as 0.15, 0.15 and 0.7 respectively for P_{birth}, P_{death} and P_{life}. 20000 iterations are simulated to let sampled parameters converge. No burn-in period is used in results in order to see all the characteristics of the algorithms. Success of the results can be increased by truncating first N_b values as burn-in period.

Two different AR(k) data are used in the simulations with orders 4 and 10. Two data sets have the same length of 10000 (data with length 15000 is created and first 5000 samples are truncated in order to provide the AR characteristics of the data. The remaining 10000 samples are used as the data set). The coefficient vectors of AR processes are $\mathbf{a}^{(4)} = $ [0.8, -0.20, 0.45, -0.55] and $\mathbf{a}^{(10)} = $ [0.9402, -0.43, 0.41679, -0.4969, 0.4771, -0.5010, 0.0505, -0.2357, 0.4024, -0.1549] both are driven with a Gaussian excitation sequence with variance of $\sigma_e^2 = 1$. Variance value of AR coefficients' prior distribution is selected as $\sigma_a^2 = 0.1$ and remains fixed at each iteration. No sampling is applied for σ_a^2.

Performance of RJMCMC in AR model estimation has been studied under three simulations. The first simulation refers to the general scenario which performs model estimation using the procedures explained in the previous sections. Figures to evaluate the estimation performance of model order and model coefficients are shown in this simulation.

Other two simulations have been studied in order to test performance of RJMCMC under different conditions. Simulation 2 studies the effect of initial model order value, $k^{(0)}$, on overall estimation procedure where the rest of the parameters are the same as explained in the previous sections. Simulation 3 studies the effect of different proposal distributions to sample auxiliary variables, u. In both of the simulations, several figures are shown to provide distinguished results according to changes in some RJMCMC parameters.

2.6.5. Simulation 1: General Simulation

In simulation 1, general simulation scenario which is explained in detail in Section 2.6 has been studied. There are two example AR models and RJMCMC has been run to estimate the model order as well as the model coefficients and excitation variance.

In Figures 2.4 and 2.5, estimation performance of RJMCMC for AR(4) model is depicted. In particular, in Figure 2.4, instantaneous estimate and the resulting posterior for the model order, k is shown. Examining the figures, RJMCMC estimates correct model order with a percentage of 90% and converges to the correct order after nearly 500 iterations. Moreover, in Figure 2.5 running mean plots in a single RJMCMC run for the three out of four model coefficients and the excitation sequence, σ_e^2 (At iteration i, the averages of the estimated parameters between iteration 1 and i have been calculated and recorded. Resulting averages represent "running means" and this term will be used for this operation for the rest of the text) is shown. Analyzing this figure shows that for all 4 parameters RJMCMC converges to correct value after 5000 iterations.

In Figures 2.6 and 2.7 the same analysis has been performed for the second example AR model, AR(10). The performance of the proposed method is very similar to that of AR(4) model. RJMCMC estimates correct model order in nearly 90% of the iterations and converges to correct order after 1000 iterations. For model coefficient estimation (Figure 2.7), it takes nearly 5000 iterations to converge to the correct value of the parameter.

2.6.6. Simulation 2: Different Initial Model Orders

In Simulation 2, the effect of the initial model order, $k^{(0)}$ on model estimation performance has been studied. Four different initial orders have been selected for both of the example AR models and model order and coefficient estimation performance of RJMCMC have been tested.

(a) Instantaneous Order Estimates

(b) Histogram for Model Order, k
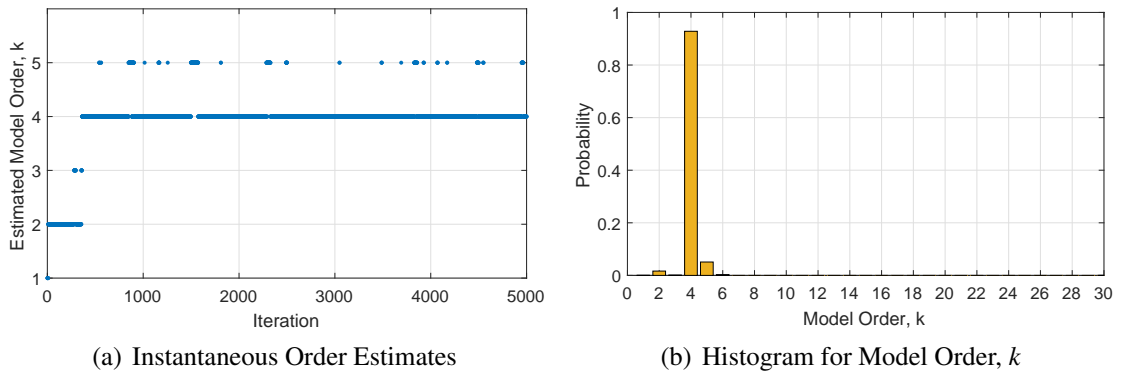
Figure 2.4. Model order estimation for AR(4)
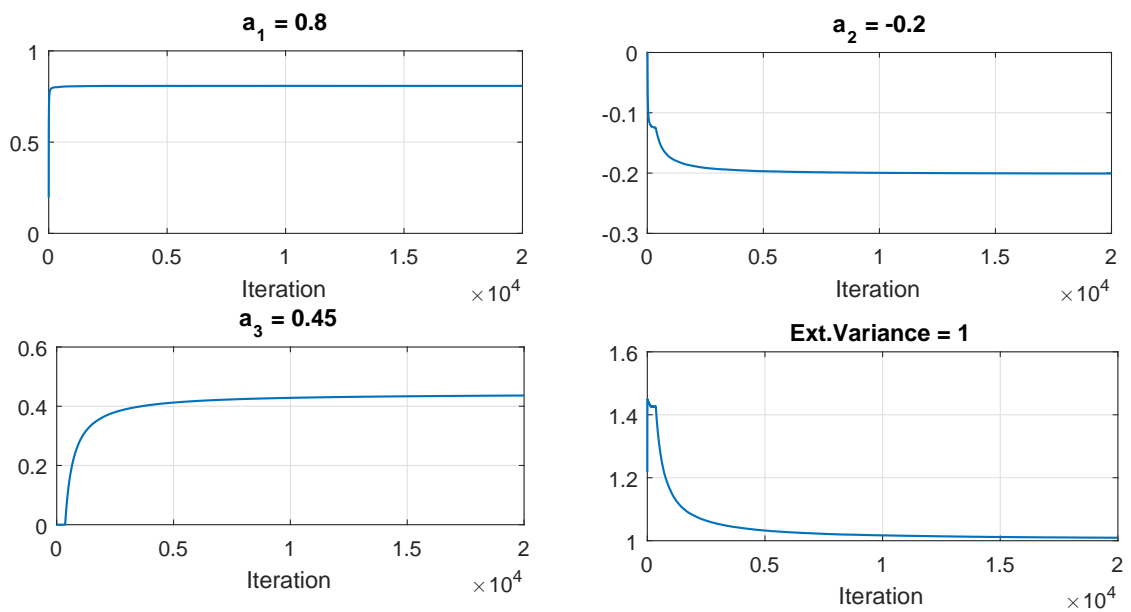
Figure 2.5. Running mean plots for AR(4) model coefficients and excitation variance

(a) Instantaneous Order Estimates　　　　(b) Histogram for Model Order, k
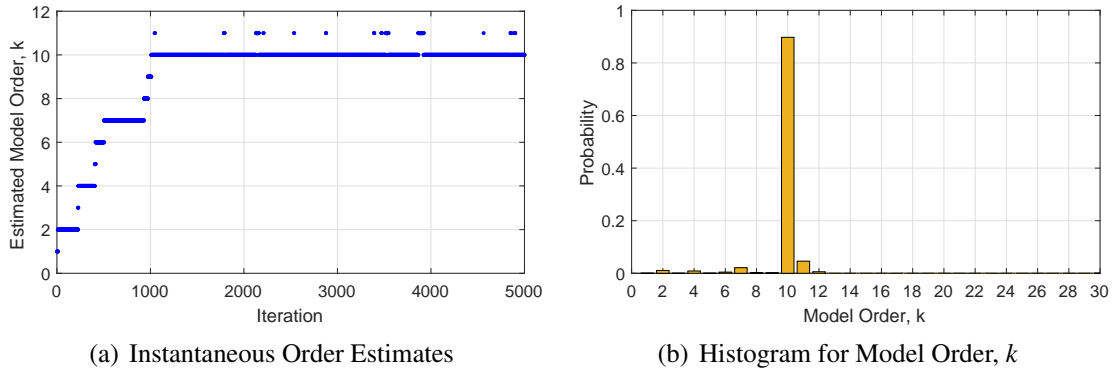
Figure 2.6. Model order estimation for AR(10)
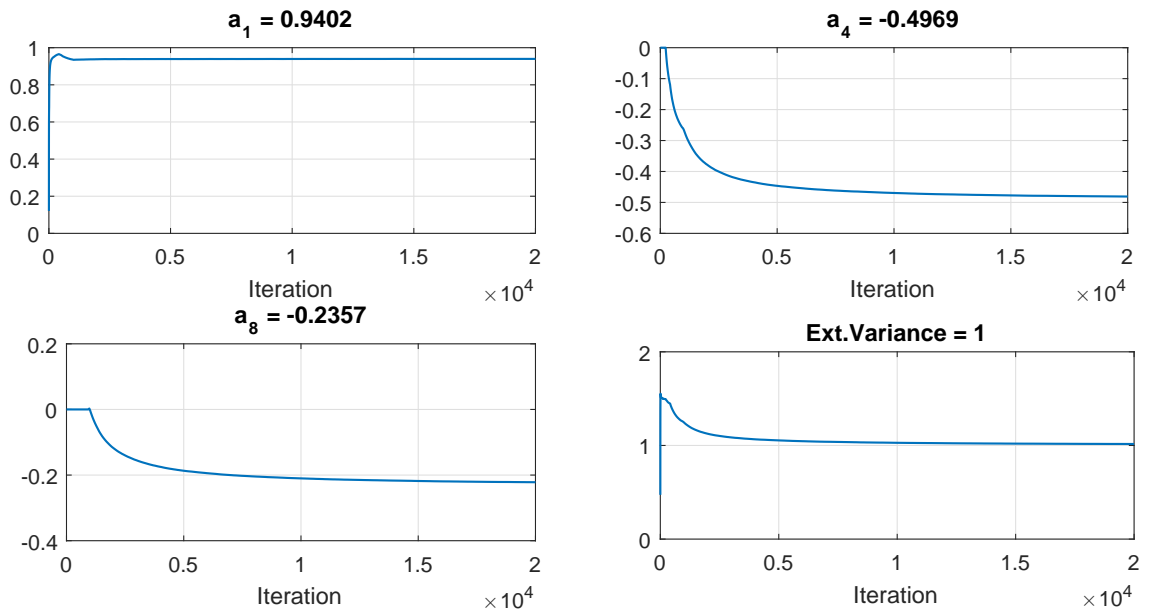
Figure 2.7. Running mean plots for AR(10) model coefficients and excitation variance

In Figures 2.8 and 2.9, instantaneous estimates for the model order are depicted for each example AR model. Initial model orders for AR(4) and AR(10) is $[1, 5, 10, 15]$ and $[1, 7, 13, 20]$, respectively. For all the initial order and AR model, RJMCMC converges to the correct value with a change on convergence time. In all the cases, it takes nearly 500 iterations to converge.

In Figures 2.10 and 2.11, running mean plots for the first model coefficient from each AR model are shown. RJMCMC converges to the correct model coefficient for all initial order selection after 2500-5000 iterations.

2.6.7. Simulation 3: Different Proposals

For the third simulation, RJMCMC performance under different proposal distributions for the auxiliary variable, u. There are three proposal distributions which are $q_1(u), q_2(u)$ and $q_3(u)$ each of which has different characteristics.

The first proposal distribution, $q_1(u)$ is selected as the same as the one defined in Section 2.6.3. Assuming that the same conditions in Section 2.6.3 are valid, for the second proposal distribution the variable u is proposed from distributions $q_2(u)$ which is defined as

$$u \sim q_2(u) = \mathcal{U}(-1, 1). \tag{2.79}$$

As long as the proposed coefficient u is independent from the recent coefficients, Jacobian can be easily computed as $|\mathcal{J}_2| = 1$.

For the third proposal distribution, $q_3(u)$, a new mapping function w_3 has been utilized. A selection of this type creates newly born coefficients to be dependent on recent coefficients and causes Jacobian of the space change is not equal to 1.

$$\mathbf{a}^{(k')} = w(\mathbf{a}^{(k)}, u) \tag{2.80}$$

$$a'_1 = a_1, \quad a'_2 = a_2, \quad \dots, a'_{k'-1} = a_k, \quad a'_{k'} = u \frac{1}{k} \sum_{i=1}^{k} a_i \tag{2.81}$$
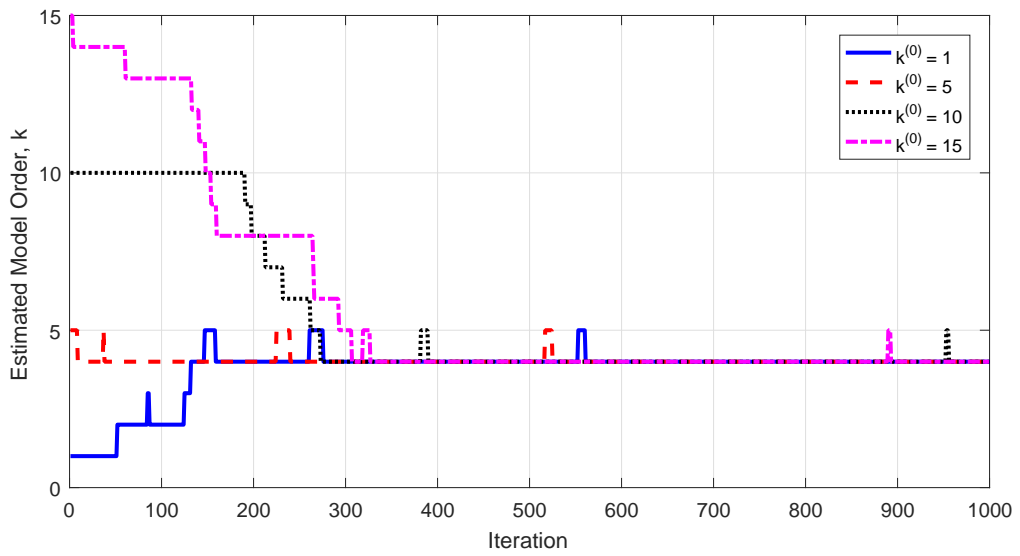
Figure 2.8. Model order estimation for different initial model orders - AR(4)
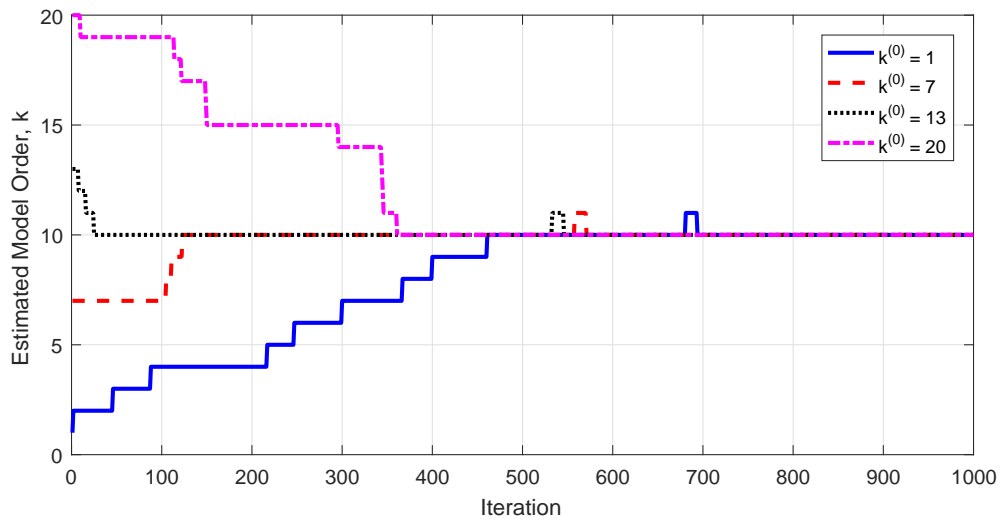
Figure 2.9. Model order estimation for different initial model orders - AR(10)
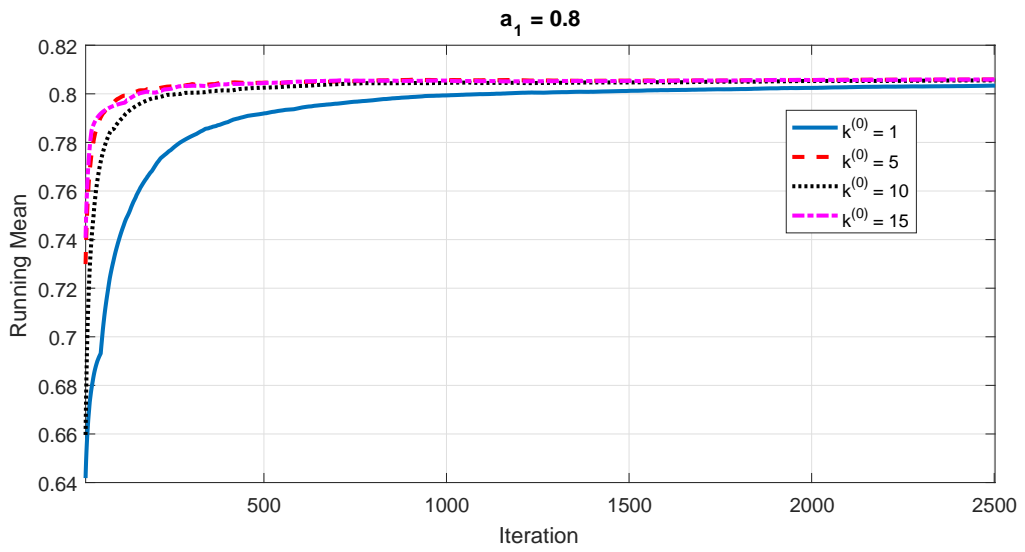
Figure 2.10. Running mean plots for model coefficient, a_1 for different initial model orders - AR(4)
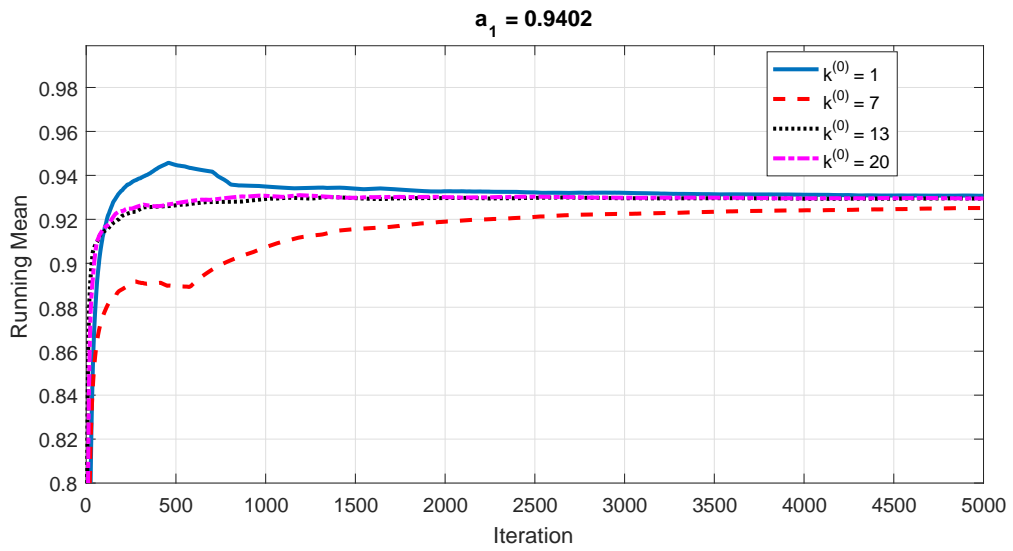
Figure 2.11. Running mean plots for model coefficient, a_1 for different initial model orders - AR(10)

where the auxiliary variable u is sampled from $q_3(u)$ which is defined as

$$u \sim q_3(u) = \mathcal{U}(-2, 2). \tag{2.82}$$

Computing the Jacobian by using the expressions in (2.81) gives

$$|\mathcal{J}_3| = \begin{vmatrix} \dfrac{\partial a'_1}{\partial a_1} & \dfrac{\partial a'_2}{\partial a_1} & \cdots & \dfrac{\partial a'_{k'}}{\partial a_1} \\ \dfrac{\partial a'_1}{\partial a_2} & \dfrac{\partial a'_2}{\partial a_2} & \cdots & \dfrac{\partial a'_{k'}}{\partial a_2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial a'_1}{\partial u} & \dfrac{\partial a'_2}{\partial u} & \cdots & \dfrac{\partial a'_{k'}}{\partial u} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \cdots & \dfrac{u}{k} \\ 0 & 1 & \cdots & \dfrac{u}{k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \dfrac{1}{k}\sum_{i=1}^{k} a_i \end{vmatrix} = \left| \dfrac{1}{k}\sum_{i=1}^{k} a_i \right|. \tag{2.83}$$

In Figures 2.12 and 2.13, instantaneous estimates for 3 different proposal distributions are shown for AR(4) and AR(10), respectively. Examining both of the figures shows that RJMCMC converge to correct model order value under different conditions. Among example proposal distributions q_1 provides faster convergence than the others. In all the cases, RJMCMC converges to correct value nearly after 1250 iterations for AR(10) model (after 500 iterations for AR(4) model), and difference of the proposal distributions becomes indistinguishable in the figures.

2.7. Conclusions

In this chapter, a previously studied toy example, which is based on the AR model estimation via RJMCMC has been implemented. We have created 3 simulation scenarios under different conditions and obtained the following results:

- Simulation 1: The general framework of the previous studies about AR model estimation has been sketched in this simulation. RJMCMC has been used to estimate the correct AR model order given data as well as its coefficients and the excitation sequence providing that the excitation sequence is normally distributed.

- Simulation 2: The effect of the initial model order has been tested in this simulation. By the help of the Markov chain properties of RJMCMC, it would converge to the correct model order even if it started from a state which is very far from the correct
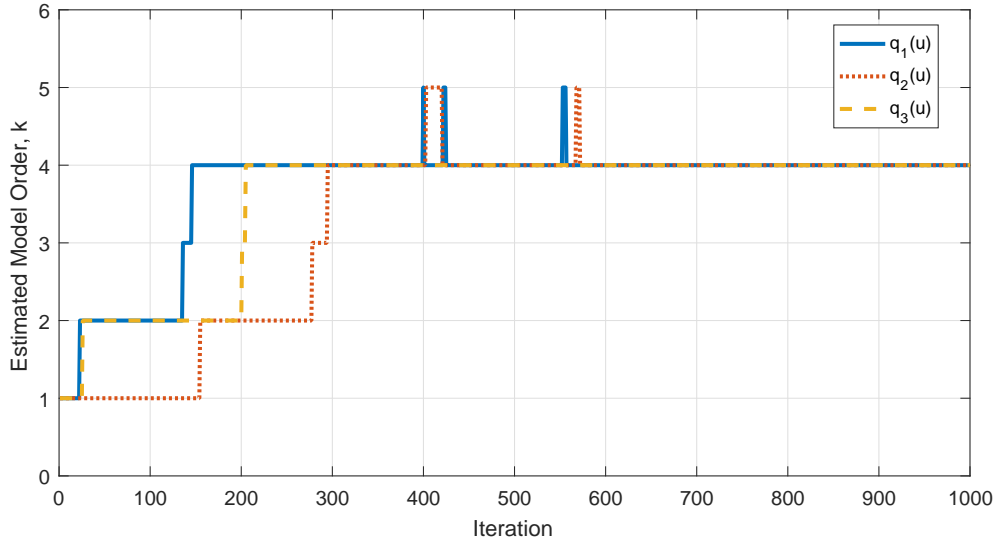
Figure 2.12. Model order estimation for different proposals - AR(4)
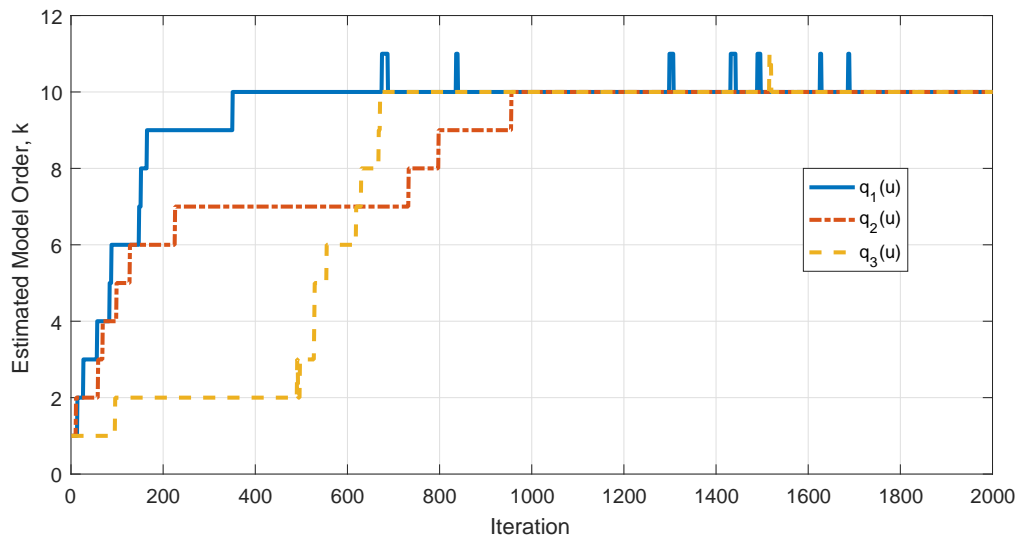
Figure 2.13. Model order estimation for different proposals - AR(10)

value.

- Simulation 3: Here using three different proposal distributions clearly shows us a very important fact about the key advantage of the RJMCMC algorithm, that is, the algorithm converges to the true model order value if the Markov chain is run long enough, even if we propose newly born coefficients from very irrelevant distributions.

This toy example and other studies in the literature have clearly demonstrated that the RJMCMC algorithm is an effective method that can be used to solve model uncertainty. Following these examples, a simple but very important question might come to mind: *Can RJMCMC be used for more complex model transitions and therefore more complex models?* This thesis will draw a framework that is *beyond trans-dimensional* for the RJMCMC algorithm and will present an approach which will be the answer to this question.

2.8. Beyond Trans-dimensional RJMCMC

In spite of RJMCMC's use in trans-dimensional cases, the original formulation in (Green, 1995) holds a wider interpretation than just sampling between spaces of different dimensions. In the beyond trans-dimensional RJMCMC point of view, the main requirements of RJMCMC stated by Green are still valid with one exception, that is, a change in parameter space definition.

In the original formulation, Green firstly derives the condition for the satisfaction of detailed balance requirements in terms of the Borel sets which the candidate models belong to. In the continuation of the derivation, he specializes his discussion to moves between spaces which differ only in dimensions and the general discussion is abandoned. However, the parameter vectors in (2.30) may belong to Borel sets which differ not only in their dimensions but also in the generic models they belong to. Thus, the algorithm can be used for much more generic implementations.

Notwithstanding, this general interpretation should be taken with caution to have a useful method. Particularly, the Borel sets should be *related* somehow, which can be conveniently set by *matching a common property (i.e. norm)* in defining the spaces. Defining proposals in this way will provide sampling more efficient candidates and help algorithm to converge faster. As an example, model transitions can be designed to provide fixed first ordered moments between spaces. Thus, this moment based approach provides a more

efficient way to explore all the candidate models within the combined space. Carrying the trained information to a new generic model space is very crucial in this framework. Otherwise, the algorithm would start to train from scratch repeatedly each time it changes states and sampling across unrelated spaces would not give us a computational advantage. In that case, one could solve for different spaces separately and compare the final results to choose the best model. Two examples one can think of firstly, are:

1) x might correspond to a linear parametric model such as AR while x' might correspond to a nonlinear model such as Volterra AR.

2) x might correspond to a pdf p_A with certain distribution parameters while x' might correspond to another pdf p_B with some other distribution parameters.

To this end, we define a combined parameter space $\varphi = \bigcup_k \varphi_k$ including more than one subspace. Assume that a move M from Markov chain state $x \in \varphi_1$ to $x' \in \varphi_2$ is defined and Borel sets $A \subset \varphi_1$ and $B \subset \varphi_2$ are related with a set of functions each of which are invertible. Particularly, for any Borel sets in both of the spaces, φ_1 and φ_2, functions $h_{12} : A \mapsto B$ and $h_{21} : B \mapsto A$ can be defined by matching a common property of the spaces. For generality, if the proposed move requires matching the dimensions, auxiliary variables $\mathbf{u}_1$ and $\mathbf{u}_2$ can be drawn from proper densities $Q_1(\cdot)$ and $Q_2(\cdot)$, respectively. Otherwise, one can set $\mathbf{u}_1$ and $\mathbf{u}_2$ to $\emptyset$. Please note that the dimensions of the parameter spaces at both sides of the transitions can be different or the same and reversible jump mechanism is still applicable.

Consequently, although the candidate spaces are of different classes, since the Borel sets are defined to be related, the assumption of Green which is given in (2.27) still holds for a symmetric measure ξ_m and densities for joint proposal distributions, $\pi(\cdot)q(\cdot, \cdot)$, can be defined with respect to this symmetric measure by satisfying the equilibrium in (2.30). Thus, the acceptance ratio can be written as

$$A(x \rightarrow x') = \min\left\{1, \frac{\pi(x')p_{M^R}Q_2(\mathbf{u}_2)}{\pi(x)p_M Q_1(\mathbf{u}_1)} \left|\frac{\partial h_{12}(\boldsymbol{\theta}_1, \mathbf{u}_1)}{\partial(\boldsymbol{\theta}_1, \mathbf{u}_1)}\right|\right\} \tag{2.84}$$

where M^R is the reverse move of M and p_M and p_{M^R} represent the probabilities of the moves. The Jacobian term appears in the equation as a result of the change of variables operation between spaces.

It is straightforward to show that the acceptance ratios in (2.44) and (2.84) are of the same kind. Hence, we add a few concluding remarks.

Remark 2.1 *The general usage of RJMCMC which is unfolded in this section makes it possible to utilize the same formulation and the same trans-dimensional framework (birth-death move pairs) in the model estimation studies of nonlinear models (i.e. Volterra series expansion based linear-in-the-parameters models). Thus, RJMCMC can explore (or jump between) the spaces of linear and nonlinear models.*

Remark 2.2 *This new utilization of RJMCMC will be named as* trans-space *rather than trans-dimensional. Trans-space RJMCMC reveals a general framework for exploring the spaces of different generic models whether or not their parameter spaces are of different dimensionality. Consequently, the trans-dimensional case is a subset of trans-space transitions.*

Remark 2.3 *Trans-space RJMCMC requires defining new types of moves due to the need for more detailed operations than just being birth and death of the parameters. These moves will be named as between-space moves and may include both birth and death of the parameters at the same time or a mapping between the parameter spaces by using their common properties. Switch move will be proposed as a between-space move, which performs a switching between the candidate spaces of the generic model classes.*

Remark 2.4 *Our general interpretation includes both the trans-dimensional RJMCMC and the MH algorithm. When searching a single parameter subspace, the trans-space approach will be simplified into an MH algorithm. Trans-dimensional transitions between the same classes are also included in this approach as well as the transitions between different generic classes all of which may have the same dimensions or not. To be more clear about the contributions of the proposed approach, in Figure 2.14, a comparison is shown visually between trans-space RJMCMC approach and MH and trans-dimensional RJMCMC. Briefly, trans-space procedure provides an approach which can be defined as beyond trans-dimensional. Transitions between different generic classes are irrespective of their dimensions of the parameter spaces. Trans-space RJMCMC reveals a generalization and unveils the great potential of the same formulation in (Green, 1995).*
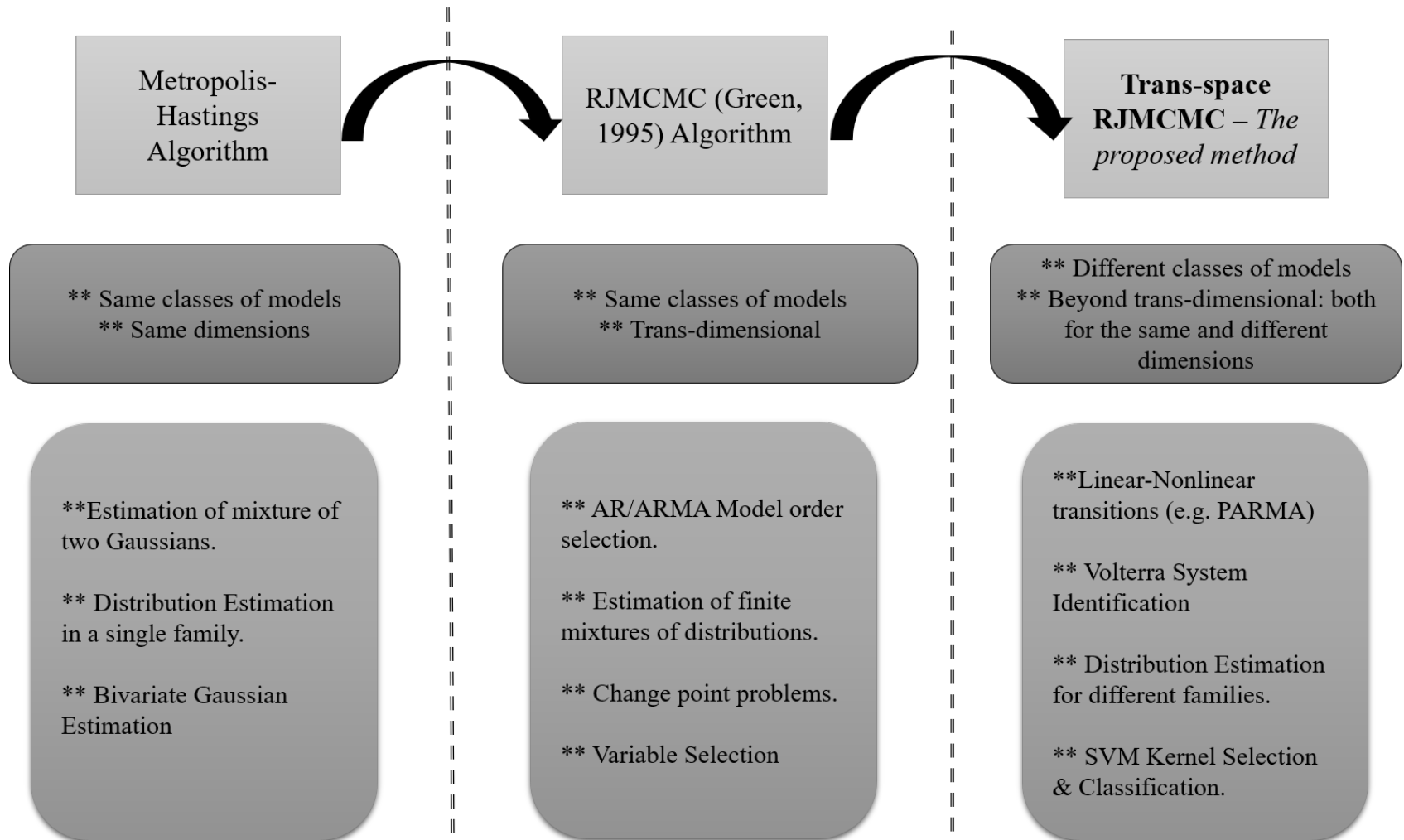
Figure 2.14. Comparison for trans-space RJMCMC, MH and RJMCMC.

CHAPTER 3

NONLINEAR (POLYNOMIAL) MODEL ESTIMATION

Due to their ease of use, linear process models have been generally used in modelling various real life phenomena. In the first modelling studies, using linear time series methods was common and there was no-need to nonlinear ones since the infinite moving average can be approximated with autoregressive processes with low orders (Potter, 1999). Nonlinear time series analysis have begun to get attention in the early 1980s, since linear time series models are unable to model many real life data sets such as asymmetric business cycles, volatility of stock markets, etc. (Tong, 2002).

There are various nonlinear time series models which have been utilized instead of their linear counterparts. One of the nonlinear time series models is the *Gaussian autoregressive conditional heteroscedasticity* (GARCH) model. The general form of a GARCH model with order k is given as (Tong, 2002)

$$X(t) = \varepsilon(t)\sigma(t) \tag{3.1}$$

where $\varepsilon(t)$ is normally distributed and $\sigma(t) = \sqrt{\alpha_0 + \sum_{i=1}^{k} \alpha_i X^2(t-i)}$. This model has been used especially in the context of econometrics and finance in order to express the changes (increase or decrease) of stocks, etc. over time.

Threshold autoregressive (TAR) model is another important example to nonlinear time series methods. The general expression of TAR models can be expressed as (Tong, 2002)

$$X(t) = \alpha_0^{(j)} + \sum_{i=1}^{k} \alpha_j^{(i)} X(t-i) + h_1^{(j)} \varepsilon(t). \tag{3.2}$$

TAR models are called as *self-excited* TAR (SETA) models when the conditional mean of $X(t)$ is piecewise linear and the conditional variance of $X(t)$ is piecewise constant (Tong, 2002). SETA models have been favourable, especially when the data sets show some periodic characteristics (Tsay, 1989; Tong, 2002). Additionally, *random coefficient au-*

toregressive (RICA) models (Robinson, 1978; Tong, 2002)

$$X(t) = \sum_{i=1}^{k} (\eta_i + B_i(t))X(t-i) + \varepsilon(t), \tag{3.3}$$

and *product* autoregressive models (Mckenzie, 1982; Tong, 2002)

$$X(t) = \varepsilon(t)X(t-1)^{\eta}, \tag{3.4}$$

can be given as examples of nonlinear time series models.

However, most of those models do not provide a solution to estimating the model parameters easily and their potential use is very limited. Since *autoregressive moving average* (ARMA) based nonlinear models including polynomial nonlinearity (or namely *polynomial* ARMA (PARMA) models) are *linear-in-the-parameters*, they differ from the other nonlinear time series models and thus employing various mathematical applications developed for linear models could be easier (Kuruoğlu, 2002). PARMA models also include polynomial autoregressive, namely PAR, and polynomial moving average, namely PMA models.

In the literature, general practice is to employ PARMA models in modelling real world problems by assuming a fixed degree of nonlinearity. It is desirable to estimate also the nonlinearity degree. However, estimating the nonlinearity degree may impose a heavy burden of algebraic operations and hence, providing a simpler approach to have an idea about the nonlinearity degree of the aforementioned models is of utmost importance.

Following the toy example presented in the previous chapter, we have seen under different simulation cases that RJMCMC can be used to solve the linear time series models uncertainty. The main objective of this chapter is to determine whether RJMCMC performs model estimation studies for linear-in-the-parameters nonlinear models and whether it can be designed to switch between linear ARMA (or AR, or MA) and nonlinear PARMA (or PAR, or PMA) models.

The approach presented in this chapter for RJMCMC can be used to solve all model uncertainties, including nonlinearity degree, in PARMA based time series models exemplified above. Trans-dimensional moves used in the toy example will also be used in this chapter, and it will be clear that the original formulation of Green and the trans-dimensional approach can be used for much more generic transitions. The results will

be presented separately for PAR (Karakuş et al., 2015), PMA (Karakuş et al., 2016) and PARMA (Karakuş et al., 2017b) models in the chronological order in which the work is done during the thesis.

3.1. Models

3.1.1. Polynomial Autoregressive (PAR) Models

Autoregressive models (AR) are time series models in which the movement of a variable is described in terms of its past values. AR models have been successful especially in representing the power spectrum of the speech/audio signals (Ganapathy et al., 2009) and in prediction studies of wind (Palomares-Salas et al., 2009), econometrics (Liew et al., 2003), etc. However, when the observed data exhibits nonlinear characteristics, performance of AR modelling has shown a significant degradation. PAR models are based on the Volterra series expansion as expressed by (3.5) which has had great success in modelling many real life phenomena such as optical channel modelling (Mhatli et al., 2015), loudspeaker system identification (Ji and Gan, 2012), short term wind speed prediction (Lee, 2011), brain signals (Lahaye et al., 2003), seismology (Bekleric, 2008) and communications (Fernandes et al., 2010).

PAR models can be represented as

$$x(l) = \mu + \sum_i^k a_i^{(1)} x(l-i) + \sum_i^k \sum_j^k a_{i,j}^{(2)} x(l-i)x(l-j) + \sum_{i,\dots}^{k,\dots} a_{i,\dots}^{(p)} x(l-i)\dots + e(l) \quad (3.5)$$

where $e(l)$ refers to the excitation sequence with distribution $\mathcal{N}(0, \sigma_e^2)$. PAR model coefficients for first order, second order and p^{th} order polynomials are represented by $a_i^{(1)}$, $a_{i,j}^{(2)}$ and $a_{i,\dots}^{(p)}$, respectively. A PAR model can be represented in the notation: $\text{P}^{(p)}\text{AR(k)}$ with the nonlinearity degree, p, and the AR memory, k.

Since PAR models are *linear-in-the-parameters*, a $\text{P}^{(p)}\text{AR}(k)$ model can be represented in matrix-vector form

$$\mathbf{x} = \mu + \mathbf{X}\mathbf{a}^{(p,k)} + \mathbf{e}, \quad \text{where} \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2\mathbf{I}_n), \quad (3.6)$$

where $\mathbf{x}$ is an n-vector of data samples, $\mathbf{X}$ is an $n \times \eta$ matrix of past samples and polynomial products of the data, $\mathbf{a}^{(p,k)}$ is an η-vector of $\mathrm{P}^{(p)}\mathrm{AR}(k)$ model coefficients, $\eta = \dfrac{(p+k)!}{p!\,k!} - 1$ refers to number of model coefficients and $\mathbf{e}$ is an n-vector of excitation sequence

$$\mathbf{x} = [x(1), x(2), \ldots, x(n)]^{T} , \tag{3.7}$$

$$\mathbf{e} = [e(1), e(2), \ldots, e(n)]^{T} , \tag{3.8}$$

$$\mathbf{X} = \begin{bmatrix} x(0) & \ldots & x(1-k) & x^2(0) & x(0)x(-1) & \ldots & x^p(1-k) \\ x(1) & \ldots & x(2-k) & x^2(1) & x(0)x(0) & \ldots & x^p(2-k) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x(n-1) & \ldots & x(n-k) & x^2(n-1) & x(n-1)x(n-2) & \ldots & x^p(n-1) \end{bmatrix} , \tag{3.9}$$

$$\mathbf{a}^{(p,k)} = \Big[a_1^{(1)}, \ldots, a_k^{(1)}, a_{1,1}^{(2)}, a_{1,2}^{(2)}, \ldots, a_{k,k}^{(2)}, \ldots, \underbrace{a_{k,k,\ldots,k}^{(p)}}_{p^{\text{TtH}}\text{ order}} \Big]^{T} . \tag{3.10}$$

In the last row of Table 3.1, the number of model coefficients, η, for different p and k pairs are shown.

3.1.2. Polynomial Moving Average (PMA) Models

Utilizing *nonlinear moving average* (NOMA) models in modelling real life applications has drawn interest after the study of Robinson in 1977 (Robinson, 1977). To name a few, NOMA based nonlinear models have been used in real life signals and systems, such as radio frequency power amplifiers (FR-PAS) (Isaksson et al., 2006), bridge aerodynamics (Kareem and Wu, 2014), finance (Zaffaroni and d'Italia, 2003) and adaptive control of the nonlinear systems (Chikkula and Lee, 2000). The main motivation to use NOMA based models in modelling is the case when the weighted sum of past values of errors (or shocks) is more important than the weighted sum of past values of data itself. Especially, in finance, in order to model volatility and exchange rates, considering the errors rather than or along with autoregression is very important (Zaffaroni and d'Italia, 2003).

Polynomial moving average (PMA) models, also represented by $\mathrm{P}^{(p)}\mathrm{MA}(q)$, are

linear-in-the-parameters Volterra based NOMA models and are expressed as

$$x(l) = \mu + \sum_{i=1}^{q} b_i^{(1)} e(l-i) + \sum_{i=1}^{q} \sum_{j=1}^{q} b_{i,j}^{(2)} e(l-i)e(l-j) + \sum_{i,\ldots}^{q,\ldots} b_{i,\ldots}^{(p)} e(l-i)\ldots + e(l), \quad (3.11)$$

where $b_i^{(1)}$, $b_{i,j}^{(2)}$ and $b_{i,\ldots}^{(p)}$ are PMA model coefficients for first order, second order and p^{th} order polynomials, respectively. MA memory of the PMA model is represented as q whereas p refers to the nonlinearity degree.

The matrix-vector representation of a $P^{(p)}MA(q)$ model is

$$\mathbf{x} = \mu + \mathbf{E}\mathbf{b}^{(p,q)} + \mathbf{e}, \quad (3.12)$$

where $\mathbf{E}$ is $n \times \gamma$ matrix, the components of which are the past samples and polynomial products of the excitation sequence and $\mathbf{b}^{(p,q)}$ refers to a vector of $P^{(p)}MA(q)$ model coefficients length of which can be calculated as $\gamma = \dfrac{(p+q)!}{p!\,q!} - 1$. $\mathbf{E}$ and $\mathbf{b}^{(p,q)}$ have the form

$$\mathbf{E} = \begin{bmatrix} e(0) & \ldots & e(1-q) & e^2(0) & e(0)e(-1) & \ldots & e^p(1-q) \\ e(1) & \ldots & e(2-q) & e^2(1) & e(1)e(0) & \ldots & e^p(2-q) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ e(n-1) & \ldots & e(n-q) & e^2(n-1) & e(n-1)e(n-2) & \ldots & e^p(n-q) \end{bmatrix}, \quad (3.13)$$

$$\mathbf{b}^{(p,q)} = \left[b_1^{(1)}, \ldots, b_q^{(1)}, b_{1,1}^{(2)}, b_{1,2}^{(2)}, \ldots, b_{q,q}^{(2)}, \ldots, \underbrace{b_{q,q,\ldots,q}^{(p)}}_{p^{th}\ \text{order}} \right]^T . \quad (3.14)$$

3.1.3. Polynomial Autoregressive Moving Average (PARMA) Models

Due to their generality spanning both AR and MA models and their ease of estimation by methods such as Box-Jenkins (Box et al., 2011), ARMA models have been preferable in time series prediction studies. Their usage area is diverse covering fields such as speech (Peng et al., 2015; Ganapathy, 2015), seismology (Kozin, 1988), video (Zhong et al., 2003), image (Cadzow et al., 1993), etc. Moreover, ARMA based models have been utilized in energy and meteorological prediction studies of solar radiation (Ji

and Chee, 2011; Benmouiza and Cheknane, 2016), electricity demand (Ohtsuka et al., 2010; Pappas et al., 2010) and wind speed (Flores et al., 2012; Erdem and Shi, 2011).

Nonlinear ARMA (NARMA) or specifically, *polynomial ARMA* (PARMA) models appear as natural alternatives to the linear ones in modelling problems including nonlinearity. NARMA based methods have also been used in the literature, for example in prediction studies of electricity load (Connor et al., 1991), wind speed (Lydia et al., 2016), in modelling a nonlinear networked control system (Zhang and Wang, 2008) and in forecasting of financial returns (Chen et al., 2015).

PARMA models, also represented by $P^{(p)}ARMA(k,q)$, can be defined as:

$$x(l) = \mu + \sum_{d=1}^{p} \sum_{\tau_1=1}^{k} \cdots \sum_{\tau_d=\tau_{d-1}}^{k} a_{\tau_1,\ldots,\tau_d}^{(d)} \prod_{j=1}^{d} x(l-\tau_j) + e(l) + \sum_{g=1}^{p} \sum_{\tau_1=1}^{q} \cdots \sum_{\tau_g=\tau_{g-1}}^{q} b_{\tau_1,\ldots,\tau_g}^{(g)} \prod_{j=1}^{g} e(l-\tau_j),$$

(3.15)

The system of equations in (3.15) can be easily represented in matrix-vector form for a $P^{(p)}ARMA(k, q)$ model as

$$\mathbf{x} = \mu + \mathbf{X}\mathbf{a}^{(p,k)} + \mathbf{E}\mathbf{b}^{(p,q)} + \mathbf{e},$$

(3.16)

where $n \times \eta$ matrix $\mathbf{X}$, η-vector $\mathbf{a}^{(p,k)}$, $n \times \gamma$ matrix $\mathbf{E}$ and γ-vector $\mathbf{b}^{(p,q)}$ have the forms given in the Sections 3.1.1 and 3.1.2.

3.2. RJMCMC Nonlinear Model Estimation Procedure

As stated in the introduction part of this chapter, the classical trans-dimensional procedures for RJMCMC have been utilized in estimating PAR, PMA and PARMA models. However, in contrast to the studies in the literature, we consider a model space, which includes different structural models, i.e. linear and nonlinear. Sampling model orders (p, k or q) requires changing the dimension at each RJMCMC iteration. There are 4 types of moves for this problem:

- **birth** of a new parameter,

- **death** of an existing parameter,

- updating the model coefficients via MH algorithm (**life** move),

- updating variance parameters via Gibbs Sampling.

Each move has probabilities P_{birth}, P_{death} and P_{life} satisfying $P_{\text{birth}} + P_{\text{death}} + P_{\text{life}} = 1$.

      General procedure requires 2 (or 3 for PARMA models) cascaded RJMCMC stages for model order parameters at each iteration. At each stage, one of the model parameters (p, k or q) has been proposed while the others remain at their recent values. Then, the acceptance ratio is calculated and the decision is tested for this candidate model order. The procedure is depicted in Algorithm 4. The intercept, μ, for all three models is chosen as 0 for simplicity.

Algorithm 4 RJMCMC scheme for PAR/PMA/PARMA Model Estimation

1: Given data **x**
2: Initialize model orders and coefficients for iteration $t = 0$
3: **for do** $t = 1 : N_{iter}$
4:     **RJMCMC for the first model order.**
5:     Propose a move m with $P_{\text{birth}}, P_{\text{death}}, P_{\text{life}}$
6:     Propose a candidate model order for selected move m
7:     Sample candidate coefficients from proposal distributions
8:     Calculate acceptance ratio, A
9:     **if** $\rho <= A$ where $\rho = \mathcal{U}(0, 1)$ **then**
10:         Update first model order with proposed value
11:         Create a new model coefficient vector from existing and proposed coefficients.
12:         Update coefficient vector with a new coefficient vector
13:     **else**
14:         Model order and coefficients remain the same. Discard proposed values.
15:     **end if**
16:     **RJMCMC for the second model order.**
17:         Use recent values for model order 1 and the coefficient vector after first stage.
18:         Repeat the same procedure in lines between 5 and 15
    for the second model order.
19:     **RJMCMC for the third model order. (*Only valid for PARMA models*)**
20:         Use recent values for model order 1 and 2 and the coefficient vector after second stage.
21:         Repeat the same procedure in lines between 5 and 15
    for the third model order.
22:     Update variance parameters via Gibbs Sampling
23: **end for**
24: **end**

      In the case of a birth move, assume that it is required to propose $\lambda = 4$ candidate model coefficients from a proposal distribution, $\chi(\mathbf{u})$, in order to satisfy the dimension matching. This proposal distribution has the form

$$\chi(\mathbf{u}) = \prod_{i=1}^{\lambda} \mathcal{U}(-\delta, \delta) \quad \text{and} \quad \delta = \frac{\nu}{E[|\mathbf{x}|]} \tag{3.17}$$

where $E[|\mathbf{x}|]$ is the expectation of the absolute value of the observed data vector $\mathbf{x}$. The upper and lower bound, δ, is chosen to depend on the given data. Thus, different proposal distributions with distinct limits can be constructed due to the fact that the magnitudes of different data sets may vary in different ranges. So, utilizing an *adhoc* choice like this adds variety for the proposed candidates. Moreover, after some trial-error process, we see that using smaller δ values increases the number of accepted RJMCMC moves. Therefore, values for v are chosen as 0.1, 0.05 and 0.02 for PAR, PMA and PARMA model estimation studies, respectively. Furthermore, each of the newly proposed model coefficients is chosen to be independent from the recent coefficients which causes a unity Jacobian.

3.3. Likelihoods and Priors for Models

An approximation of the likelihood function of ARMA models has been provided by performing a recursive estimation procedure in the model itself for the unobserved excitation values in (Box et al., 2011). This approach has been used in model estimation studies for Bayesian analysis of ARMA based time series models (Troughton and Godsill, 1998; Ehlers and Brooks, 2004; Eğri et al., 2010).

On the other hand, in a previous study (Ivanov, 1987), it has been stated that outputs of Volterra based models follow Gaussian characteristics for white inputs and narrowband models. In addition, a Gaussian approximate likelihood has been used in restoration of nonlinearly distorted AR signals when the whole system follows a nonlinear AR model characteristics in (Troughton and Godsill, 2001). For PMA and PARMA model estimation studies, we have experimentally verified the Gaussianity by testing the output processes of the models.

Thus, an approximate likelihood function for a $\mathrm{P}^{(p)}\mathrm{ARMA}(k,q)$ has been used in this study as,

$$f(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi\sigma_e^2)^{(n-q_{\max})}}} \exp\left(\frac{-1}{2\sigma_e^2} \sum_{t=q_{\max}+1}^{n} e_t^2\right) \tag{3.18}$$

$$\approx \mathcal{N}(\mathbf{x} - \mathbf{X}\mathbf{a}^{(p,k)} - \mathbf{E}\mathbf{b}^{(p,q)}|\mathbf{0}, \sigma_e^2\mathbf{I}_{n-q_{\max}}). \tag{3.19}$$

It is straightforward to obtain likelihood functions for PAR and PMA models by using the same expression in (3.18) for PAR models as $q = 0$ and for PMA models as

$$k = 0$$

$$f(\mathbf{x}|\boldsymbol{\theta}) \approx \mathcal{N}(\mathbf{x} - \mathbf{Xa}^{(p,k)}|\mathbf{0}, \sigma_e^2 \mathbf{I}_n), \quad \text{for PAR models,} \tag{3.20}$$

$$f(\mathbf{x}|\boldsymbol{\theta}) \approx \mathcal{N}(\mathbf{x} - \mathbf{Eb}^{(p,q)}|\mathbf{0}, \sigma_e^2 \mathbf{I}_{n-q_{\max}}), \quad \text{for PMA models.} \tag{3.21}$$

In the previous RJMCMC studies, generally uniform priors are selected for model order parameters (Troughton and Godsill, 1998; Ehlers and Brooks, 2004; Eğri et al., 2010). Correspondingly, in the model estimation study in this chapter, we assume that each model is equiprobable and we choose the noninformative uniform priors for model orders p, k, and q with upper bounds $p_{\max}, k_{\max}$ and $q_{\max}$, respectively:

$$f(p) = \mathcal{U}(1, p_{\max}), \quad f(k) = \mathcal{U}(1, k_{\max}), \quad f(q) = \mathcal{U}(1, q_{\max}). \tag{3.22}$$

In order to provide conditional conjugacy, priors for model coefficients are assumed to be normally distributed. Besides, the excitation variance and variance of the model coefficients are assumed as inverse-Gamma distributed. These choices of priors for variances are based on the conditional conjugacy of the inverse-Gamma distribution. Then, the full posterior conditional distribution becomes also inverse-Gamma (Ehlers and Brooks, 2004). All the priors are given as

$$f(\mathbf{a}^{(p,k)}|p, k, \sigma_a^2) = \mathcal{N}(\mathbf{a}^{(p,k)}|\mathbf{0}, \sigma_a^2 \mathbf{I}_\eta), \tag{3.23}$$

$$f(\mathbf{b}^{(p,q)}|p, q, \sigma_b^2) = \mathcal{N}(\mathbf{b}^{(p,q)}|\mathbf{0}, \sigma_b^2 \mathbf{I}_\gamma), \tag{3.24}$$

$$f(\sigma_a^2) = \mathcal{IG}(\sigma_a^2|\alpha_a, \beta_a), \tag{3.25}$$

$$f(\sigma_b^2) = \mathcal{IG}(\sigma_b^2|\alpha_b, \beta_b), \tag{3.26}$$

$$f(\sigma_e^2) = \mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e), \tag{3.27}$$

where $\mathcal{IG}(\cdot)$ refers to an inverse Gamma distribution and $\alpha_a, \beta_a, \alpha_b, \beta_b, \alpha_e, \beta_e$, are hyperparameters for variances.

3.4. PAR Model Estimation

Bayesian hierarchy for the posterior density $f(\boldsymbol{\theta}|\mathbf{x})$ corresponding to parameter vector $\boldsymbol{\theta}$ consisting of the parameters $\{p, k, \mathbf{a}^{(p,k)}, \sigma_e^2, \sigma_a^2\}$ can be expressed as

$$f(p, k, \mathbf{a}^{(p,k)}, \sigma_e^2, \sigma_a^2|\mathbf{x}) \propto f(\mathbf{x}|p, k, \mathbf{a}^{(p,k)}, \sigma_e^2)f(\mathbf{a}^{(p,k)}|p, k, \sigma_a^2)f(\sigma_a^2)f(\sigma_e^2)f(k)f(p). \quad (3.28)$$

In case of a birth move (birth of a new variable) from AR memory parameter k to k' for $k' > k$ where the nonlinearity degree p is fixed, the acceptance ratio, $\alpha_{\text{birth}} = \min\{1, r_{\text{birth}}\}$, can be computed as

$$r_{\text{birth}} = \frac{f(\mathbf{x}|p, k', \mathbf{a}^{(p,k')}, \sigma_e^2)}{f(\mathbf{x}|p, k, \mathbf{a}^{(p,k)}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(p,k')}|p, k', \sigma_a^2)}{f(\mathbf{a}^{(p,k)}|p, k, \sigma_a^2)} \times \frac{P_{\text{death}}}{P_{\text{birth}}\chi(\mathbf{u})} \times \left|\frac{\partial \mathbf{a}^{(p,k')}}{\partial(\mathbf{a}^{(p,k)}, \mathbf{u})}\right|. \quad (3.29)$$

For a death move, which changes states from k to k' where $k' < k$, no new parameters are needed to be proposed. We remove the coefficients which belong to AR memory k of parameter vector $\mathbf{a}^{(p,k)}$.

The acceptance ratio of a death move, $\alpha_{\text{death}}(k \to k') = \min\{1, r_{\text{death}}\}$, is calculated directly via its reverse move $\alpha_{\text{birth}}(k' \to k)$. By definition, if

$$\alpha_{\text{birth}}(k' \to k) = \min\{1, r'_{\text{birth}}\}, \quad (3.30)$$

then we can directly write

$$\alpha_{\text{death}}(k \to k') = \min\{1, 1/r'_{\text{birth}}\}. \quad (3.31)$$

A classical MH algorithm is applied to update the PAR coefficients when the proposed and recent model orders are equal to each other causing no dimension change between the parameter spaces. Acceptance ratio is defined as $\alpha_{\text{life}} = \min\{1, r_{\text{life}}\}$ where r_{life} is

$$r_{\text{life}} = \frac{f(\mathbf{x}|p, k', \mathbf{a}^{(p,k')}, \sigma_e^2)}{f(\mathbf{x}|p, k, \mathbf{a}^{(p,k)}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(p,k')}|p, k', \sigma_a^2)}{f(\mathbf{a}^{(p,k)}|p, k, \sigma_a^2)} \times \frac{q(\mathbf{a}^{(p,k)}|p, k', \mathbf{a}^{(p,k')})}{q(\mathbf{a}^{(p,k')}|p, k, \mathbf{a}^{(p,k)})}. \quad (3.32)$$

Since PAR models are linear-in-the-parameters and are written in matrix-vector form (3.6), the proposal distribution which is derived for AR model estimation study in (Ehlers and Brooks, 2003) and utilized as proposal distribution (2.66) in the toy example in the previous chapter, is valid for PAR models after performing required changes. Updated values of parameters are proposed from the distribution $q(\mathbf{a}^{(p,k')}|p, k, \mathbf{a}^{(p,k)})$ which is defined below

$$\mathbf{a}^{(p,k')} \sim q\left(\mathbf{a}^{(p,k')}|p, k, \mathbf{a}^{(p,k)}\right) = \mathcal{N}\left(\mathbf{a}^{(p,k')}|\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n^{-1}\right), \tag{3.33}$$

where $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ are

$$\boldsymbol{\mu}_n = \mathbf{a}^{(p,k)} \quad \text{and} \quad \boldsymbol{\Sigma}_n = \sigma_e^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \sigma_a^{-2}\mathbf{I}_\eta. \tag{3.34}$$

At each iteration, the excitation variance σ_e^2 is updated by using (3.37). The updating mechanism follows Gibbs sampling methodology. Following the same procedure as in (2.71), the conditional posterior for σ_e^2 is derived as (Troughton and Godsill, 1998)

$$f(\sigma_e^2|\mathbf{x}, p, k, \mathbf{a}^{(p,k)}) \propto f(\mathbf{x}|k, \mathbf{a}^{(p,k)}, \sigma_e^2)f(\sigma_e^2) \tag{3.35}$$

$$f(\sigma_e^2|\mathbf{x}, p, k, \mathbf{a}^{(p,k)}) \approx \mathcal{N}(\mathbf{e}|\mathbf{0}, \sigma_e^2\mathbf{I}_n)\mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e) \tag{3.36}$$

$$= \mathcal{IG}(\sigma_e^2|\alpha_n, \beta_n) \tag{3.37}$$

where $\alpha_n = \alpha_e + \frac{1}{2}n$ and $\beta_n = \beta_e + \frac{1}{2}\mathbf{e}^{\mathrm{T}}\mathbf{e}$.

3.4.1. Simulation & Results

In order to study performance of RJMCMC algorithm in PAR model estimation, 11 different PAR models (3 linear and 8 nonlinear) are generated synthetically each of which has a length of 1000 samples (firstly, 50000 samples of data is generated and last 1000 samples are selected to ensure that the selected model coefficients are suitable and model generates bounded data, i.e. with finite range). Each data set is driven with a Gaussian excitation sequence with a variance of σ_e^2. Variance parameter of PAR coefficients is selected as $\sigma_a^2 = 0.01$ and no sampling is applied for σ_a^2.

Initial values (also the lower bounds) for model orders, k_0 and p_0 are set to 1. Upper bounds for these parameters are set to 8 and 7, respectively. Additionally, hyperparameters for the prior distribution of excitation variance are selected as $\alpha_e = 1$ and $\beta_e = 2$. Birth and death moves are equiprobable with probabilities of $P_{\text{birth}} = P_{\text{death}} = 0.15$. In order to speed up the convergence of the model coefficient estimations to the correct values during the operation of the algorithm, the life move, which is the coefficient update move, is chosen more likely than the other within-model moves as $P_{\text{life}} = 0.7$. RJMCMC is run for 10000 iterations to let sampled parameters converge.

In order to compare the model selection performance of RJMCMC, results obtained by two popular model order selection methods with names AIC and BIC are provided. The equations for these are given as:

$$\text{AIC} = 2N + n\log(\text{RSS}/n), \tag{3.38}$$

$$\text{BIC} = \log(n)N + n\log(\text{RSS}/n), \tag{3.39}$$

where N is number of parameters for the model, n data length and RSS is *the residual sum of squares* which is $\text{RSS} = \mathbf{x}^{\text{T}}\mathbf{x} - \mathbf{x}^{\text{T}}\mathbf{X}(\mathbf{X}^{\text{T}}\mathbf{X})^{-1}\mathbf{X}^{\text{T}}\mathbf{x}$.

In Figure 3.1 running means of the coefficients of $P^{(3)}\text{AR}(2)$ model and the variance of the excitation sequence are plotted. PAR model for model orders $(3, 2)$ has nine coefficients; however, in the figure only five of them are plotted as well as estimated σ_e^2. Examining all the sub-figures, it can be stated that after nearly 2000 iterations, RJMCMC converges to the estimated values.

In Figures through 3.2 and 3.5, the instantaneous estimates and the joint histograms of the estimated parameters p and k for models $P^{(2)}\text{AR}(3)$ and $P^{(3)}\text{AR}(1)$ are shown. Algorithm decides true order pair, $(p, k) = (2, 3)$, nearly 80% of the iterations and nearly 60% for $(p, k) = (3, 1)$. For both of the models, RJMCMC converges to the correct model orders within 1000 iterations.

Model order selection performances of RJMCMC, AIC and BIC are shown in Table 3.1. Particularly, RJMCMC is superior to AIC for all 11 example models. On the other hand, BIC performs better than RJMCMC especially for linear models. However, this performance of BIC is caused by its bias which penalizes the model complexity. One can clearly see that the cost of this bias becomes evident in model selection performance of BIC for highly nonlinear models shown in the last four columns of Table 3.1 which cannot reach even 5%. In order to demonstrate this effect, the percentages of model selec-
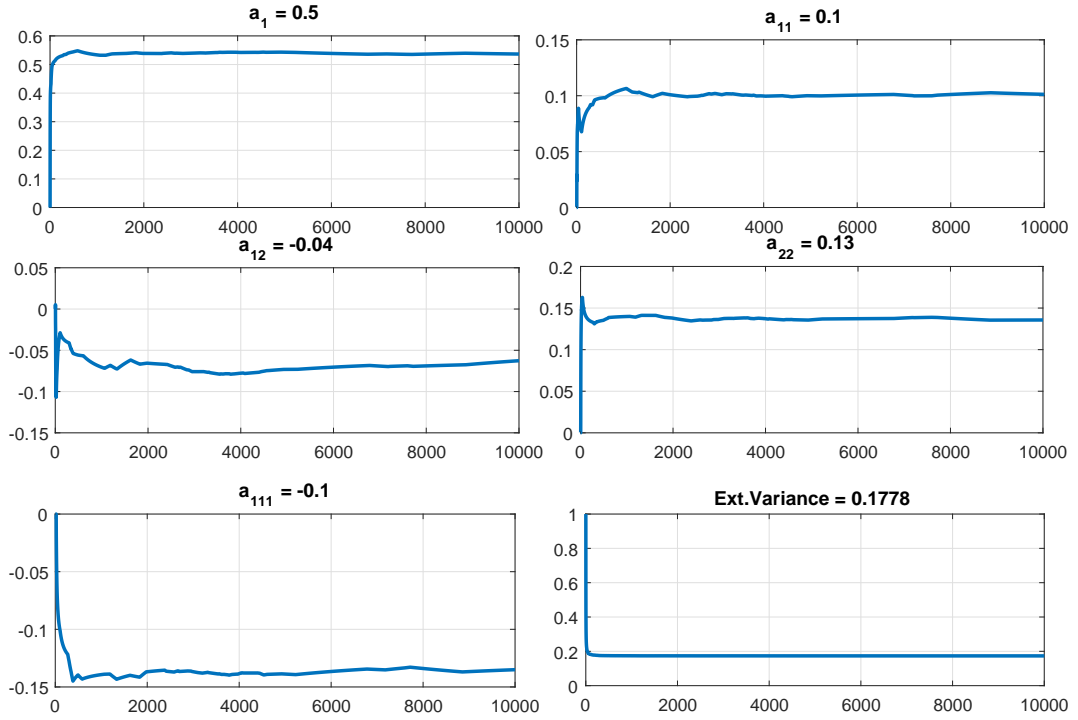
Figure 3.1. Running Mean Plots for $P^{(3)}AR(2)$ Model Estimation. Horizontal axes for each sub-figure refer to RJMCMC iteration.

tion for RJMCMC, AIC and BIC plotted as a function of nonlinearity degree, p, in Figure 3.6 for models $P^{(2)}AR(2)$ and $P^{(3)}AR(2)$. Figure 3.6 and Table 3.1 depict that RJMCMC and BIC achieve very close results for models with small number of coefficients whereas RJMCMC outperforms both AIC and BIC when the nonlinearity degree is increased accompanied by an increasing number of model coefficients.

Coefficient estimation performance are also shown in the last row of Table 3.1 in terms of *normalized mean square error* (NMSE) values after 100 RJMCMC runs. The NMSE is defined as

$$\text{NMSE} = \frac{1}{w} \sum_{i=1}^{w} \frac{(a_i - \hat{a}_i)^2}{\|\mathbf{a}\|_2^2} \tag{3.40}$$

where a_i and $\hat{a}_i$ is the ith element of the η-dimensional coefficient vector $\mathbf{a}$ and its estimate $\hat{\mathbf{a}}$. In addition, $\|\mathbf{a}\|_2$ is the l_2-norm of $\mathbf{a}$. Examining the NMSE figures show that RJMCMC estimates model coefficients for all example PAR models around an average NMSE error of 10^{-2}.

58
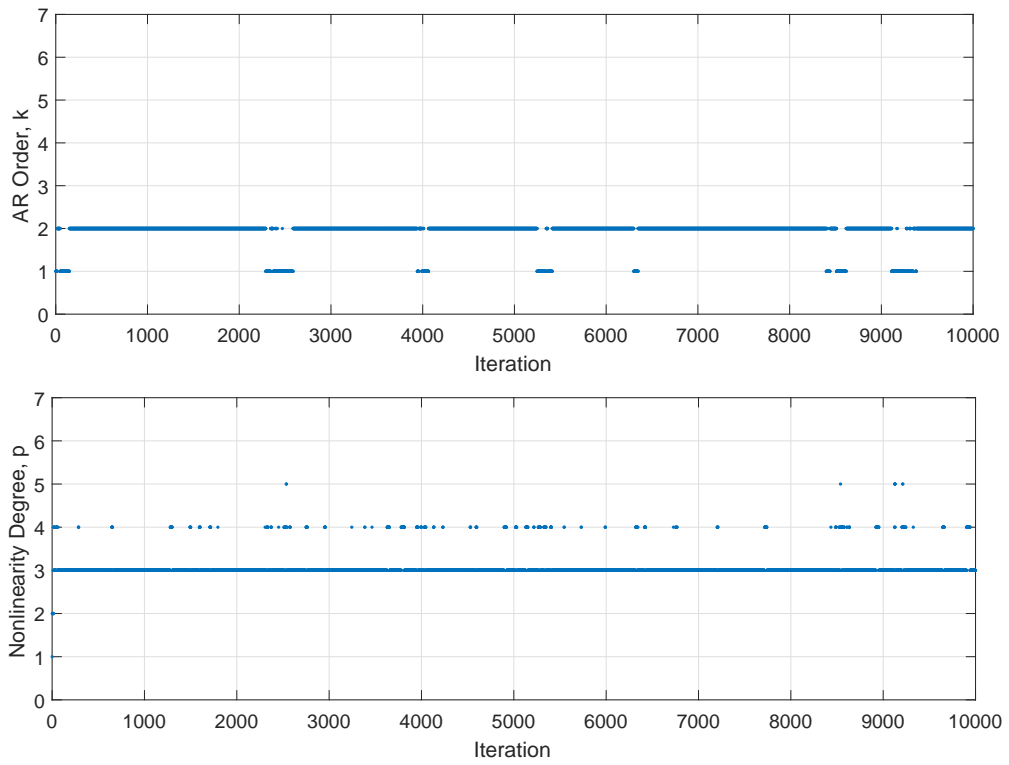
Figure 3.2. Instantaneous plot for nonlinearity degree, p, and the AR order, k, for $P^{(2)}AR(3)$
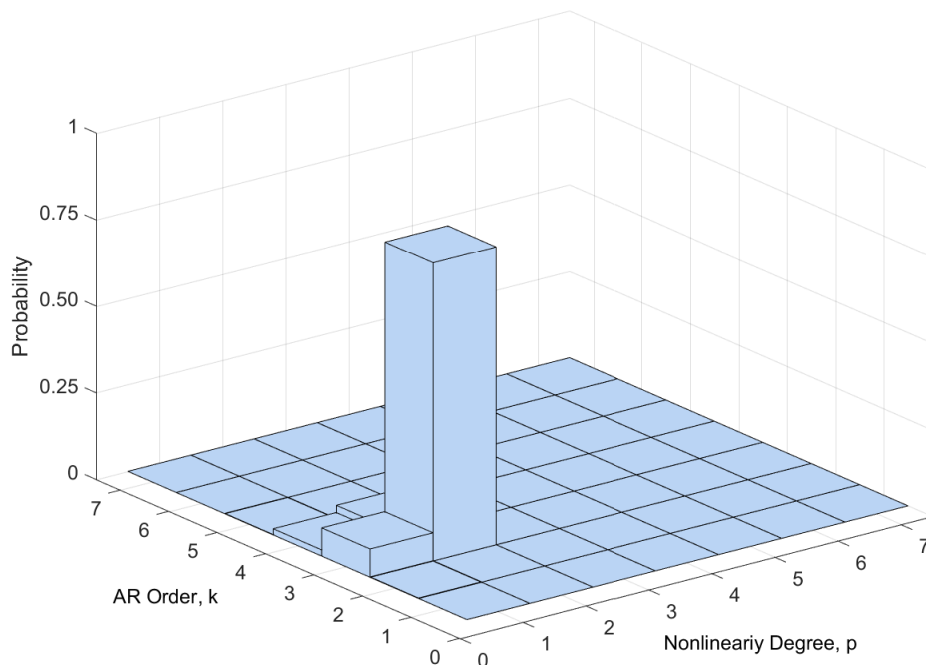
Figure 3.3. Estimated joint histogram for $P^{(2)}AR(3)$ Model
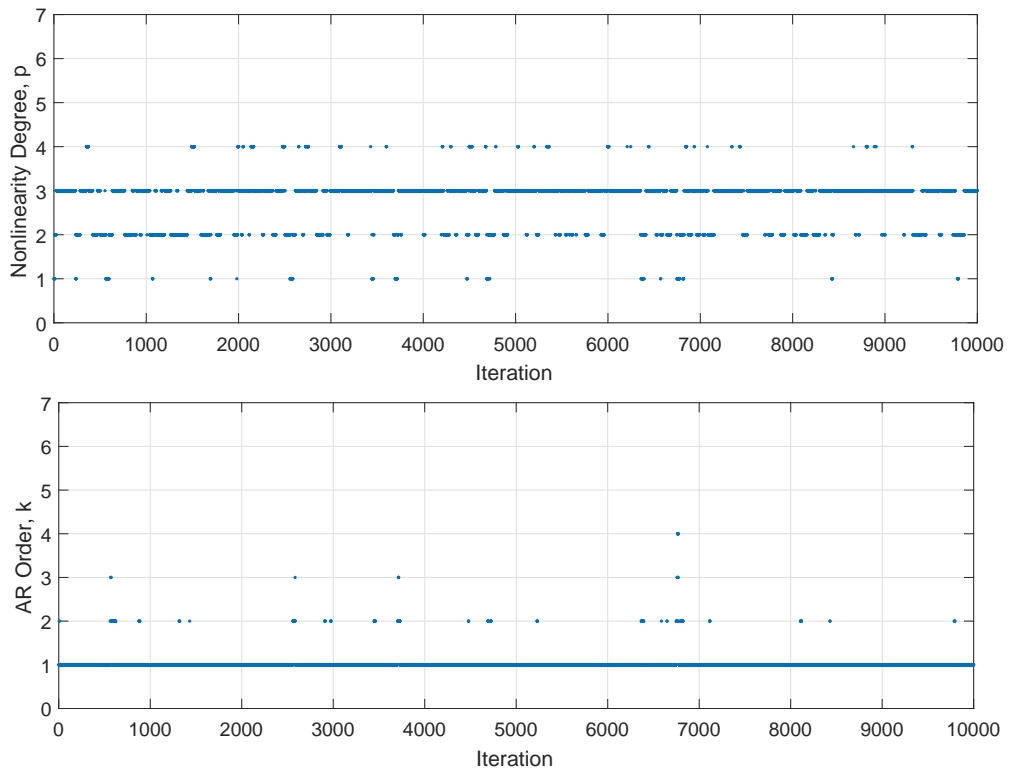
Figure 3.4. Instantaneous plot for nonlinearity degree, p and the AR order, k for $P^{(3)}AR(1)$
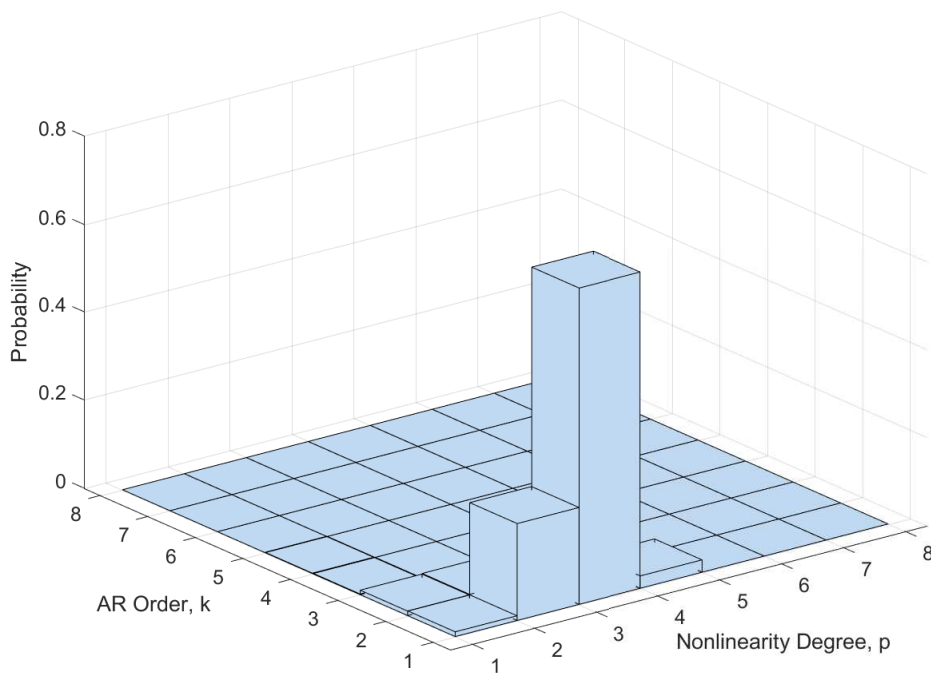
Figure 3.5. Estimated joint histogram for $P^{(3)}AR(1)$ Model

Table 3.1. Detected Percentage of True Models

	P$^{(1)}$AR(1)	P$^{(1)}$AR(2)	P$^{(1)}$AR(3)	P$^{(2)}$AR(1)	P$^{(2)}$AR(2)	P$^{(2)}$AR(3)
RJMCMC	96%	**100%**	**98%**	82%	90%	**88%**
AIC	36%	45%	57%	51%	77%	73%
BIC	**100%**	**100%**	**98%**	**100%**	**96%**	66%
# of Coef.	1	2	3	2	5	9
Avg. NMSE of Coef. Vector Estimate	0.0010	0.0039	0.0034	0.0037	0.0038	0.0017

	P$^{(3)}$AR(1)	P$^{(3)}$AR(2)	P$^{(3)}$AR(3)	P$^{(4)}$AR(1)	P$^{(5)}$AR(1)	
RJMCMC	76%	**76%**	**75%**	**60%**	**84%**	
AIC	66%	51%	43%	17%	17%	
BIC	**92%**	4%	1%	2%	2%	
# of Coef.	3	9	19	4	5	
Avg. NMSE of Coef. Vector Estimate	0.0049	0.0076	0.0200	0.0338	0.0073	

3.5. PMA Model Estimation

With the help of Bayes theorem, the joint posterior density, or namely RJMCMC target distribution, $f(\theta|\mathbf{x})$, is expressed as

$$f(p, q, \mathbf{b}^{(p,q)}, \sigma_e^2, \sigma_b^2|\mathbf{x}) \propto f(\mathbf{x}|p, q, \mathbf{b}^{(p,q)}, \sigma_e^2)f(\mathbf{b}^{(p,q)}|p, q, \sigma_b^2)f(\sigma_b^2)f(\sigma_e^2)f(q)f(p). \quad (3.41)$$

In PMA model estimation study, RJMCMC moves are the same with the ones described in PAR model estimation section. Firstly, for a **birth** move from q to q' where p is fixed, the acceptance ratio is $\alpha_{\text{birth}} = \min\{1, r_{\text{birth}}\}$ where r_{birth} is

$$r_{\text{birth}} = \frac{f(\mathbf{x}|p, q', \mathbf{b}^{(p,q')}, \sigma_e^2)}{f(\mathbf{x}|p, q, \mathbf{b}^{(p,q)}, \sigma_e^2)} \times \frac{f(\mathbf{b}^{(p,q')}|p, q', \sigma_b^2)}{f(\mathbf{b}^{(p,q)}|p, q, \sigma_b^2)} \times \frac{P_{\text{death}}}{P_{\text{birth}}\chi(\mathbf{u})} \times \left|\frac{\partial \mathbf{b}^{(p,q')}}{\partial(\mathbf{b}^{(p,q)}, \mathbf{u})}\right|. \quad (3.42)$$

When the proposed model order q' is lower than the recent one, q ($q' < q$), no new parameters are proposed and a **death** move will be applied. The same procedure as in PAR is executed and the acceptance ratio of the death move appears as $\alpha_{\text{death}}(q \rightarrow q') = \min\{1, 1/r'_{\text{birth}}\}$ where r'_{birth} refers to the ratio of a birth move from q' to q.

When the newly proposed model order is equal to the recent value, e.g. $q' = q$, a
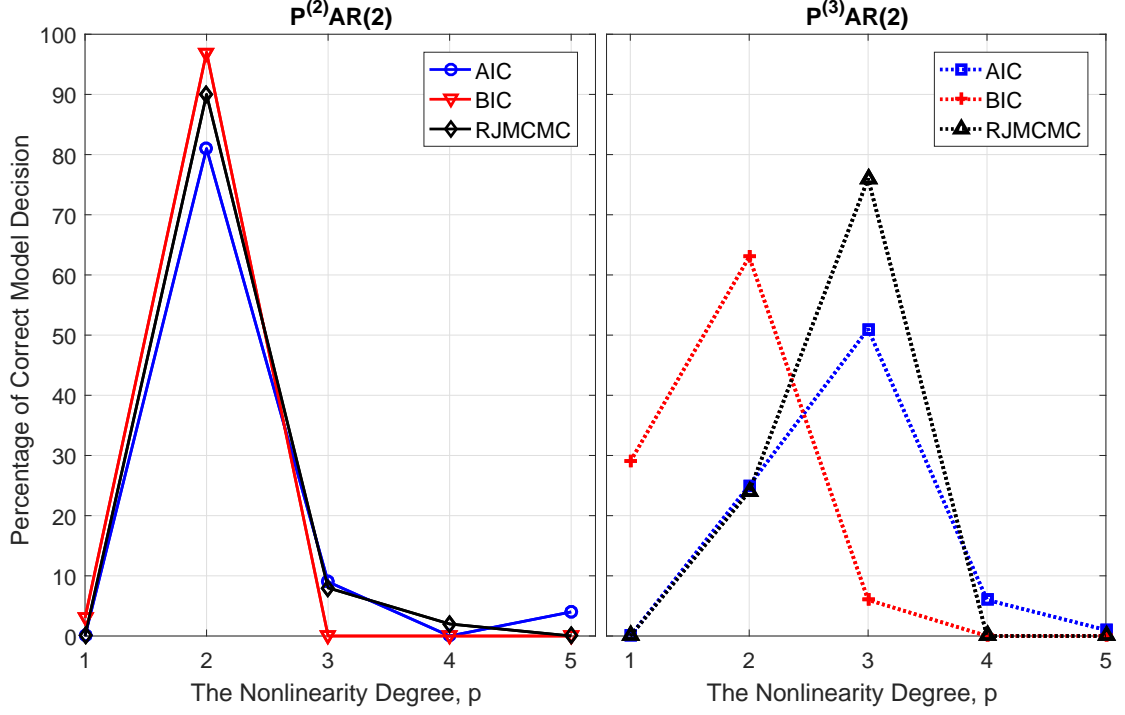
Figure 3.6. Percentage of correct model decision comparison

life move will be applied. Acceptance ratio is defined as $\alpha_{\text{life}} = \min\{1, r_{\text{life}}\}$. So, r_{life} is

$$r_{\text{life}} = \frac{f(\mathbf{x}|p, q', \mathbf{b}^{(p,q')}, \sigma_e^2)}{f(\mathbf{x}|p, q, \mathbf{b}^{(p,q)}, \sigma_e^2)} \times \frac{f(\mathbf{b}^{(p,q')}|p, q', \sigma_b^2)}{f(\mathbf{b}^{(p,q)}|p, q, \sigma_b^2)} \times \frac{\psi(\mathbf{b}^{(p,q)}|p, q', \mathbf{b}^{(p,q')})}{\psi(\mathbf{b}^{(p,q')}|p, q, \mathbf{b}^{(p,q)})}. \quad (3.43)$$

In order to update model coefficients, a proposal distribution of $\psi(\cdot)$ is defined as

$$\mathbf{b}^{(p,q')} \sim \psi(\mathbf{b}^{(p,q')}|p, q, \mathbf{b}^{(p,q)}), \quad (3.44)$$

$$= \mathcal{N}(\mathbf{b}^{(p,q')}|\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n^{-1}), \quad (3.45)$$

where $\boldsymbol{\mu}_n = \sigma_e^{-2}\boldsymbol{\Sigma}_n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{x}$ and $\boldsymbol{\Sigma}_n = \sigma_e^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{X} + \sigma_b^{-2}\mathbf{I}_\gamma$.

Excitation variance, σ_e^2, and variance of model coefficients, σ_b^2, are updated at each iteration via Gibbs Sampling. Constructing the full conditional distribution for σ_e^2 follows the same steps in AR model selection example in Section 2.6.2.3 and is chosen to

be inverse Gamma as (Troughton and Godsill, 1998)

$$f(\sigma_e^2|\mathbf{x}, p, q, \mathbf{b}^{(p,q)}) \propto f(\mathbf{x}|p, q, \mathbf{b}^{(p,q)}, \sigma_e^2)f(\sigma_e^2) \tag{3.46}$$

$$\approx \mathcal{N}(\mathbf{e}|\mathbf{0}, \sigma_e^2\mathbf{I}_n)\mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e) \tag{3.47}$$

$$= \mathcal{IG}(\sigma_e^2|\alpha_{en}, \beta_{en}) \tag{3.48}$$

where $\alpha_{en} = \alpha_e + \frac{1}{2}n$, $\beta_{en} = \beta_e + \frac{1}{2}\mathbf{e}^{\mathrm{T}}\mathbf{e}$.

Similarly, the full conditional distribution for σ_b^2 is obtained as (Troughton and Godsill, 1998)

$$f(\sigma_b^2|\mathbf{x}, p, q, \mathbf{b}^{(p,q)}) \propto f(\mathbf{b}^{(p,q)}|\sigma_b^2)f(\sigma_b^2) \tag{3.49}$$

$$\approx \mathcal{N}(\sigma_b^2|\mathbf{0}, \sigma_b^2\mathbf{I}_\gamma)\mathcal{IG}(\sigma_b^2|\alpha_b, \beta_b) \tag{3.50}$$

$$= \mathcal{IG}(\sigma_b^2|\alpha_{bn}, \beta_{bn}) \tag{3.51}$$

where $\alpha_{bn} = \alpha_b + \frac{1}{2}\gamma$ and $\beta_{bn} = \beta_b + \frac{1}{2}(\mathbf{b}^{(p,q)})^{\mathrm{T}}\mathbf{b}^{(p,q)}$.

Model estimation procedure for PMA requires past samples of the excitation sequence. However, this sequence is an unobserved sequence and thus it is required to be estimated. For this purpose, an initial excitation vector, $\mathbf{e}^{(0)}$, is created and the complete excitation vector $\mathbf{e}$ is estimated at each iteration via (3.11) starting from $e(q+1)$ up to $e(n)$ by using $\mathbf{e}^{(0)}$. Hence, the aforementioned initial excitation vector, $\mathbf{e}^{(0)}$ is sampled from the distribution

$$\mathbf{e}^{(0)} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2\mathbf{I}_q). \tag{3.52}$$

3.5.1. Simulation & Results

In order to study the RJMCMC performance on PMA model estimation, 6 different PMA models (2 linear and 4 nonlinear) each of which has 500 samples, are generated synthetically. Each data set is driven with a Gaussian excitation sequence with a variance of σ_e^2. So as to provide the output sequence of each PMA models are normally distributed,

Kolmogorov-Smirnov and Kullback-Leibler tests have been performed. Hyperparameters are set to $\alpha_e = \alpha_b = 1$ and $\beta_e = \beta_b = 2$. The initial model orders q_0 and p_0 are set to 1 and both of upper bounds p_{max} and q_{max} are set to 6. RJMCMC move probabilities are selected as 0.15, 0.15 and 0.7, respectively, for P_{birth}, P_{death} and P_{life}. 10000 iterations are run to let sampled parameters converge.

AIC/BIC calculations have not been performed for PMA and PARMA simulations. These results will not be used for PMA and PARMA because of the need for a troublesome estimation step for the unobserved excitation sequence in AIC/BIC calculations. We also believe that the performance degradation of these model selection methods, shown in the PAR results (Figure 3.6), will be further increased, taking into account the error caused by the estimation of the unobserved excitation data in the PMA and PARMA models.

Table 3.2. Model Estimation Results

	$P^{(1)}MA(4)$	$P^{(1)}MA(5)$	$P^{(2)}MA(2)$
Percentage of Detection	70%	70%	100%
Avg. NMSE of Coeff. Vector Estimate	0.0287	0.0448	0.0174
# of Coeff. (w)	4	5	5
	$P^{(2)}MA(3)$	$P^{(3)}MA(2)$	$P^{(4)}MA(1)$
Percentage of Detection	100%	65%	70%
Avg. NMSE of Coeff. Vector Estimate	0.0198	0.0308	0.0531
# of Coeff. (w)	9	9	4

Table 3.2 shows the percentage of detection for true model order pairs and estimated model coefficient error values in terms of NMSE values. The NMSE for PMA model coefficients is defined as:

$$\text{NMSE} = \frac{1}{w} \sum_{i=1}^{w} \frac{(b_i - \hat{b}_i)^2}{\|\mathbf{b}\|_2^2}. \tag{3.53}$$

RJMCMC is run 20 times for each PMA model and average NMSE values after 20 repetitions are presented in Table 3.2. Examining the results in Table 3.2, for all example PMA models RJMCMC estimates true model order pairs with at least 65% of percentage within 20 simulations. Moreover, RJMCMC achieves reasonable performance on model coefficients estimation by succeeding average NMSE of around 10^{-2} for all models.

In Figures 3.7 and 3.8, the instantaneous estimates and the joint posterior density

of the model orders p and q for models $P^{(1)}MA(4)$ and $P^{(2)}MA(2)$ are plotted in a single RJMCMC run. For both of the models, RJMCMC estimates the true model orders over 50% of the iterations.

3.6. PARMA Model Estimation

In PARMA model estimation study, the target distribution of RJMCMC $f(\boldsymbol{\theta}|\mathbf{x})$ is decomposed via Bayes Theorem for the parameter vector $\boldsymbol{\theta} = \{p, k, q, \mathbf{a}^{(p,k)}, \mathbf{b}^{(p,q)}, \sigma_e^2, \sigma_a^2, \sigma_b^2\}$

$$f(\boldsymbol{\theta}|\mathbf{x}) \propto f(\mathbf{x}|p, k, q, \mathbf{a}^{(p,k)}, \mathbf{b}^{(p,q)}, \sigma_e^2) f(\mathbf{a}^{(p,k)}|p, k, \sigma_a^2) \times$$
$$f(\mathbf{b}^{(p,q)}|p, q, \sigma_b^2) f(\sigma_a^2) f(\sigma_b^2) f(\sigma_e^2) f(p) f(k) f(q). \quad (3.54)$$

Assume that a **birth** move has been proposed from AR order k to k' when p and q are fixed. Then, resulting acceptance ratio is $\alpha_{\text{birth}} = \min\{1, r_{\text{birth}}\}$ where r_{birth} can be given as

$$r_{\text{birth}} = \frac{f(\mathbf{x}|p, k', q, \mathbf{a}^{(p,k')}, \mathbf{b}^{(p,q)}, \sigma_e^2)}{f(\mathbf{x}|p, k, q, \mathbf{a}^{(p,k)}, \mathbf{b}^{(p,q)}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(p,k')}|p, k', \sigma_a^2)}{f(\mathbf{a}^{(p,k)}|p, k, \sigma_a^2)} \times \frac{P_{\text{death}}}{P_{\text{birth}} \chi(\mathbf{u})} \times \left| \frac{\partial \mathbf{a}^{(p,k')}}{\partial(\mathbf{a}^{(p,k)}, \mathbf{u})} \right|. \quad (3.55)$$

When a **death** move has been proposed from k to k', no new parameters are being proposed. Since birth and death moves are reversible move pairs, $\alpha_{\text{death}}(k \to k')$ can be written in terms of $\alpha_{\text{birth}}(k' \to k)$.

When $k' = k$, a **life** move will be proposed with acceptance ratio, $\alpha_{\text{life}} = \min\{1, r_{\text{life}}\}$ with

$$r_{\text{life}} = \frac{f(\mathbf{x}|p, k', q, \mathbf{a}^{(p,k')}, \mathbf{b}^{(p,q)}, \sigma_e^2)}{f(\mathbf{x}|p, k, q, \mathbf{a}^{(p,q)}, \mathbf{b}^{(p,q)}, \sigma_e^2)} \times \frac{f(\mathbf{a}^{(p,k')}|p, k', \sigma_a^2)}{f(\mathbf{a}^{(p,k)}|p, k, \sigma_a^2)} \times \frac{\psi(\mathbf{a}^{(p,k)}|p, k', \mathbf{a}^{(p,k')})}{\psi(\mathbf{a}^{(p,k')}|p, k, \mathbf{a}^{(p,k)})} \quad (3.56)$$

where the proposal distribution $\psi(\cdot)$ is

$$\psi(\mathbf{a}^{(p,k)}|p, k', \mathbf{a}^{(p,k')}) = \mathcal{N}(\mathbf{a}^{(p,k')}|\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n^{-1}), \quad (3.57)$$

with $\boldsymbol{\mu}_n = \sigma_e^{-2} \boldsymbol{\Sigma}_n^{-1} \mathbf{X}^{\mathrm{T}} (\mathbf{x} - \mathbf{E} \mathbf{b}^{(p,q)})$ and $\boldsymbol{\Sigma}_n = \sigma_e^{-2} \mathbf{X}^{\mathrm{T}} \mathbf{X} + \sigma_a^{-2} \mathbf{I}_\eta$.
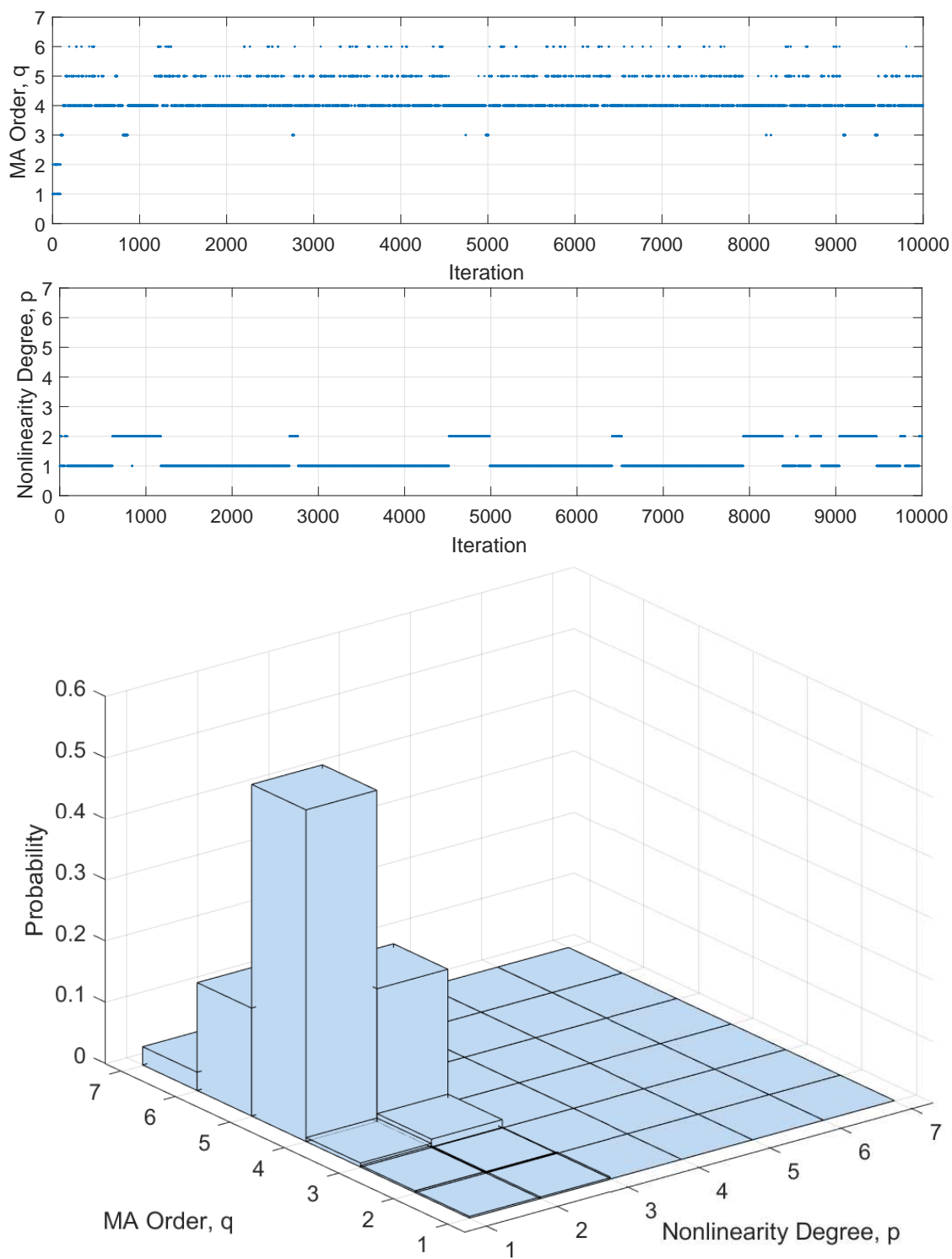
Figure 3.7. The instantaneous model order estimates and the joint posterior density of the model orders of $P^{(1)}MA(4)$
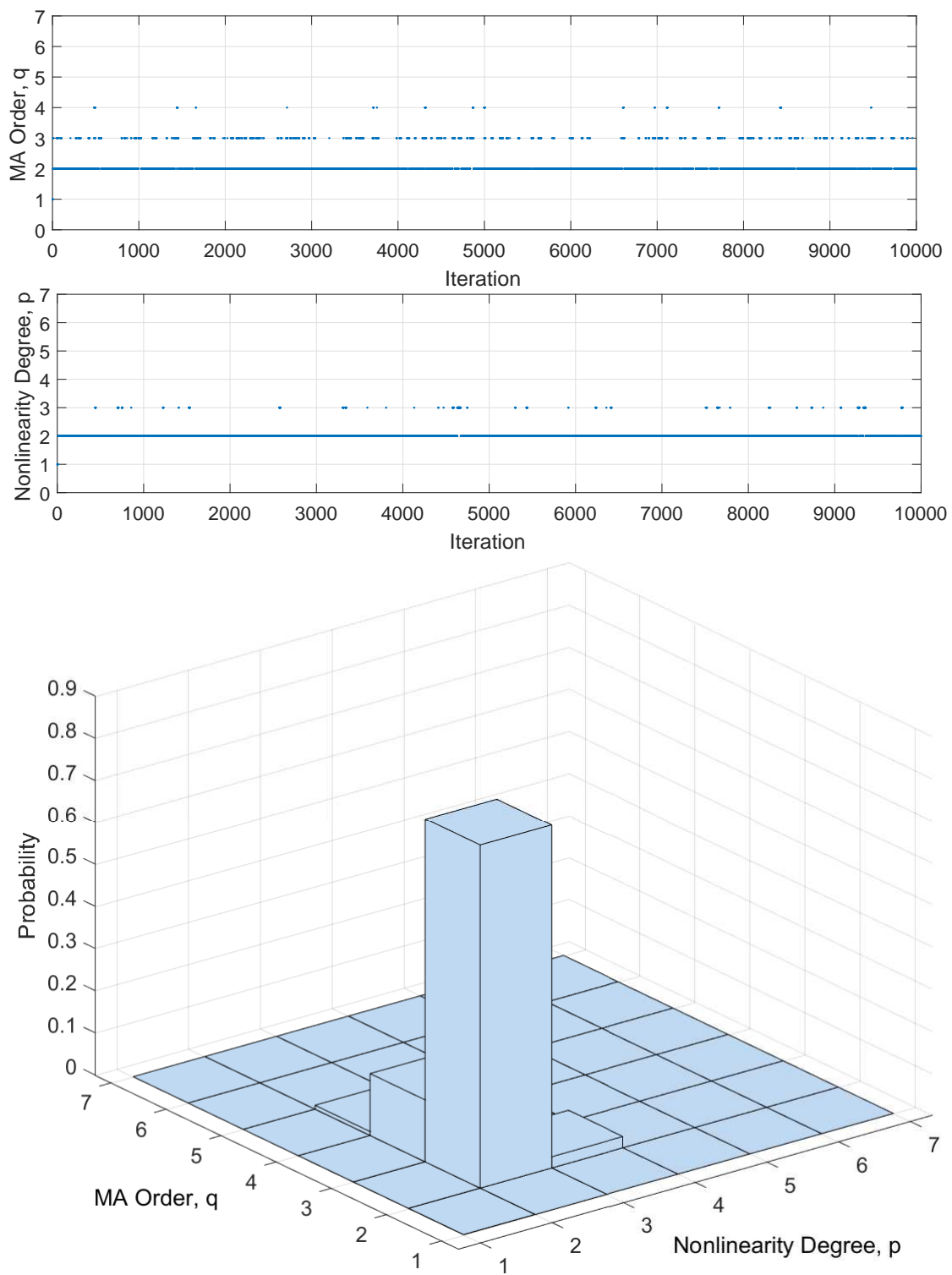
Figure 3.8. The instantaneous model order estimates and the joint posterior density of the model orders of $P^{(2)}MA(2)$

At each iteration several update moves are performed to update variance parameters and initial unobserved excitation vector $\mathbf{e}^{(0)} = [e_1, e_2, \ldots, e_q]$ via Gibbs sampling. The full conditional distribution for σ_e^2 is (Troughton and Godsill, 1998)

$$f(\sigma_e^2|\mathbf{x}, p, q, \mathbf{b}^{(p,q)}) = \mathcal{IG}(\sigma_e^2|\alpha_{en}, \beta_{en}), \tag{3.58}$$

where $\alpha_{en} = \alpha_e + \frac{1}{2}n, \beta_{en} = \beta_e + \frac{1}{2}\mathbf{e}^T\mathbf{e}$. Similarly, the full conditionals for coefficient vector variances appear as

$$f(\sigma_a^2|\mathbf{x}, p, k, \mathbf{a}^{(p,k)}) = \mathcal{IG}(\sigma_a^2|\alpha_{an}, \beta_{an}), \tag{3.59}$$

$$f(\sigma_b^2|\mathbf{x}, p, q, \mathbf{b}^{(p,q)}) = \mathcal{IG}(\sigma_b^2|\alpha_{bn}, \beta_{bn}), \tag{3.60}$$

where parameters for these inverse Gamma functions are $\alpha_{an} = \alpha_a + \frac{1}{2}\eta$, $\beta_{an} = \beta_a + \frac{1}{2}(\mathbf{a}^{(p,k)})^T\mathbf{a}^{(p,k)}$, $\alpha_{bn} = \alpha_b + \frac{1}{2}\gamma$ and $\beta_{bn} = \beta_b + \frac{1}{2}(\mathbf{b}^{(p,q)})^T\mathbf{b}^{(p,q)}$.

Initial excitation vector $\mathbf{e}^{(0)}$ can be sampled from the distribution given in (3.52).

3.6.1. Simulation & Results

In order to study the performance, 10 PARMA models have been generated for simulations. Each model generates 20 zero mean ($\mu = 0$) data sets each of length 750 samples. Each data set is driven with a Gaussian excitation sequence of variance of σ_e^2.

Hyperparameters $\alpha_a, \alpha_b, \alpha_e, \beta_a, \beta_b$ and β_e are selected as 1. Each of the initial values p_0, k_0 and q_0 is also set to 1. Upper bounds, $p_{\max}, k_{\max}$ and $q_{\max}$, for model orders are selected as 5, 6 and 6, respectively. RJMCMC offers birth and death model moves with probability of 0.15 and life model move with a probability of 0.7. Apart from the procedures in PAR and PMA model estimation, RJMCMC performs a training procedure by applying 10 consecutive life moves to update the newly estimated model coefficients if the proposed model order is accepted. Each RJMCMC run performs 25000 iterations including a burn-in period of 10000 iterations. For PAR and PMA, there has been no need to discriminate a burn-in period since the algorithm attains the characteristic of steady state behavior in a very small number of samples compared to simulation length.

In order to measure the estimation performance of the proposed method for model coefficients, NMSE has been used. The expression for NMSE in estimating PARMA

model coefficients is

$$\text{NMSE} = \frac{1}{\eta + \gamma} \sum_{i=1}^{\eta+\gamma} \frac{(h_i - \hat{h}_i)^2}{\|\mathbf{h}\|_2^2}, \tag{3.61}$$

where vector $\mathbf{h}$ includes both AR and MA model coefficients vectors $\mathbf{a}^{(p,k)}$ and $\mathbf{b}^{(p,q)}$ and $\hat{\mathbf{h}}$ is its estimate.

Table 3.3 demonstrates results after 20 RJMCMC runs. These results reveal that for all 10 PARMA models, RJMCMC estimates correct model orders with the highest percentages. In addition, RJMCMC exhibits great performance on model coefficients estimation with NMSE values between 6×10^{-4} and 1×10^{-2} for all the models.

Table 3.3. Model estimation results

	(1,3,0)	(1,0,4)	(1,3,1)	(1,2,2)	(2,1,1)
% Perf. of Detection	85	65	80	80	95
Avg. NMSE	0.0015	0.0022	0.0081	0.0145	0.0079
# of Coeff.	3	4	4	4	4
	(2,2,1)	(2,3,1)	(3,1,0)	(3,0,1)	(3,1,1)
% Perf. of Detection	90	100	70	80	80
Avg. NMSE	0.0050	0.0056	0.0006	0.0011	0.0030
# of Coeff.	7	11	3	3	6

Figure 3.9 depicts the estimated joint posteriors of model orders for 3 example PARMA models in a single RJMCMC run. Each sub-figure shows 5 most frequently estimated models and their probabilities after burn-in period. Examining the figure, we can clearly state that models which have the highest probabilities are the correct models and these posteriors depict that RJMCMC estimates the correct models.

In Figure 3.10, instantaneous estimates for each model order p, k and q for a $P^{(2)}$ARMA(2,1) model, are shown for a single RJMCMC run with 25000 iterations. Burn-in period is also shown in Figure 3.10 in order to see the transient character of the algorithm.

Figure 3.11 shows the posterior probabilities for model coefficients of $P^{(1)}$ARMA(3,1) in a single RJMCMC run. Vertical black line corresponds to the correct model coefficient value for each sub-figure. Coefficient estimates for only correctly estimated model orders after burn-in period have been used to obtain the distributions in this figure. Resulting pos-
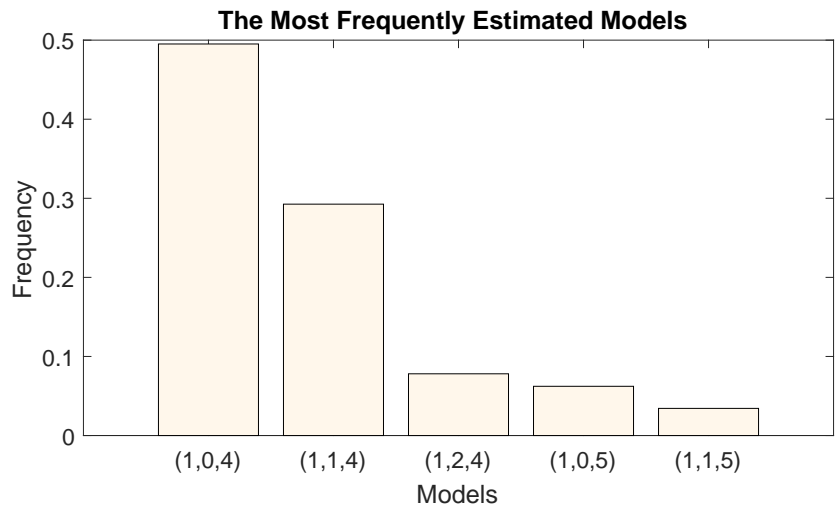
(a) $P^{(1)}$ARMA(0,4)
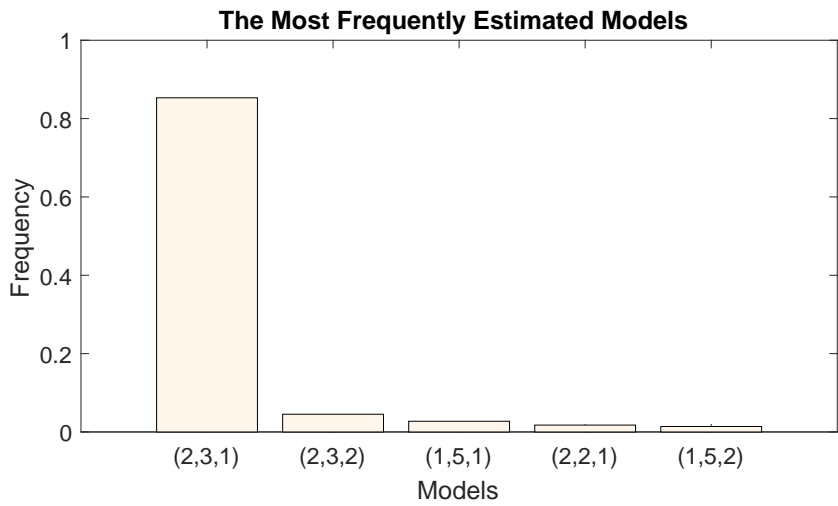
(b) $P^{(2)}$ARMA(3,1)
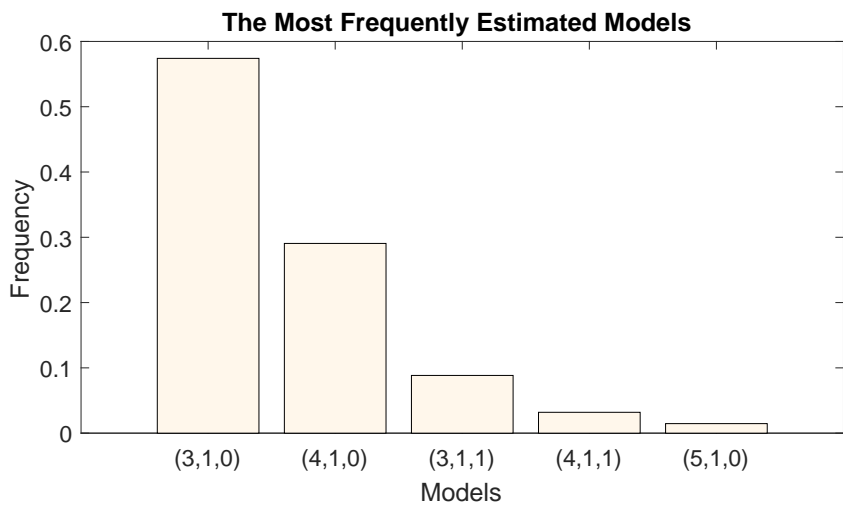
(c) $P^{(3)}$ARMA(1,0)

Figure 3.9. Estimated joint posteriors for model orders in a single RJMCMC run.
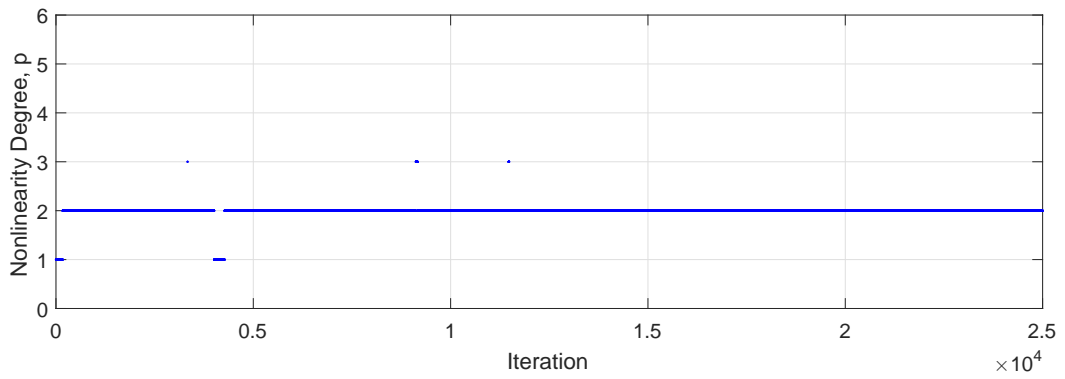
(a) Nonlinearity Degree, p
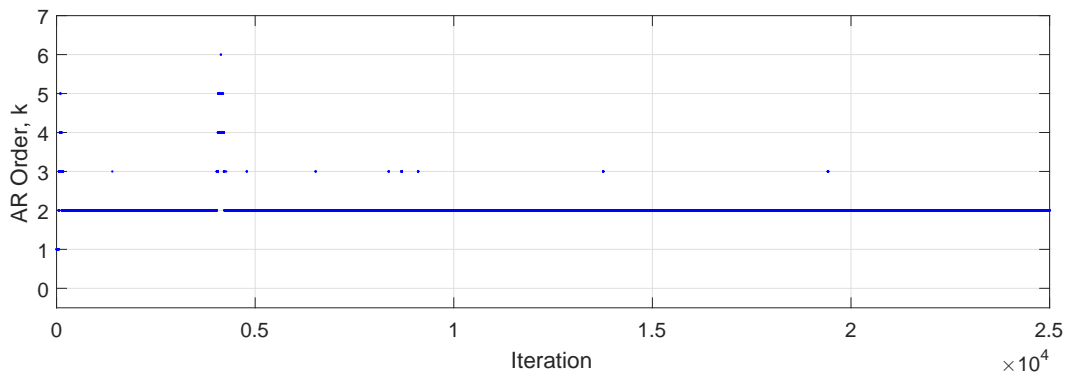
(b) AR Order, k
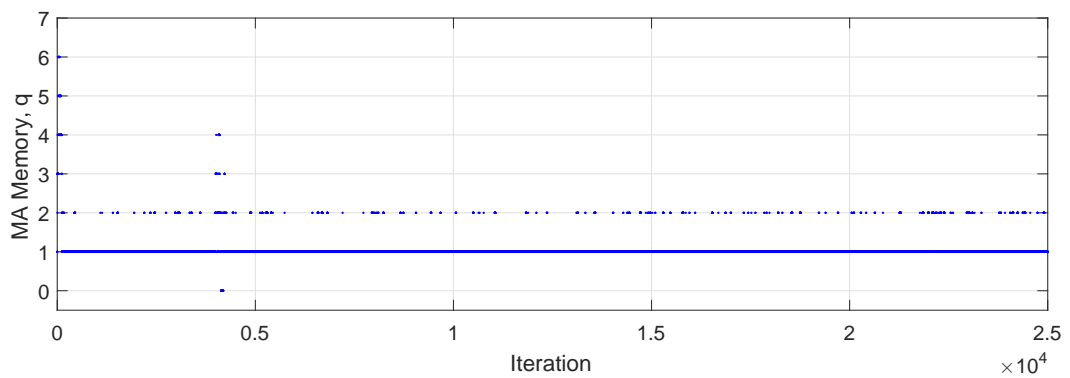
(c) MA Order, q

Figure 3.10. The instantaneous model order estimates of $P^{(2)}ARMA(2,1)$.

teriors show that RJMCMC estimates model coefficients with a remarkable performance and the correct model coefficient values stand in 95% *confidence interval* (CI) ($\pm 2\sigma$) for b_1 and a_1 and 68% CI ($\pm\sigma$) for a_2 and a_3 of the estimated posteriors.
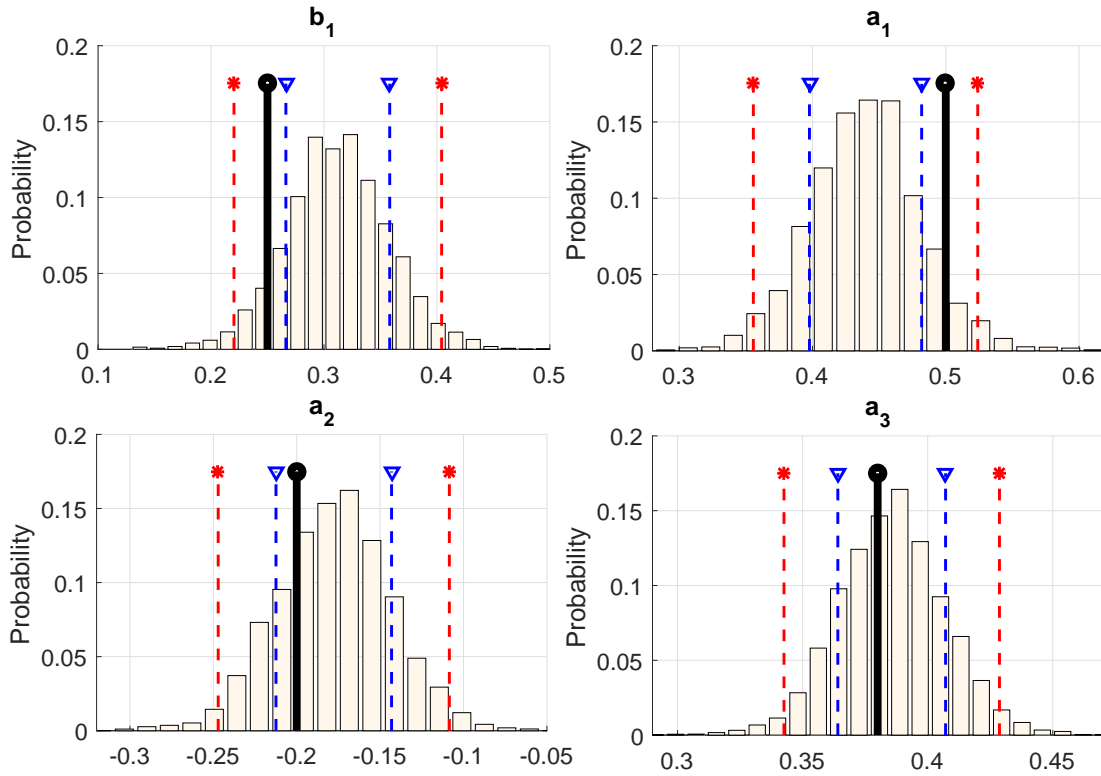
Figure 3.11. Estimated posteriors for $P^{(1)}$ARMA(3,1) coefficients (where the selected coefficient vector is $\mathbf{b}^* = [0.25, 0.5, -0.2, 0.38]$). *Vertical lines with "o" marker are the correct coefficient values, vertical dashed lines with "∇" and "$*$" markers refer to $\pm\sigma$ and $\pm 2\sigma$ CIs, respectively.*

3.7. Conclusions on Nonlinear Model Estimation

In this chapter of the thesis, we have presented a general framework for the estimation problems of Volterra based linear-in-the-parameters nonlinear models, namely PAR, PMA, and PARMA by demonstrating the classical trans-dimensional RJMCMC procedure in an anomalous way. Numerical results demonstrate promising performance of the proposed method in estimating model orders and corresponding model coefficients concurrently for all three nonlinear time series models. Particularly, model order selection performance of the proposed method provides remarkable performance compared to

AIC and BIC, especially when the system nonlinearity degree is high. Additionally, the coefficient estimation results of the proposed method are also very successful. Despite not knowing the correct model orders, this automatic estimation procedure also reveals that the proposed method is an important complete estimation method.

The given toy example about linear AR model uncertainty in the previous chapter, has been extended to a nonlinear model estimation study, and thus the ability of RJMCMC for making transitions between different structural models is demonstrated. This result also suggests that the proposed method is an important method compared to the other classical methods in nonlinear model selection applications, especially due to its ability to estimate the polynomial nonlinearity degree of the models.

Another important contribution of the proposed method is that it can be used as a nonlinearity test procedure for given sets of data in terms of polynomial nonlinearity. This contribution is due to the ability of the proposed method to switch between the linear and nonlinear models. The applicability of the classical trans-dimensional method in much more complex applications is clearly demonstrated.

Contrary to its classical usage, the extension from linear models to nonlinear ones, presented in this chapter, provides a very broad application framework addressed by RJMCMC. If the necessary modifications are made, the RJMCMC can be a model estimation method that can be successfully used in the solution of much more difficult problems. To name a few, the proposed use of RJMCMC in this section, are not limited to PARMA based models. In addition, further applications are also not limited to Gaussian innovation cases, and thus, this approach can also be applied to model estimation studies with non-Gaussian innovation sequences.

CHAPTER 4

BAYESIAN VOLTERRA SYSTEM IDENTIFICATION

Models for nonlinear systems can be classified as nonparametric and parametric. Nonparametric models include state-space and nonlinear phase-space models all of which generally have infinite number of parameters. Infinite ordered Volterra and Wiener models are also located within this group. In parametric nonlinear system models, one can express the input-output relationship in terms of a mathematical function which has a finite number of parameters. Polynomial nonlinear models, which can also be defined as truncated Volterra models, are important parametric nonlinear models (Nowak, 2002).

Among these nonlinear system models, Volterra models (Schetzen, 1980) are favorable due to their linear-in-the-parameters property. In addition, there are several reasons which make Volterra models appealing in nonlinear modelling. Firstly, Volterra models provide approximation on continuously differentiable transfer functions with Taylor series expansion. This makes them adaptable for representing many nonlinear systems. Furthermore, various nonlinear differential equations such as Lotka-Volterra, Schrödinger (Kang, 2004) can be rewritten as a Volterra system. Secondly, the inverse of a Volterra model is also Volterra type and thus they provide easy-to-implement solutions for system identification problems (Le Caillec, 2011).

A discrete time Volterra model with the output $y(l)$ is given by (Alper, 1965)

$$y(l) = \mu + \sum_{m=1}^{p} \sum_{\tau_1=1}^{q} \ldots \sum_{\tau_m=\tau_{m-1}}^{q} h_{\tau_1,\ldots,\tau_m}^{(m)} \prod_{j=1}^{m} x(l - \tau_j) \qquad (4.1)$$

where $x(\cdot)$ is the input of the model and $h_{\tau_1,\ldots,\tau_m}^{(m)}$ refers to the mth order discrete Volterra model coefficients (or namely kernels). The nonlinearity degree is denoted by p and q indicates the system memory. The notation of $V(p, q)$ can be used for Volterra models.

Examining (4.1), the matrix-vector form of the Volterra models is

$$\mathbf{y} = \mu + \mathbf{X}\mathbf{h}^{(\mathbf{p},\mathbf{q})} \qquad (4.2)$$

where μ refers to the intercept and the $\eta \times 1$ coefficient vector $\mathbf{h}^{(\mathbf{p},\mathbf{q})}$ and $n \times \eta$ data matrix

X are

$$\mathbf{h}^{(p,q)} = \begin{bmatrix} h_1^{(1)}, & h_2^{(1)}, & \ldots, & h_q^{(1)}, & h_{1,1}^{(2)}, & h_{1,2}^{(2)}, & \ldots, & h_{q,q}^{(2)}, & \ldots, & h_{q,q,\ldots,q}^{(p)} \end{bmatrix}^T \qquad (4.3)$$

$$\mathbf{X} = \begin{bmatrix} x(0) & x(-1) & \ldots & x(1-q) & x^2(0) & x(0)x(-1) & \ldots & x^2(1-q) & \ldots & x^p(1-q) \\ x(1) & x(0) & \ldots & x(2-q) & x^2(1) & x(1)x(0) & \ldots & x^2(2-q) & \ldots & x^p(2-q) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x(n-1) & x(n-2) & \ldots & x(n-q) & x^2(n-1) & x(n-1)x(n-2) & \ldots & x^2(n-q) & \ldots & x^p(n-q) \end{bmatrix}$$

$$(4.4)$$

where n specifies the data length. The number of Volterra coefficients, also denoted by η, can be calculated for a V(p,q) model by

$$\eta = \binom{p+q}{p} - 1 = \frac{(p+q)!}{p!\, q!} - 1. \qquad (4.5)$$

Volterra system models have been popular models and have achieved remarkable success in modelling nonlinear systems in many real life applications. In signal processing, Volterra models are widespread models and their application areas cover speech, image, communications, audio, mechanical systems, etc. To name a few: in audio, Volterra model has been used for parametric loudspeaker system identification in (Ji and Gan, 2012) and for acoustic echo cancellation in (Contan et al., 2013). Second order Volterra models have been utilized in identification of drift oscillations of moored vessels objects in (Koh and Powers, 1985). Nonlinear communication channels in satellite links have been modelled with sparse third order (cubic) Volterra systems which were estimated via adaptive algorithms (Kalouptsidis et al., 2011). Authors have proposed an indirect Volterra model for predistorter application in compensation of a given system (Eun and Powers, 1997). Volterra system models were applied also in (Mhatli et al., 2015) to coherent optical fiber systems outperforming the adaptive reference methods on equalizing the fiber link channel effects. Volterra systems with complex coefficients have been used in blind identification of single-input-single-output (SISO) communication channels with second order nonlinearity in (Mileounis and Kalouptsidis, 2009) and for linear time invariant (LTI) finite impulse response (FIR) multiple-input-multiple-output (MIMO) systems in (Kotoulas et al., 2011). Generally in the literature, as shared by the mentioned applications, VSI methodology has been applied to the models with predetermined nonlinearity degree and system memory. However, a prior knowledge of the nonlinearity degree is

an unrealistic assumption for most of these applications and estimating the nonlinearity degree is extremely important.

Recently, adaptive approaches are favourable to solve VSI problems. Particularly, the aforementioned methods generally perform estimation studies about Volterra system coefficients based on *nonlinear least mean squares* (NLMS) (Glentis et al., 2014; Chaudhary et al., 2016), *nonlinear recursive least squares* (NRLS) (Shi and Shi, 2011; Claser et al., 2016), *least mean pth power* (LMP) (Lu et al., 2016), and extended Kalman filters (Bouilloc and Favier, 2012; Batselier et al., 2016). In addition to adaptive approaches, genetic algorithms (Abbas and Bayoumi, 2006; Merabti and Massicotte, 2014), QR decomposition (Shoaib et al., 2010), neuro-fuzzy (Loussifi et al., 2016) and neural network (Fortuna et al., 2003) architectures have also been utilized in VSI.

Several Bayesian approaches have been used in SI problems in the literature. A MH based MCMC approach has used to identify dynamical systems in (Ninness and Henriksen, 2010), authors (Beck, 2010) have proposed an identification scheme based on probability logic with Bayesian updating. In addition to these, in order to identify nonlinear dynamic systems *transitional Markov chain Monte Carlo* (TMCMC) (Green and Worden, 2015) and *simulated annealing* (SA) (Green, 2015) are also used.

In the previous chapter, an approach has been presented which uses RJMCMC in estimating the linear-in-the-parameters nonlinear time series methods. Particularly, a complete approach has been established that RJMCMC can be used for transitions between linear/nonlinear models using the classical trans-dimensional approach and to estimate the nonlinearity degrees of the PAR/PMA/PARMA models as well.

The procedure in this chapter will take the approach in the previous chapter one step further, suggesting a more general RJMCMC move than the classical trans-dimensional moves. This new RJMCMC procedure will be called as *trans-structural*. It is *not trans-dimensional* because performs transitions between spaces regardless of their dimension sizes. It *is trans-structural* because it performs transitions across multiple dimensions simultaneously and between different structural model classes, such as linear and nonlinear. The trans-structural RJMCMC (Karakuş et al., 2017a) will be used for the identification of Volterra system models. Thus, a solution will be presented for the fixed nonlinearity degree problem in the above mentioned literature examples.

4.1. Nonlinear System Identification

System identification is a task of determining the relationship between the input and output of the unknown system in question, generally based on noisy output measurements. Assume we have an unknown system with input $x(t)$ and output $y(t)$, identification of this system is to estimate an expression for relation F such as (Nowak, 2002):

$$y(t) = F[x(t)] + e(t), \tag{4.6}$$

where $e(t)$ refers to noise on system output. When the relation F is a nonlinear operation, the task is named as *nonlinear system identification*.

Another point of view defines nonlinear system identification to select the best nonlinear model F from a set of candidate models $\mathcal{F}$. The selection criterion may be based on different metrics or cost functions, e.g. sum of squared errors. Thus, the identification process simply becomes a *model selection* operation from a set of models which minimizes the cost function (Nowak, 2002).

Up to this chapter of the thesis, we have demonstrated the estimation of models with polynomial nonlinearity by extending the methods derived for linear AR models. This model estimation process may also be referred simply as nonlinear model selection. Therefore, it is very clear that the RJMCMC usage we have shown in the previous chapters has a direct link between system identification process and can also be used in identifying a nonlinear system in terms of Volterra system models by making the necessary modifications.

In the literature, methods used to identify unknown systems provide information about the uncertainties, especially in terms of formulations or mathematical expressions, in case there is a little (or no) physical information about the systems. When the system subject to be identified is linear, SI methods are well established and easy to implement. However, it has been already stated that most of the real life applications exhibit nonlinear characteristics. Hence, approaches to the solution for a nonlinear SI problem can be challenging since the aforementioned systems may be highly nonlinear and the number of possible model structures might be very high (Green and Worden, 2015).

4.2. Trans-Structural RJMCMC for Volterra systems identification

In the previous chapter, we have stated that one of the main aims of this thesis is to show the formulation which Peter Green derived in (Green, 1995) offers wider interpretation of transitions between spaces which are not limited to trans-dimensional jumps. Thus, by using the reversible jump mechanism of Green, it is possible to explore spaces of the same dimensionality but with different structures, or both different dimensions and structures. Defining transitions between linear and nonlinear spaces may be given as an example for this type. In this chapter, transitions of this type are named as *trans-structural* and particularly, other than being trans-dimensional, trans-structural RJMCMC unveils the great potential of RJMCMC within deeper applications, i.e. across multiple dimensions simultaneously and between spaces of the same dimensionality.

Firstly, a state space $\mathcal{X} = \bigcup_k \{k\} \times \mathcal{R}^{n_k}$ is supposed to be the union of k subspaces which includes models with indicator k, $X_k = \{k\} \times R^{n_k}$ and each model can be defined as different types. Here, the term different types, is used for the cases e.g. linear and nonlinear models or models which are driven with different probability distributions, etc. Assume two subspaces $\mathcal{X}_1$ and $\mathcal{X}_2$ of different types where the dimensions n_1 and n_2 may be equal. The subspaces $\mathcal{X}_1$ and $\mathcal{X}_2$ have parameters spaces $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ and both have proper densities in $\mathcal{R}^{n_1}$ and $\mathcal{R}^{n_2}$.

Define a move type "m", to perform a transition from state $x \in \mathcal{X}_1$ to state $x' \in \mathcal{X}_2$, with probability p_m. A transition kernel is used to perform this transition which is applied in two steps as indicated in (2.10) and thus the detailed balance should be satisfied. Unlike the classical trans-dimensional approach, trans-structural RJMCMC transition between spaces of different types performs transition across multiple dimensions at once and may include both *birth* of new parameters and *death* of existing parameters at the same time. Hence, number of parameters, (or namely dimension of the parameter space) may be the same for both of the spaces. Briefly, transitions of this type switch the model spaces of different types, and hence will be named as **switch** moves in trans-structural RJMCMC concept.

In order to express the trans-structural RJMCMC approach more clearly, herein a toy example is given. In particular, two different Volterra models are implemented each of which has the same number of model coefficients but different structures. First model is a linear Volterra model say V(1,2), and the other one is nonlinear, say V(2,1). The general

expressions of the models from (4.1), are given below:

$$y(l) = h_1^{(1)} x(l-1) + h_2^{(1)} x(l-2), \tag{4.7}$$

$$y(l) = h_1^{(1)} x(l-1) + h_{1,1}^{(2)} x^2(l-1). \tag{4.8}$$

Then suppose, the data, $\mathbf{y}$, observed from one of the candidate models is given initially. For both of the candidate models two parameter subspaces $\mathcal{X}_1 = \{1\} \times \mathcal{R}^2$ and $\mathcal{X}_2 = \{2\} \times \mathcal{R}^2$ can be written easily from the definition above. The parameter subspaces for the models, V(1,2) and V(2,1) are $x = (1, h_1^{(1)}, h_2^{(1)}) \in \mathcal{X}_1$ and $x' = (2, h_1^{(1)}, h_{1,1}^{(2)}) \in \mathcal{X}_2$, respectively. A single between-model move is proposed which switches subspaces with probability p_m. The algorithm retains the same subspace with probability $1 - p_m$ and performs a within-model move to update model coefficients.

Particularly, in case of a transition from model V(1,2) to V(2,1), although the parameter dimensions are the same, just one of the model coefficients is common, say $h_1^{(1)}$. The other coefficient of the candidate model, say $h_{1,1}^{(2)}$, should be proposed and $h_2^{(1)}$ will be removed. For the reverse move, m^R, performing a transition from model V(2,1) to V(1,2) with probability, p_{m^R}, the same procedure will be applied, that $h_2^{(1)}$ will be proposed and $h_{1,1}^{(2)}$ will be removed. Consequently, we can easily write

$$\text{Move } m \rightarrow \hat{h}_1^{(1)} = h_1^{(1)}, \quad \hat{h}_{1,1}^{(2)} = u, \quad h_2^{(1)} \rightarrow \text{removed} \tag{4.9}$$

$$\text{Reverse move } m^R \rightarrow \hat{h}_1^{(1)} = h_1^{(1)}, \quad \hat{h}_2^{(1)} = u', \quad h_{1,1}^{(2)} \rightarrow \text{removed} \tag{4.10}$$

where coefficients with hats, e.g $\hat{h}_1^{(1)}$ refer to the coefficients of the candidate model and variables u and u' are proposed from the densities q_1 and q_2, respectively. Hence, the acceptance ratio for this transition appears as

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(x'|\mathbf{y}) p_{m^R} q_2(u')}{\pi(x|\mathbf{y}) p_m q_1(u)} \left| \frac{\partial(x', u')}{\partial(x, u)} \right| \right\}. \tag{4.11}$$

This derivation of a toy example constructed so far has been implemented in a computer simulation and in particular, RJMCMC has been utilized to decide the true model which the observed data $\mathbf{y}$ comes from. Both of the input and the output data sets are initially provided to the algorithm and the algorithm performs identification of the

unknown system. 100 distinct data sets from each model have been generated and examining the results of this simulation, for both of the models RJMCMC estimates the true models with %100 performance. Estimated joint histograms of model orders belonging to a single realization are shown in Figure 4.1 for both of the models.

4.2.1. Defining The Likelihood

When the input sequence is normally distributed, the output of a Volterra system is not guaranteed to be Gaussian due to the nonlinear operations performed on the input sequence. However, it was stated in a previous study (Ivanov, 1987) that output distribution of a narrowband Volterra system with white inputs is Gaussian. Hence, when the system memory of a Volterra system tends to infinity, its output is Gaussian.

Besides, in Bayesian approaches in SI studies, it is a widespread convention to express the likelihood as a measure how well the estimated model fits the observed data. Thus, in this study, we assume that the prediction error can be expressed as (Hadjidoukas et al., 2014)

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}. \tag{4.12}$$

where $\hat{\mathbf{y}} = [\hat{y}(1), \hat{y}(2), \dots, \hat{y}(n)]$ refers to the model prediction and the observed system output is $\mathbf{y} = [y(1), y(2), \dots, y(n)]$.

In previous studies (Green, 2015; Simoen et al., 2013; Hadjidoukas et al., 2014; Goulet and Smith, 2013), a zero mean Gaussian error-prediction model has been assumed. In Figure 4.2, distributions of the prediction error sequences for three example Volterra models which are implemented for the simulation case 1, are depicted. In order to provide a measure about the Gaussianity of the prediction error sequences, *Kullback-Leibler* (KL) divergence has been calculated with the fitted Gaussian distributions. Examining both the visual and the numerical results in the sub-figures of Figure 4.2, it can be easily stated that the prediction error distributions for all three Volterra models are Gaussian in terms of KL divergence values all of which are lower than 0.05.
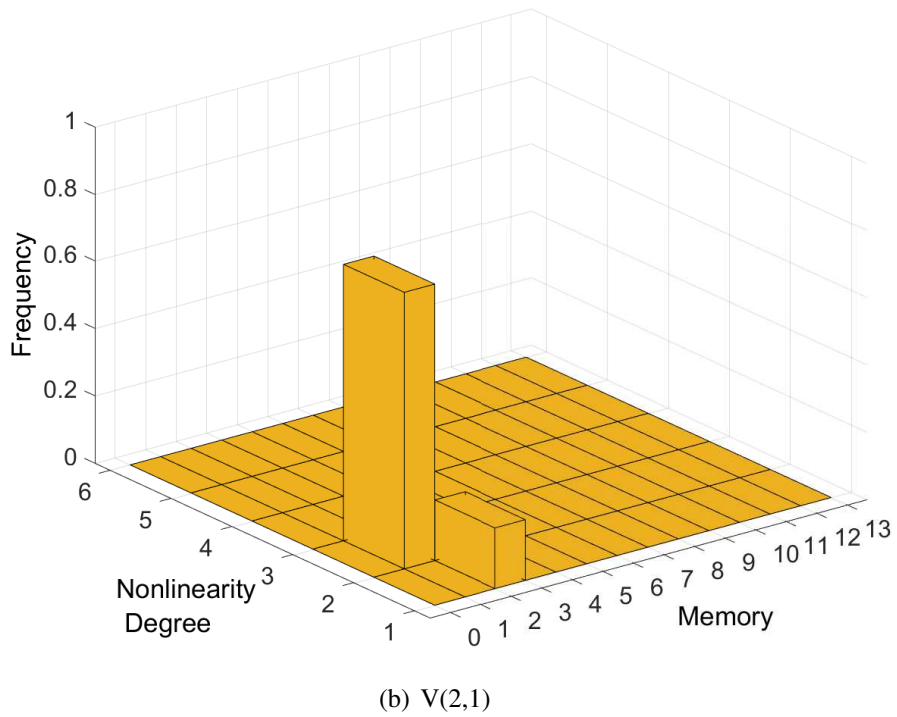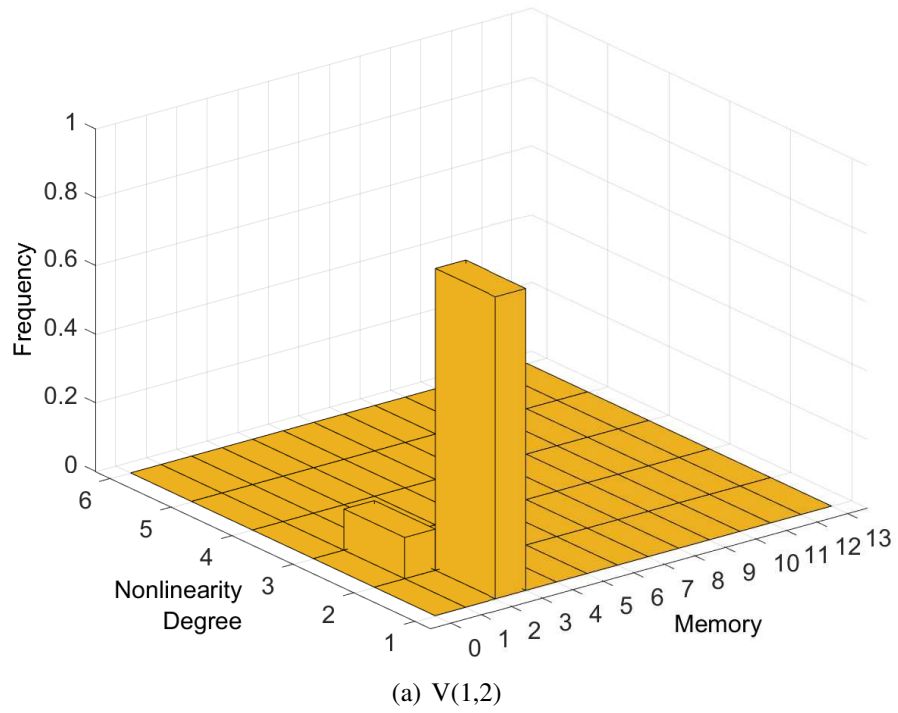
(a) V(1,2)

(b) V(2,1)

Figure 4.1. Toy example model estimation histograms - (a) V(1,2)   (b) V(2,1).
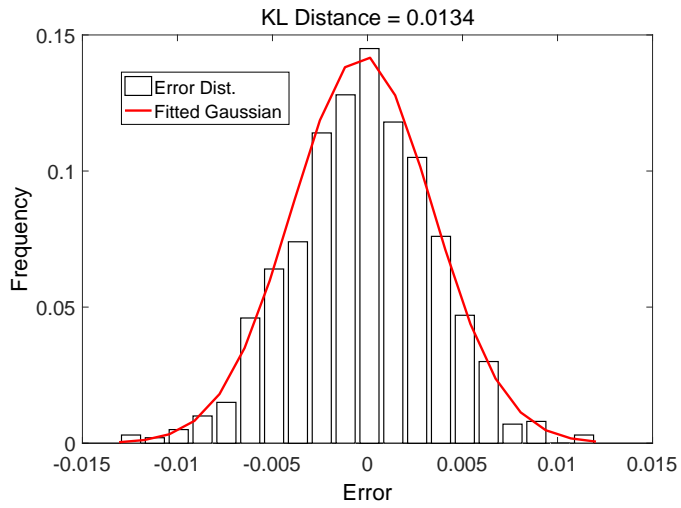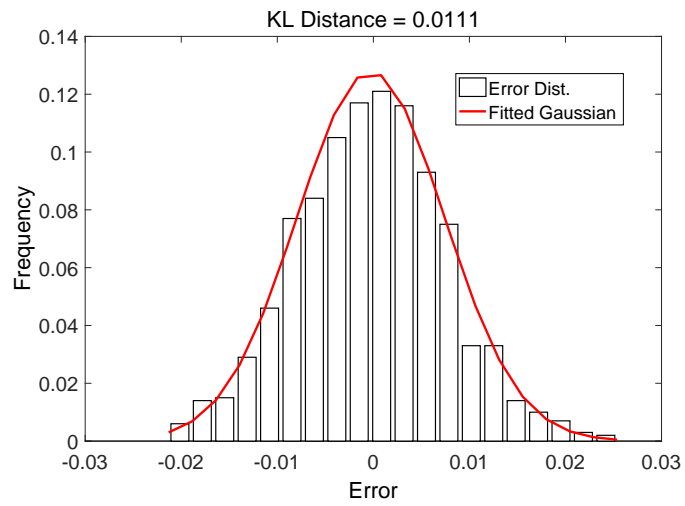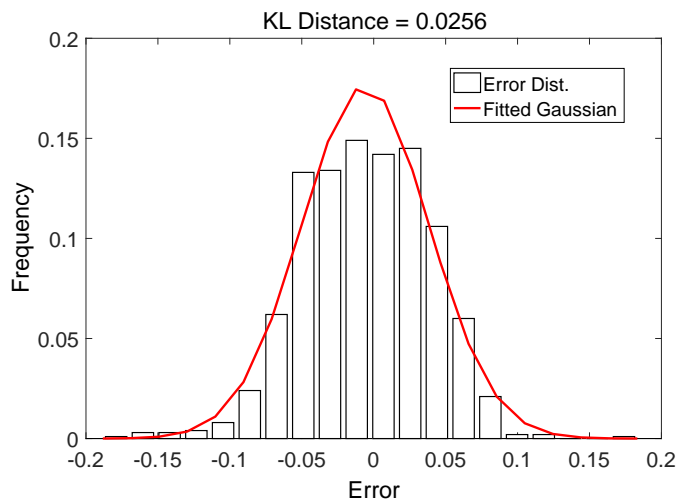
Figure 4.2. Prediction error histograms and fitted Gaussians for models - (a) V(1,10) (b) V(2,5)   (c) V(3,3).

Thus, using a Gaussian error-prediction model the likelihood function appears as

$$f(\mathbf{y}|\theta) = (2\pi\sigma_e^2)^{-n/2} \exp\left(\frac{1}{2\sigma_e^2} \sum_{t=1}^{n}(y_t - \hat{y}_t)^2\right) \qquad (4.13)$$

$$\approx \mathcal{N}(\mathbf{e}|\mathbf{0}, \sigma_e^2\mathbf{I}_n) \qquad (4.14)$$

where θ is the parameter vector consisting of $\{p, q, \mathbf{h}^{(p,q)}, \sigma_e^2, \sigma_h^2\}$, n refers to the data length, $\mathbf{e} = [e(1), e(2), \ldots e(n)]$ is the prediction error and σ_e^2 is the error variance.

4.2.2. Hierarchical Bayes Model

The joint posterior distribution of RJMCMC for VSI, $f(\theta|\mathbf{x})$, can be decomposed via Bayes theorem for the parameter vector $\theta = \{p, q, \mathbf{h}^{(p,q)}, \sigma_e^2, \sigma_h^2\}$ as

$$f(p, q, \mathbf{h}^{(p,q)}, \sigma_e^2, \sigma_h^2|\mathbf{y}) \propto f(\mathbf{y}|p, q, \mathbf{h}^{(p,q)}, \sigma_e^2)f(\mathbf{h}^{(p,q)}|p, q, \sigma_h^2)f(\sigma_h^2)f(\sigma_e^2)f(q)f(p). \quad (4.15)$$

4.2.3. Prior Selection

In the absence of real prior information, using noninformative priors is very common (Green and Richardson, 2001). To name a few: in time series model estimation studies using uniform priors for model orders has been a common choice (Troughton and Godsill, 1998; Ehlers and Brooks, 2004; Eğri et al., 2010). Moreover, results which are obtained by using uniform priors can be easily converted to other priors, that is achieved by using the identity (Richardson and Green, 1997)

$$f^*(k, \theta^{(k)}|y) \propto f(k, \theta^{(k)}|y)\frac{f^*(k)}{f(k)} \qquad (4.16)$$

where $f^*(\cdot|y)$ refers to the posterior density for the prior f^*.

Consequently, we are assuming that the model orders are independent and each model is equiprobable with upper bounds p_{max} and q_{max} for model orders p and q, respec-

tively. Therefore, uniform priors are chosen for the model memory q and the nonlinearity degree p

$$f(q) = \mathcal{U}(1, q_{\max}) \quad \text{and} \quad f(p) = \mathcal{U}(1, p_{\max}). \tag{4.17}$$

Volterra model coefficients are assumed to be normally distributed a priori and for the variances, σ_e^2 and σ_h^2, we use conjugate priors (Ehlers and Brooks, 2004)

$$f(\mathbf{h}^{(p,q)}|p, q, \sigma_h^2) = \mathcal{N}(\mathbf{h}^{(p,q)}|\mathbf{0}, \sigma_h^2 \mathbf{I}_\eta), \tag{4.18}$$

$$f(\sigma_h^2) = \mathcal{IG}(\sigma_h^2|\alpha_h, \beta_h), \tag{4.19}$$

$$f(\sigma_e^2) = \mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e). \tag{4.20}$$

4.2.4. RJMCMC Methodology for VSI

In this thesis, in Chapter 2, we have presented the trans-dimensional RJMCMC derivation for linear AR, which is frequently discussed in the literature. This derivation has been extended in Chapter 3 and been utilized to estimate models with polynomial nonlinearity. These nonlinear models, PAR/PMA/PARMA, are linear-in-the-parameters models and follow the Volterra series expansion. Although the stochastic modeling and system identification procedures differ from each other in their essence, the approach derived in Chapters 2 and 3 can also be used to identify the Volterra system models.

So, as mentioned in previous chapters, when we develop RJMCMC for VSI, we propose a new RJMCMC move that we have not used before. This move is called a switch move, and the classical trans-dimensional birth and death moves used in the previous chapters are combined into a single move to perform transitions across multiple dimensions. In addition, switch move is one of the most important contributions in this thesis as it offers to switch between the same size but different structural models.

Remark 4.1 *Please note that the same dimensional transitions of this type cannot be defined as a special case. On the contrary, as shown clearly in the toy example above, the dimension of the parameter spaces during a transition can be the same, since it provides transitions across multiple parameter dimensions simultaneously (e.g. V(1,2) → V(2,1))*

and since it performs birth and/or death moves at the same time. In this case, even though the total space size is the same between different models, the reversible jump mechanism can still be used. For this reason, switch move should be seen as a much more general move which also includes trans-dimensional moves. This situation can be obviously seen in Figure 4.3.

Figure 4.3 depicts the difference between the trans-structural approach presented in this chapter and the trans-dimensional approach in terms of RJMCMC model moves. As can be clearly seen from the Figure 4.3, the classic birth and death steps can be performed in a switch move, as well as the transitions between different models of the same size, which have not been used for RJMCMC in previous studies. The general RJMCMC implementation steps for VSI study is given in Algorithm 5.
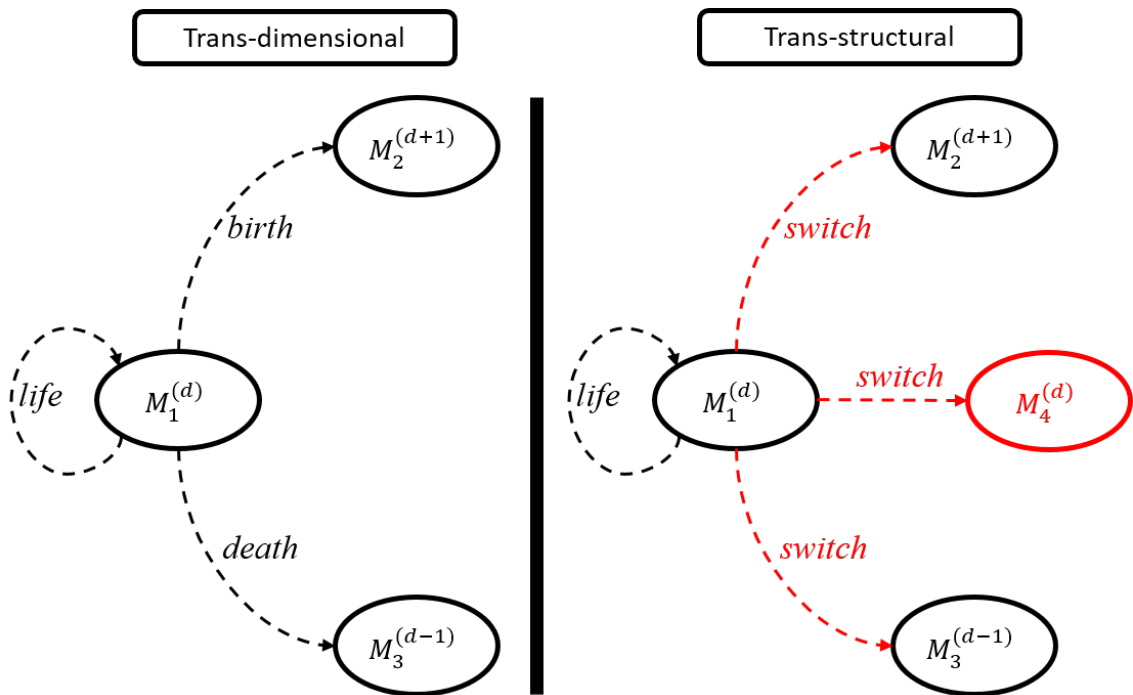
Figure 4.3. Trans-structural vs. trans-dimensional RJMCMC. $M_i^{(d)}$ refers to ith model with d-dimensional parameter space.

4.2.4.1. Switch Move

Between-model move, or namely switch move, explores the spaces of different Volterra models at each time it is proposed. Candidate models may have different struc-

---
Algorithm 5 RJMCMC Algorithm for VSI

1: Given I/O data sets **x**
2: Initialize parameters for iteration $t = 0$
3: **for do** $i = 1 : N_{iter}$
4:     *Choose* Life or Switch move with probabilities $P_{\text{life}}, P_{\text{switch}}$.
5:     **if** Switch is selected **then**
6:         *Sample* candidate model orders, p' and q'.
7:         *Perform Switch* move to *calculate* acceptance ratio, $A = \alpha_{\text{switch}}$.
8:     **elseif** Life is selected **then**
9:         *Candidate* model orders $\rightarrow p' = p^{(i-1)}, \quad q' = q^{(i-1)}$.
10:         *Perform Life* move in Section 5.3.5.1 to *calculate* acceptance ratio, $A = \alpha_{\text{life}}$.
11:     **end if**
12:     *Sample* a random variable $R \sim \mathcal{U}(0, 1)$
13:     **if** $R \leq A$ **then**
14:         $p^{(i)} = p'$ and $q^{(i)} = q'$.
15:         *Update* other parameters with proposed ones.
16:     **else**
17:         $p^{(i)} = p^{(i-1)}$ and $q^{(i)} = q^{(i-1)}$.
18:         *Do not change* the values of other parameters.
19:     **end if**
20:     *Perform Update* move in Section 4.2.4.3 to update variances, σ_e^2 and σ_h^2.
21: **end for**
---

tures and their space dimension may be different or the same.

The acceptance ratio for a **switch** move from (p, q) to (p', q'), is defined as $\alpha_{\text{switch}} = \min\{1, r_{\text{switch}}\}$ where r_{switch} is expressed as:

$$r_{\text{switch}} = \frac{f(\mathbf{y}|p', q', \mathbf{h}^{(p',q')}, \sigma_e^2)}{f(\mathbf{y}|p, q, \mathbf{h}^{(p,q)}, \sigma_e^2)} \times \frac{f(\mathbf{h}^{(p',q')}|p', q', \sigma_h^2)}{f(\mathbf{h}^{(p,q)}|p, q, \sigma_h^2)} \times \frac{\chi(\mathbf{u}')}{\chi(\mathbf{u})} \times \left| \frac{\partial(\mathbf{h}^{(p',q')}), \mathbf{u}')}{\partial(\mathbf{h}^{(p,q)}, \mathbf{u})} \right|. \quad (4.21)$$

where $\chi(\cdot)$ is the proposal distribution for the auxiliary variables and will be defined in (4.31). After performing switch move, to turn back to the previous state another switch move should be proposed. This means that the reverse move of the switch move is itself. Thus, the ratio for the move probabilities in the original formulation is equal to 1 and invisible in (4.21).

In particular, in (4.21) first two terms refer to the likelihood and the prior ratios, respectively. Proposal ratio is the third term whereas the magnitude of the Jacobian is the fourth term.

4.2.4.2. Life Move

The proposed and accepted coefficients within the previously performed switch moves, are updated in the **life** move. In cases when the algorithm remains at the same model, a life move is performed. Acceptance ratio of the life move is $\alpha_{\text{life}} = \min\{1, r_{\text{life}}\}$ and hence, r_{life} is expressed as

$$r_{\text{life}} = \frac{f(\mathbf{y}|p, q, \widehat{\mathbf{h}}^{(p,q)}, \sigma_e^2)}{f(\mathbf{y}|p, q, \mathbf{h}^{(p,q)}, \sigma_e^2)} \times \frac{f(\widehat{\mathbf{h}}^{(p,q)}|p, q, \sigma_h^2)}{f(\mathbf{h}^{(p,q)}|p, q, \sigma_h^2)} \times \frac{\psi(\mathbf{h}^{(p,q)}|p, q, \widehat{\mathbf{h}}^{(p,q)})}{\psi(\widehat{\mathbf{h}}^{(p,q)}|p, q, \mathbf{h}^{(p,q)})}. \tag{4.22}$$

Updating procedure in this type of move is performed via sampling from the distribution $\psi(\cdot)$

$$\widehat{\mathbf{h}}^{(p,q)} \sim \psi(\widehat{\mathbf{h}}^{(p,q)}|p, q, \mathbf{h}^{(p,q)}) \tag{4.23}$$

$$= \mathcal{N}(\widehat{\mathbf{h}}^{(p,q)}|\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n^{-1}) \tag{4.24}$$

where $\boldsymbol{\mu}_n = \sigma_e^{-2}\boldsymbol{\Sigma}_n^{-1}\mathbf{X}^T\mathbf{y}$ and $\boldsymbol{\Sigma}_n = \sigma_e^{-2}\mathbf{X}^T\mathbf{X} + \sigma_h^{-2}\mathbf{I}_\eta$.

4.2.4.3. Update Move - Updating Variances

As stated in Section 4.2.1, VSI mechanism in this chapter includes an error term. The variance of this error term, σ_e^2 is updated at each iteration via Gibbs Sampling. The full conditional distribution for σ_e^2 is constructed as (Troughton and Godsill, 1998)

$$f(\sigma_e^2|\mathbf{y}, p, q, \mathbf{h}^{(p,q)}) \propto f(\mathbf{y}|p, q, \mathbf{h}^{(p,q)}, \sigma_e^2)f(\sigma_e^2) \tag{4.25}$$

$$\approx \mathcal{N}(\mathbf{e}|\mathbf{0}, \sigma_e^2\mathbf{I}_n)\mathcal{IG}(\sigma_e^2|\alpha_e, \beta_e) \tag{4.26}$$

$$= \mathcal{IG}(\sigma_e^2|\alpha_{en}, \beta_{en}), \tag{4.27}$$

where $\alpha_{en} = \alpha_e + \frac{1}{2}n$ and $\beta_{en} = \beta_e + \frac{1}{2}\mathbf{e}^T\mathbf{e}$.

It is straightforward to derive the full conditional distribution for σ_h^2 as (Troughton

and Godsill, 1998)

$$f(\sigma_h^2|\mathbf{y}, p, q, \mathbf{h}^{(p,q)}) \propto f(\mathbf{h}^{(p,q)}|\sigma_h^2)f(\sigma_h^2) \tag{4.28}$$

$$\approx \mathcal{N}(\sigma_h^2|\mathbf{0}, \sigma_h^2\mathbf{I}_\eta)\mathcal{IG}(\sigma_h^2|\alpha_h, \beta_h) \tag{4.29}$$

$$= \mathcal{IG}(\sigma_h^2|\alpha_{hn}, \beta_{hn}) \tag{4.30}$$

where $\alpha_{hn} = \alpha_h + \frac{1}{2}\eta$ and $\beta_{hn} = \beta_h + \frac{1}{2}(\mathbf{h}^{(p,q)})^{\mathrm{T}}\mathbf{h}^{(p,q)}$ and η is the number of parameters as defined in (4.5).

4.2.4.4. Proposing Candidates

Each RJMCMC iteration requires to select one of the switch or life move firstly with probabilities $\mathrm{P_{switch}}$ and $\mathrm{P_{life}}$. Uniform prior is selected for all candidate switchable models (with probability $\mathrm{P_{switch}}/\rho$ for ρ possible models).

An *adhoc* selection is made based on the data observed at the system output for the proposal distribution, $\chi(\mathbf{u})$. A vector of the candidate coefficients with a length of λ is proposed from this proposal distribution which is assumed to be a multivariate Gaussian distribution

$$\chi(\mathbf{u}) = \mathcal{N}\left(\mathbf{0}, \left(\frac{\sigma_h^2}{\zeta E[|\mathbf{y}|]}\right)I_\lambda\right) \tag{4.31}$$

where $E[|\mathbf{y}|]$ is the expected value of the absolute value of the data vector $\mathbf{y}$ and ζ is the modulation constant.

A selection of this kind is not a unique selection for the proposal distribution. Interested users can choose a uniform distribution and the like, which may change the mixing of the algorithm. We have tested several selections and decided to choose a proposal distribution of this kind. The data dependency comes from the variance of this proposal distribution which is chosen to depend on the data. In experimental analysis, there are two distinct simulations (Simulation 1 and 2) the first of which has four cases (See Table 4.2) in order to test performance of the algorithm under different noise effects. Thus, to take these various additional noise sequences into consideration in the proposals, $E[|\mathbf{y}|]$ has been used in $\chi(\cdot)$. Furthermore, in the case study of Simulation 2, which

is about nonlinear communication channel estimation, various modulation schemes have been used. This information is also used in $\chi(\cdot)$ via parameter ζ. This data and modulation dependent *adhoc* choice adds variety and provides an efficient sampling procedure for the candidates. The modulation constant ζ is selected as $\log_2(M)$ for M-ary modulations (Simulation 2) whereas is equal to 1 for no modulation case (Simulation 1).

Additionally, the transition between the most recent model coefficient space to the candidate space is chosen to be performed through an identity function which makes the Jacobian equals to unity. An example for a model transition from V(1, 2) to V(1,3) which is required to sample a single candidate coefficient, u is shown as

$$\hat{h}_1^{(1)} = h_1^{(1)} \tag{4.32}$$

$$\hat{h}_2^{(1)} = h_2^{(1)} \tag{4.33}$$

$$\hat{h}_3^{(1)} = u. \tag{4.34}$$

4.3. Experimental Analysis

In this section, an experimental analysis to study the performance of the proposed method is performed. In Figure 4.4, the block diagram of the proposed Bayesian VSI scheme has been depicted for a system input and output sequences which are given as **x** and **y**, respectively. In addition, the additive noise sequences for input and output are **u** and **w**, respectively.

Estimated model order parameter pair $(\widehat{p}, \widehat{q})$ and the resulting model coefficient vector $\widehat{\mathbf{h}^{(p,q)}}$ have been utilized to perform 1-step ahead prediction for the output of the system, $\widehat{\mathbf{y}}$, by using the Volterra model expression in (4.1).

4.3.1. Simulation 1: Synthetically Generated Data

The proposed method has been studied for synthetically generated data sets in simulation case 1. Particularly, three Volterra models, V(1, 10), V(2, 5), V(3, 3) have been implemented, model coefficients of which are shown in Table 4.1. The input sequence is a Gaussian process of zero mean and unity variance for all the models and outputs with lengths of 1000 samples from each model have been collected. The intercept, μ, is chosen
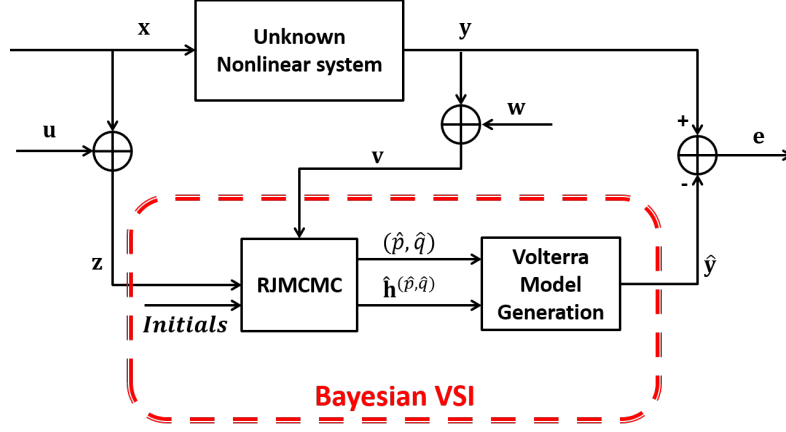
Figure 4.4. The proposed method VSI block diagram.

as 0 for simplicity. Four cases have been employed to study the performance of the RJMCMC based VSI procedure under different noise conditions (See Table 4.2).

Table 4.1. Details for Volterra models in Simulation 1

$V(p,q)$	$\mathbf{h}^{(p,q)} = [\mathbf{h}^{(1)}, \quad \mathbf{h}^{(2)}, \quad \dots, \quad \mathbf{h}^{(p)}]^{\mathrm{T}}$	Calculated SNR(dB) values[(*)]
$V(1,10)$	$\mathbf{h}^{(1)} = [0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5]$	14.13/22.62/10.42/14.19
$V(2,5)$	$\mathbf{h}^{(1)} = [0.7, 0, 0.2, 0, -0.7]$ $\mathbf{h}^{(2)} = [0, 0.1, 0, 0, -0.25, 0.15, 0, 0.42, 0.02, 0, 0.7, 0, -0.31, 0, 0.28]$	13.52/22.24/10.42/13.58
$V(3,3)$	$\mathbf{h}^{(1)} = [-0.06, 0.2331, -1.3619]$ $\mathbf{h}^{(2)} = [0, 0.7, 0, 0.3, -0.25, 0.15]$ $\mathbf{h}^{(3)} = [0.5, 0, 0, -0.44, 0.15, -0.25, 0, -0.37, 0, 0.58]$	17.69/26.33/10.44/17.77

[(*)] Calculated SNR values in dB are presented for Case 2/Case 3/Case 4-Input/Case 4-Output, respectively.

Initial values for hyperparameters of prior distribution of σ_e^2, are selected as $\alpha_e = 1$ and $\beta_e = 1$ and those for σ_h^2, are selected as $\alpha_h = 35$ and $\beta_h = 2$. The initial model orders for p_0 and q_0 are set to 1 and upper bounds p_{max} and q_{max} are set to 5 and 12, respectively. Model moves are selected as equally likely, and thus P_{switch} and P_{life} are both 0.5. Calculated *signal-to-noise ratio* (SNR) values in decibels for each model and each case has been depicted in Table 4.1.

Model order estimation performance of RJMCMC is compared to two widely used model order selection methods, which are AIC and BIC. In particular, AIC rewards goodness of fit, however, penalizes the number of estimated parameters of the model. On the other hand, BIC is more informed then AIC and its penalty term is more stringent than

that of AIC. So, BIC tends to choose smaller models than AIC.

RJMCMC has also a similar penalization when the algorithm tries to add redundant variables. For example, increasing order by one and setting the additional coefficient to zero does not change the likelihood, but the prior takes a lower value than before, yielding a posterior probability lower than the previous one (Meyer-Gohde et al., 2014).

Table 4.2. Cases for Simulation 1

	Details
Case 1	Both I/O are noise free
Case 2	Output is corrupted by a white Gaussian noise process of mean 0 and variance 0.1
Case 3	Output is corrupted by a colored Gaussian noise process. The white noise in Case 2 is filtered by an FIR filter, and the output of the filter is used to corrupt the output.
Case 4	Both I/O are corrupted by white Gaussian noise processes of mean 0 and variance 0.1

Table 4.3. Percentage of detecting correct model orders

		Case 1	Case 2	Case 3	Case 4
	RJMCMC	**100%**	**100%**	**100%**	**100%**
V(1,10)	AIC	99%	84%	89%	76%
	BIC	**100%**	**100%**	**100%**	**100%**
	RJMCMC	**100%**	99%	**100%**	**93%**
V(2,5)	AIC	93%	68%	85%	0%
	BIC	99%	**100%**	**100%**	11%
	RJMCMC	**100%**	**100%**	**100%**	**89%**
V(3,3)	AIC	98%	83%	93%	0%
	BIC	99%	**100%**	**100%**	13%

Table 4.3 presents the model selection percentages of RJMCMC and the reference methods AIC and BIC after 100 realizations. In each RJMCMC realization the most visited model after burn-in period is decided to be the detected model. Examining Table 4.3, AIC always falls short of selecting the true model order pair as compared to that of RJMCMC and BIC. RJMCMC and BIC get generally the same percentages, however, for the nonlinear models (V(2,5) and V(3,3)), RJMCMC performs better. Especially,

for case 4, performance of RJMCMC is superior for nonlinear models and the detection performance is at least %89, whereas BIC achieves at most %13 for the same models.

Figure 4.5 shows the joint posterior density of the model orders, p and q for randomly selected cases in a single example realization. It is easy to state that RJMCMC estimates true model orders higher than %50 for each example realizations.

In addition to model order selection performance, next we compare the success of RJMCMC in estimating model coefficients to NLS estimate which is obtained via the augmented data matrix $\mathbf{X}$. NLS has been given the correct model orders p, and q and performs estimation

$$\hat{\mathbf{h}}_{\mathrm{NLS}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}, \qquad (4.35)$$

where vector $\mathbf{y}$ refers to the output and $\mathbf{X}$ has the form defined in (4.4). The performance comparison is based on the model coefficient estimation of RJMCMC and NLS in terms of error measure, NMSE.

Table 4.4. Performance comparison of model coefficient estimation in terms of NMSE

		Case 1	Case 2	Case 3	Case 4
V(1,10)	RJMCMC	5.89E-07	2.36E-06	2.47E-06	1.43E-03
	Informed NLS	2.42E-09	8.46E-07	7.86E-07	1.26E-03
V(2,5)	RJMCMC	6.76E-08	2.06E-05	1.12E-07	1.42E-03
	Informed NLS	8.42E-09	1.93E-05	7.73E-08	1.32E-03
V(3,3)	RJMCMC	1.69E-04	1.84E-04	1.74E-04	6.07E-03
	Informed NLS	6.76E-08	2.28E-07	3.90E-08	3.46E-03

In Table 4.4, coefficient estimation for all models and all cases are depicted in terms of NMSE values. According to the results in Table 4.4, coefficients estimation with NLS has slightly lower error values than RJMCMC for all the cases. However, the NMSE figures of NLS are hypothetical since they are based on unavailable perfect model order estimates. So, the performance of the proposed method is remarkable since it performs estimation on model orders and coefficients at the same time.

Figure 4.6 demonstrates the estimated output data histograms for each case and each synthetically generated Volterra model data. Examining each subplot in Figure 4.6 shows that real data means stand in the high probability ranges of estimated data distributions.
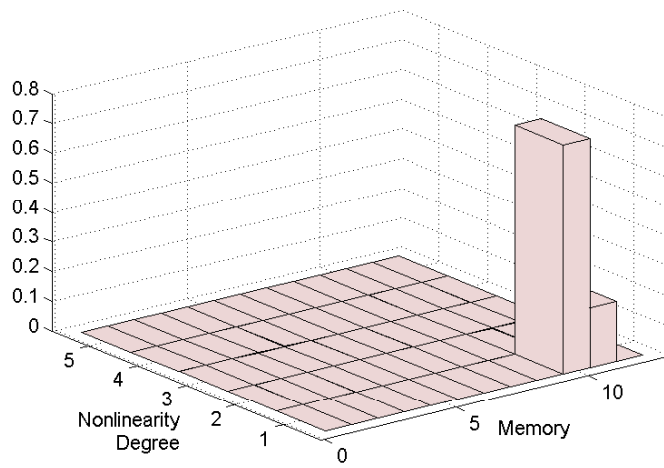
(a) V(1,10) - Case 2
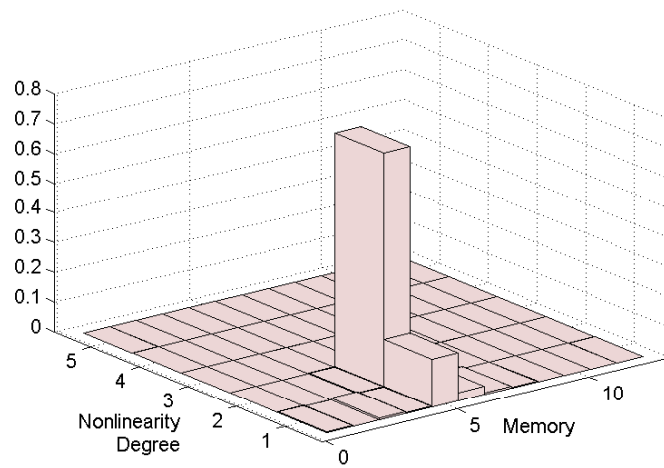
(b) V(2,5) - Case 3
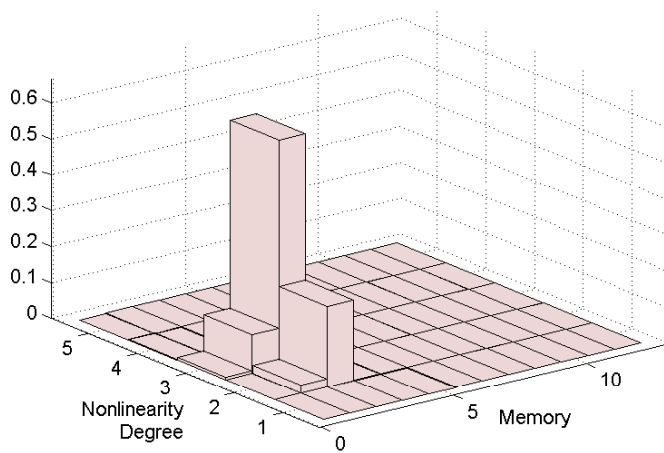
(c) V(3,3) - Case 4

Figure 4.5. The joint posterior density of the model orders of (a) - V(1,10), (b) - V(2,5), (c) - V(3,3).
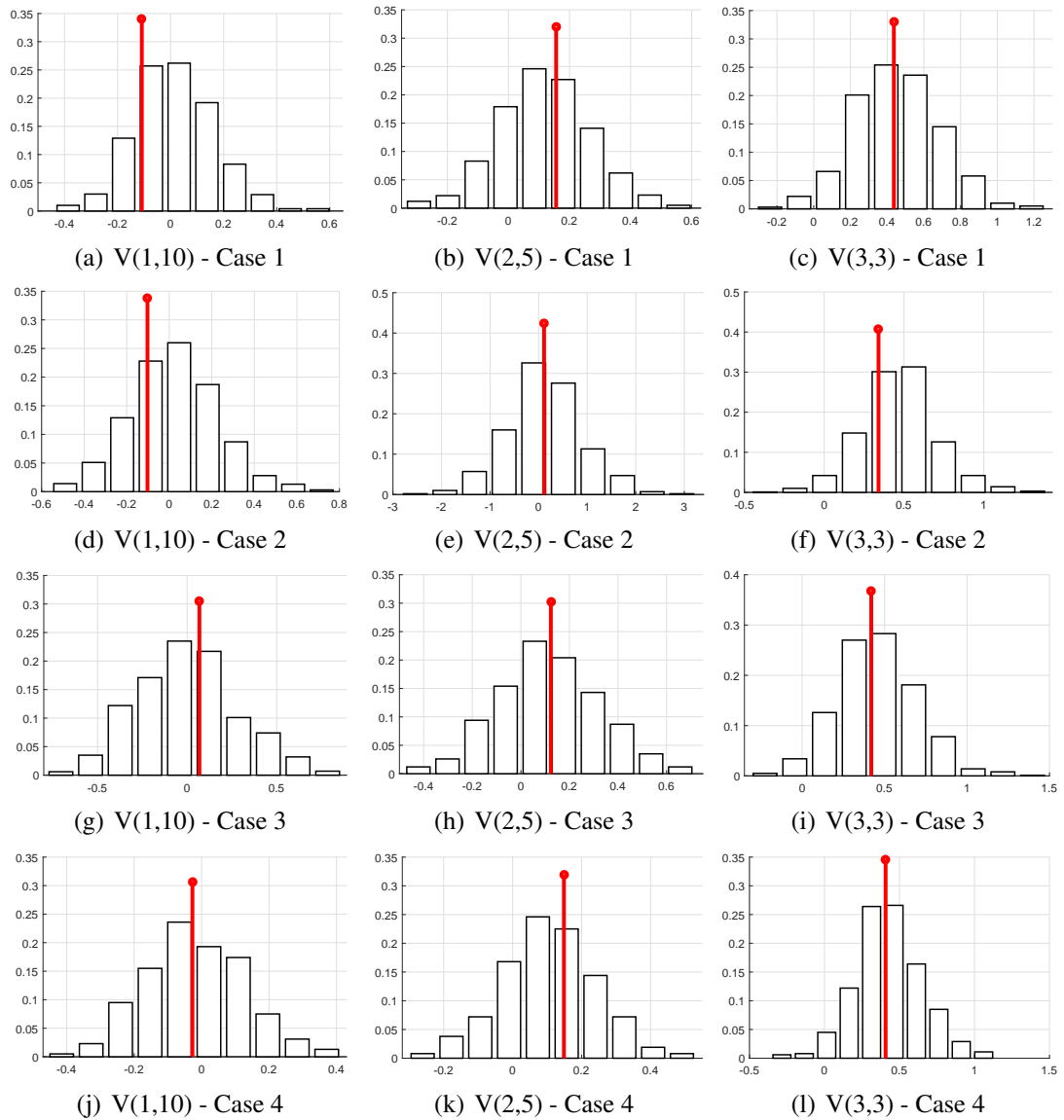
Figure 4.6. Estimated output histograms for all cases and all models in simulation 1 via RJMCMC. Real data mean values are plotted with a vertical line. Each column shows the results for simulated models and each row shows the results for simulation cases.

Table 4.5. RJMCMC Computational Gain

| | Case 1 | | Case 2 | | Case 3 | | Case 4 | |
	Total*	Avg.**	Total*	Avg.**	Total*	Avg.**	Total*	Avg.**
V(1,10)	16	12.37	18	12.65	16	12.31	17	12.51
V(2,5)	20	10.22	15	9.08	20	10.98	20	13.3
V(3,3)	18	8.11	20	8.06	19	8.5	26	9.79

Each RJMCMC run has performed 30000 iterations, and number of visited Volterra models has been recorded for each run.
* *Numbers at **Total** cells represent the total number of distinct Volterra models visited after 100 RJMCMC runs.*
** *Numbers at **Avg.** cells refer to the average number of Volterra models visited at a single run after 100 RJMCMC runs.*
Model space includes 60 Volterra models.

As mentioned before, RJMCMC avoids performing exhaustive searches in its learning process, instead searches model space by using the likelihood, the priors and the data by visiting only plausible models. Calculations on computational gain of RJM-CMC for Simulation 1 are shown in Table 4.5. Firstly, "Total" columns show that at most 50% of the candidate models (in all the cases these are including the correct models) have been visited. Then, analysing the "Avg." columns show that the number of models in search subset is less than the total amount and thus it is easily stated that RJMCMC decides "true model" by examining at most only one fifth of the model space (at most 12-13 models over 60 possible models). Hence, this exhibits the computational gains of RJMCMC compared with the other model selection methods AIC, BIC or the sampling algorithms such as nested sampling, TMCMC, etc. where all perform exhaustive searches on model space.

4.3.2. Simulation 2: Nonlinear Channel Estimation

Due to high-power amplifiers at the transmitter side and filtering operations at the receiver side, in many communication systems it is very common to observe nonlinear input-output characteristics, most of which can be approximated via Volterra series. A nonlinear communication channel is expressed in terms of discrete time baseband Volterra

model with symmetric coefficients as (Bouilloc and Favier, 2012; Mileounis et al., 2009)

$$y(l) = \sum_{v=1}^{\frac{p+1}{2}} \sum_{m_1=1}^{q} \cdots \sum_{m_{2v-1}=m_{2v-2}}^{q} h_{m_1,\dots,m_{2v-1}}^{(2v-1)} \prod_{i=1}^{v} x(l-m_i) \prod_{j=v+1}^{2v-1} x^*(l-m_j) \qquad (4.36)$$

where $x(l)$ and $y(l)$ refer to the complex input and output envelopes of the system, p and q are the nonlinearity degree and the memory of the channel, respectively. The $(2v - 1)$st-order Volterra coefficient is represented as $h_{m_1,\dots,m_{2v-1}}^{(2v-1)}$. In addition, in (Benedetto and Biglieri, 1999) it is expressed that powers of even-ordered terms do not contribute to the output. Thus, only odd-ordered terms $\{p = 1, 3, \dots\}$ are taken into account for baseband Volterra representation in (4.36).

OFDM technique is very widespread in many modern communication systems such as *asymmetric digital subscriber line* (ADSL) modems, *digital video broadcasting* (DVB) and recent mobile communication systems in 4G. Despite this popularity, OFDM is very vulnerable to nonlinearities due to its high peak-to-average power ratio (Mileounis et al., 2009). For these reasons, an OFDM communication system which transmits through a nonlinear communication channel has been implemented. Then, the proposed Bayesian VSI model has been used to perform a nonlinear channel estimation in terms of Volterra series.

In particular, to study the performance, a baseband Volterra model in (4.36) is assumed to represent the unknown nonlinear communication channel with nonlinearity degree of 3 and memory of 2. Message sequences are uniformly distributed and modulated via M-QAM modulations for $M = 4, 16, 64$ (4QAM will be denoted as QPSK for the rest of the text). Following modulation, symbols are sent through an OFDM system with 512 sub-carriers. Resulting symbols have been parallel-to-serial converted and transmitted through the nonlinear channel. After adding white Gaussian noise, the transmitted corrupted signal has been received at the receiver.

Pilot messages have been used to apply a VSI procedure. Thus, both pilot OFDM output and the corrupted received signal are known at the receiver as input and output of the unknown system, respectively. Performing the procedure expressed in the previous sections, RJMCMC estimates the nonlinearity degree, the system memory and the corresponding channel coefficients. Initial hyperparameters are selected as $\alpha_e = 1$, $\beta_e = 1$, $\alpha_h = 35$ and $\beta_h = 2$. The initial orders are $q_0 = 1$ and $p_0 = 1$ and the upper bounds are $q_{max} = 12$ and $p_{max} = 5$. Performance of the proposed method is studied under different additive noise conditions with *symbol-to-noise ratio* (E_s/N_0) values between -5 dB and

25 dB. In a single RJMCMC run, 20000 iterations are performed and a total of 100 Monte Carlo runs have been simulated. Results are presented as average of these repetitions in order to remove random realization effects. The simulated nonlinear channel coefficients are $h_1 = [0.5, \quad 0.3]$ and $h_3 = [-0.7, \quad -0.2, \quad 0.34, \quad -0.27]$ for linear and cubic terms of the baseband Volterra model in (4.36), respectively.

Percentage of correctly estimated model orders for varying E_s/N_0 values are shown in Figure 4.7. Analyzing Figure 4.7, when E_s/N_0 is higher than 0 dB, the true nonlinear channel, V(3,2) is correctly estimated by RJMCMC, with a remarkable performance of at least 99% after 100 RJMCMC runs for all the modulations while below 0 dB estimation performance is at least 85% times of the repetitions.

In Figure 4.8 the log-scale NMSE values for the estimated channel coefficients are depicted in increasing order of E_s/N_0. Examining Figure 4.8 shows that NMSE values are very close for both Informed NLS method and RJMCMC for all the cases. For lower E_s/N_0 values, estimation performances are lower as expected and NMSE values are around 10^{-3} for E_s/N_0 of 0 dB, when NMSE values for all the cases are below 10^{-5} at E_s/N_0 of 25 dB.
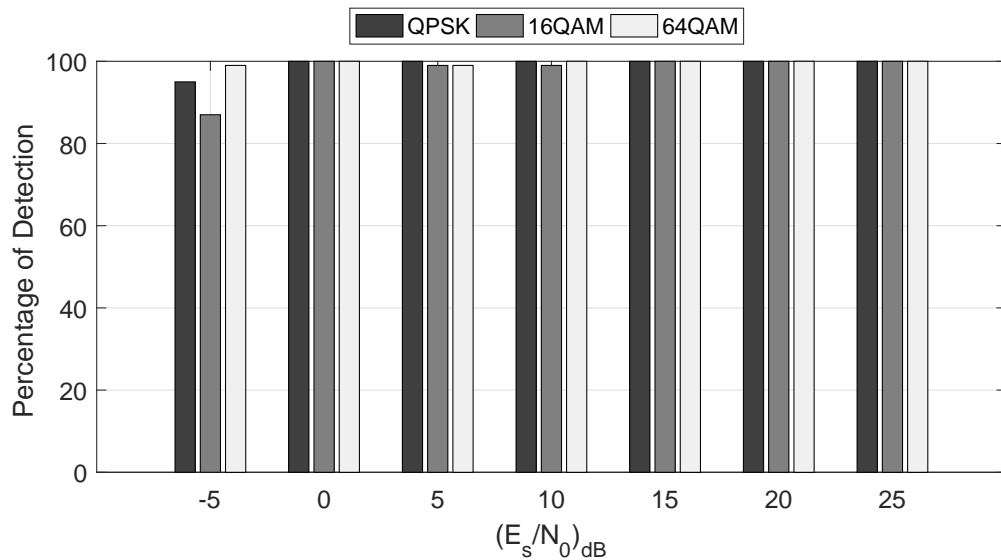
Figure 4.7. Percentage of correctly estimated model order via RJMCMC for varying E_s/N_0 (Nonlinear channel, V(3,2)).
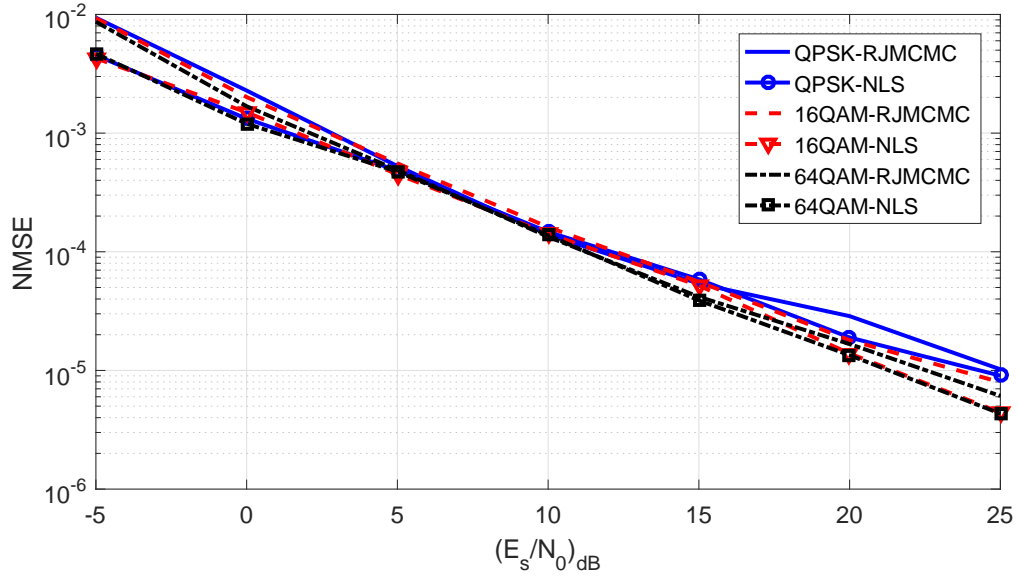
Figure 4.8. NMSE values for estimation of channel coefficients of RJMCMC & In-
formed NLS for QPSK, 16-QAM and 64-QAM modulation schemes.

4.4. Conclusions on VSI

In this chapter, the proposed method, the trans-structural RJMCMC (or trans-space RJMCMC for different structural model), is used for the study of the Volterra system models, and in summary the following results are obtained:

- This methodology has provided a compact method for VSI which estimated the nonlinearity degree of the unknown system as well as the system memory and model coefficients.

- In cases when the systems in question have varying degree of nonlinearities or no prior information regarding nonlinearity degree is available, using the proposed method in this chapter provides solution by estimating the nonlinearity degree of the unknown systems in terms of Volterra systems.

- RJMCMC shows remarkable performance on nonlinear channel estimation in an OFDM communication system. Performance results for QPSK, 16-QAM and 64-QAM, are satisfactory for both channel model selection and coefficient estimation studies.

- These results demonstrate the potential of RJMCMC in identifying nonlinear sys-

tems and nonlinear communication channels in terms of Volterra models.

Up to this chapter, the classical formulation of RJMCMC has been shown to be able to switch between different generic models by applying the necessary modifications. With the approach which has been presented in this chapter, the classical interpretation of the RJMCMC has been extended one step further. This very important extension also incorporates a much more general RJMCMC move, which is newly proposed and includes the classic trans-dimensional RJMCMC moves, e.g. birth and death, as its sub-moves.

Up to this point, the objectives mentioned in the introduction part of this thesis have been discussed through the Volterra models. In particular, the linear/nonlinear model transitions and estimating the nonlinearity degree, which is generally assumed to be known in the literature, clearly shows the importance of the applications so far. The methods developed so far, have shown very clearly that RJMCMC can be used for transitions between more general and completely different spaces, apart from structurally different transitions such as linear and nonlinear.

CHAPTER 5

DISTRIBUTION MODELLING

In various real-life modelling problems, we have limited prior information regarding which model family is more suitable for the problem. In such cases, a method that would allow one to choose between different model families on the fly would be useful and thus would eliminate the need for modelling with each candidate model class separately and comparing to find the best. Such a method provides computational gains especially when the number of parameters and candidate model classes is high. An example is the choice between different *probability density function* (pdf) models for noise or signals.

The pdf estimation problem is a frequently encountered problem in signal processing and statistics and their application fields such as image processing and telecommunications. In communication systems, channel modelling has been an important issue so as to characterize the whole system. However, for most of the cases, performing a deterministic channel modelling might be impossible and to represent real life systems, statistical channel models are very important. In addition, in applications of noise reduction operations in image processing, power-line communication systems, etc. dealing with a suitable statistical model beforehand is also important for the methods to be developed. Despite this importance, estimating the correct (or suitable) probability distribution along with its parameters within a number of generic distribution models may necessitate testing each candidate in order to choose the best possible model for the observed data/noise.

In the literature, the general practice is to assume the generic distribution family beforehand and estimate its parameters given the observed data. The most common choice is the Gaussian probability model especially in communications, network modelling, digital images, mainly due to its analytical ease. However, it has been reported in the literature that the noise exhibits non-Gaussian and impulsive characteristics in application areas such as wireless communications (Bhatti et al., 2009; Blackard et al., 1993), *power line communications* (PLC) (Lin et al., 2013; Alsusa and Rabie, 2013), *digital subscriber lines* (xDSL) (Al-Naffouri et al., 2011; Fantacci et al., 2010), image processing (Simoncelli, 1997; Achim et al., 2003) and seismology (Yue and Peng, 2015). In the case of non-Gaussian impulsive noise/data, various model families can be listed, for example, Middleton Class A, Bernoulli-Gaussian, α-Stable, Generalized Gaussian (GG), Student's

t, etc.

Other than the impulsive distributions, Poisson and negative binomial distributions have been widely used in modelling count data (Ver Hoef and Boveng, 2007; Hastie and Green, 2012). In addition, Poisson distribution has been also popular in queueing theory as well as exponential distribution (Brown et al., 2005; Wolff, 1982; Medhi, 2002). Other popular type of distributions are envelope distributions which are positive valued distributions, such as Rayleigh, Rice, Weibull, Nakagami, log-normal and have been popular to model positive valued real data sets. In the literature, distributions of this type have been used in applications such as fading channel modelling in communications (Patzold et al., 1998; Choudhary and Robinson, 2014; Rodrigo-Peñarrocha et al., 2016), SAR imaging (Delignon and Pieczynski, 2002; Kuruoglu and Zerubia, 2004; Gao, 2010; Li et al., 2011), wind speed distribution modelling (Chou and Corotis, 1983; Baïle et al., 2011; Harris and Cook, 2014; Drobinski et al., 2015).

Previous studies mentioned above clearly demonstrate the importance of estimating or having an idea about which distribution an observed set of data comes from. Even if the candidates are of different classes, selecting the best possible generic distribution model in a large number of candidate model spaces can be performed by utilizing the trans-space RJMCMC methodology presented in the previous chapters. This can be achieved by avoiding performing an exhaustive search on the whole model space that is, trying each model class separately. This chapter contributes to the literature with a new interpretation on RJMCMC beyond trans-dimensional sampling and also presents a generalization to the procedures presented in Chapters 3 and 4. The usage of RJMCMC methodology in statistical models PAR, PMA and PARMA and VSI problems considered in the previous chapters can be classified as an extension of common RJMCMC methodology of making trans-dimensional jumps to performing multiple trans-dimensional jumps, either in a cascaded fashion as in the cases of PAR, PMA and PARMA or concurrently as in the case of VSI. However, jumping to be introduced in this chapter is between different parameter spaces (classes) and transition between spaces of different probability distribution families constitutes a suitable application area and a good example of this type of transition.

5.1. Trans-distributional RJMCMC

The proposed method can use different names which explain the applications clearly according to the cases that trans-space sampling is performed. For example, trans-

space RJMCMC used for identification of Volterra systems in the previous chapter has been named as *trans-structural* since linear and nonlinear Volterra models follow different characteristics in a structural manner (even nonlinear Volterra models with different nonlinearity degrees, e.g. quadratic Volterra models and cubic Volterra models). Although the candidate models are structurally different from each other, all candidate models are Volterra system models. Thus, the connection between models during RJMCMC transitions is established by means of the Volterra model in (4.1) and reformulation of the RJMCMC procedures derived in Chapter 3 for nonlinear time series models can be easily performed due to the linear-in-the-parameters nature of the Volterra system models in the trans-structural approach.

On the other hand, the trans-space RJMCMC application in this chapter presents a different approach, in particular, it explores the spaces of different distribution families and thus will be named as *trans-distributional*. While the candidate distributions in a single family can be connected to each other with the help of defined expressions of those families, the transitions between different families are the most important challenge to the trans-distributional approach because of the different mathematical expressions and properties of all of the distribution families.

The trans-distributional approach necessitates additional considerations in order not to lose the potential benefits offered by RJMCMC, unlike other trans-space approaches presented up to this chapter. Sampling across unrelated spaces would not give us a computational advantage. In that case, one could solve for different spaces separately and compare the final results to choose the best model. Therefore, the Borel sets should be *related* somehow. The relation in question can be conveniently established, e.g. through *matching norms* in defining the spaces. Defining proposals in this way will provide sampling more efficient candidates and help algorithm to converge faster (an example will be discussed in Section 5.3.5.2).

5.2. Distribution Families

5.2.1. Impulsive Distribution Families

As mentioned in the above sections, although the Gaussian assumption is often used in the literature, it is of vital importance to use impulse distribution models in many

real-life problems. To give an example, coding schemes or channel capacity estimates obtained using a wrong noise model can cause serious loss in communication systems (Wang et al., 2011). Besides, the assumption of a wrong distribution in any other application area can cause any developed method to perform below expectations. The general belief that the impulsive distribution families have relatively more complex estimation methods than the Gaussian distribution limit the utilization of impulsive distributions. Therefore, it is very important to accurately predict which impulsive distribution family an observed data comes from.

As impulsive distribution families, three families have been utilized in this thesis, namely, SαS, GG and Student's t (details are given in Appendix A). These three families cover many different noise modelling studies and include Gaussian distribution as a special member, and many real life noise measurements can be modelled using these distribution families.

For example, SαS family has various demonstrated application areas such as PLC (Laguna-Sanchez and Lopez-Guerrero, 2015), SAR imaging (Achim et al., 2003), near optimal receiver design (Kuruoglu et al., 1998), modelling of contourlet transform subbands (Sadreazami et al., 2014), seismic amplitude data modelling (Yue and Peng, 2015), as noise model for molecular communication (Farsad et al., 2015), reconstruction of non-negative signals (Tzagkarakis and Tsakalides, 2010) (please see (Nolan, 2010) and references therein for detailed applications).

GG distributions have found applications in wavelet based texture retrieval (Do and Vetterli, 2002), image modelling in terms of Markov random fields (Bouman and Sauer, 1993), multicomponent texture discrimination in color images (Verdoolaege and Scheunders, 2011), wheezing sound detection (Le Cam et al., 2009), modelling sea-clutter data (Novey et al., 2010).

Student's t distribution is an alternative to Gaussian distribution especially for small populations where the validity of central limit theorem is questionable. Student's t distribution has been used in applications of finance (Patton, 2006; Engle and Bollerslev, 1986), full-waveform inversion of seismic data (Aravkin et al., 2011), independent vector analysis for speech separation (Liang et al., 2013), medical image segmentation (Nguyen and Wu, 2012), growth curve modelling (Zhang et al., 2013).

5.2.2. Envelope Distribution Families

Envelope distribution families in this case study are continuous distribution families which are supported on interval, $[0, \infty)$. Positive valued measurements have been modelled by using distributions of this type. Members, such as Nakagami, Rayleigh, Rice, Weibull, log-normal, have been utilized to model various applications of communication system channels, atmospherical instances such as wind speed, speckle noise of SAR and ultrasound images, etc.

In particular, Nakagami distribution has been very popular in modelling communication fading channel along with Weibull distribution (Patzold et al., 1998; Choudhary and Robinson, 2014; Rodrigo-Peñarrocha et al., 2016). Additionally, Nakagami distribution has been also used in classification and characterization of ultrasonic images (Shankar et al., 2001; Shankar, 2001), wind speed modelling (Alavi et al., 2016), speckle suppression (Ghofrani et al., 2001).

In addition to its usage in communications, Weibull distribution has been extensively used to model wind speed distribution (Chou and Corotis, 1983; Baïle et al., 2011; Harris and Cook, 2014; Drobinski et al., 2015). Additionally, it has been also used in modelling survival data (Mudholkar et al., 1996) and failure data (Cordeiro et al., 2010).

$\mathcal{K}$-distribution has been popular in SAR imagery (Gao, 2010; Bian and Mercer, 2015), communications to model fast fading channels and shadowing effects (Reig and Rubio, 2013; Tuli et al., 2014; Li, 2016), modelling polarimetric radar data (Yueh et al., 1989), atmospheric general circulation model (Sekiguchi and Nakajima, 2008).

Gamma distribution has been used to model various signals in areas such as fading channel of communication systems (Shankar, 2004; Al-Ahmadi and Yanikomeroglu, 2010), SAR imagery (Gao et al., 2017), rainfall (Aksoy, 2000), bacterial-gene expressions (Friedman et al., 2006). Generalized (heavy tailed) Rayleigh distribution (Kuruoglu and Zerubia, 2004) has been proposed in order to model urban scenes SAR images with a better performance than the classical distributions Rayleigh, Weibull and $\mathcal{K}$.

5.3. RJMCMC for Impulsive Distribution Estimation

In this case study, we have applied RJMCMC to problems in which a stochastic process, $\mathbf{x}$, is given whose impulsive distribution is to be found. For this purpose, we have defined a reversible jump mechanism which estimates the distribution family among three

impulsive distribution families, namely, SαS, GG and Student's *t*.

In the literature, RJMCMC usage in this problem has been limited and it has been used to be examples of trans-dimensional approach deciding between two specific distributions (Hastie and Green, 2012; Barker and Link, 2013). Particularly, when modelling count data, reversible jump mechanism has been applied to choose between Poisson and negative binomial distributions in (Hastie and Green, 2012). This study deals with the question whether the count data is over-dispersed relative to Poisson distribution. In (Barker and Link, 2013) an approach which is a combination of Gibbs sampler and RJMCMC has been used to decide between Poisson and geometric distributions by using a universal parameter space called "palette". Both of the studies have utilized RJMCMC in distribution estimation; however, the number of candidate distributions was limited to two. In both of the studies, Poisson distribution is a special member of the distribution families in question (or, there is a direct relation between Poisson and negative binomial or geometric distributions) and the proposed methods in these questions can be handled with a single family search.

The proposed method, *trans-distributional* RJMCMC, is much more general than the examples above and aims to fit a distribution to a given process **x** among various distributions by identifying the distribution's family and estimating its shape and scale parameters. Two types of between-class moves have been defined, namely *intra-class-switch* and *inter-class-switch*. These moves propose model class changes *within* and *between* probability distribution families, respectively.

5.3.1. Parameter Space

RJMCMC construction for impulsive data modelling begins by first defining the parameter space. Parameter space has been defined on the common parameters for all three distribution families. These are: *shape, scale* and *location* parameters (α, γ and δ, respectively). In addition to them, *the family identifier*, *k*, which refers to the estimated distribution family is added to the parameter space. The *k* values of the distributions SαS, GG and Student's *t* are 1, 2 and 3, respectively. Therefore, the parameter vector $\boldsymbol{\theta}$ can be formed as: $\boldsymbol{\theta} = \{k, \alpha, \delta, \gamma\}$.

In this case study, the observed data from all three families are assumed to be symmetric around the origin for simplicity. Therefore, δ, is set to 0. Hence, its effect will be invisible in the simulations. Consequently, parameter vector $\boldsymbol{\theta}$ is reduced to: $\boldsymbol{\theta} = \{k, \alpha, \gamma\}$.

5.3.2. Hierarchical Bayesian Model

The target distribution, $f(\boldsymbol{\theta}|\mathbf{x})$ can be decomposed to likelihood times priors due to Bayes Theorem as

$$f(\boldsymbol{\theta}|\mathbf{x}) \propto f(\mathbf{x}|k, \alpha, \gamma)f(\alpha|k)f(k)f(\gamma) \tag{5.1}$$

where $f(\mathbf{x}|k, \alpha, \gamma)$ represents the likelihood and $f(\alpha|k)$, $f(k)$, and $f(\gamma)$ are the priors.

5.3.3. Likelihood

We assume that the stochastic process $\mathbf{x}$ with a length of n comes from one of the distributions in candidate families (SαS, GG and Student's t). Then, the likelihood corresponds to a pdf from one of these distributions

$$f(\mathbf{x}|k, \alpha, \gamma) = \begin{cases} \prod_{i=1}^{n} \text{S}\alpha\text{S}(\gamma), & k = 1 \\ \prod_{i=1}^{n} \text{GG}_{\alpha}(\gamma), & k = 2 \\ \prod_{i=1}^{n} t_{\alpha}(\gamma), & k = 3 \end{cases} \tag{5.2}$$

5.3.4. Priors

Priors are selected as the following

$$f(\gamma) = \mathcal{IG}(a, b), \tag{5.3}$$

$$f(k) = \mathbb{I}_{\{1/3, 1/3, 1/3\}} \quad \text{for } k = 1, 2, 3, \tag{5.4}$$

$$f(\alpha|k) = \begin{cases} \mathcal{U}(0, 2) & k = 1, \\ \mathcal{U}(0, \alpha_{\text{max,GG}}) & k = 2, \\ \mathcal{U}(0, \alpha_{\text{max},t}) & k = 3, \end{cases} \tag{5.5}$$

where *a* and *b* represent the hyperparameters for scale parameter and they are generally selected to take small values such as $1, 0.1$ in the literature. The upper bounds for the shape parameters of GG and Student's *t* distributions are defined as $\alpha_{\text{max,GG}}$ and $\alpha_{\text{max},t}$, respectively.

Choosing an inverse gamma prior for scale parameter has been a general practice especially for Gaussian problems. Due to the lack of prior work about conjugate priors for distributions other than the Gaussian case and since Gaussian distribution is common for all three families, an inverse gamma conjugate prior for scale parameters is chosen for simplicity. Furthermore, all families are equiprobable *a priori* and shape parameter is uniformly distributed between lower and upper bounds.

5.3.5. Model Moves

Two RJMCMC model moves have been defined to perform trans-distributional transitions discussed in the previous sections. These are: *life* and *switch* moves. Life move performs classical MH algorithm to update scale parameter γ while k and α remain the same. Switch move performs exploring the other distribution spaces. For this purpose, two types of switch moves have been defined: *intra-class-switch* and *inter-class-switch*. Intra-class-switch performs exploring the distributions in the same family, while inter-class-switch explores spaces of different families. At each RJMCMC iteration, one of the moves is chosen with probabilities P_{life}, $P_{\text{intra-cl-sw}}$ and $P_{\text{inter-cl-sw}}$, respectively. In Figure 5.1 the flow diagram of the proposed method is depicted where the parameter N refers to the maximum number of iterations.

5.3.5.1. Life Move

Life move defines a transition from parameter space (k, α, γ) to (k', α', γ') and only proposes a candidate for the scale parameter, γ ($\alpha' = \alpha$ and $k' = k$). The proposal distribution for scale parameter γ' is chosen as

$$q(\gamma'|\gamma) = \mathcal{T}\mathcal{N}(\gamma, \xi_{scale}) \quad \text{for interval } (0, \gamma + 1] \tag{5.6}$$

where $\mathcal{TN}(\gamma, \xi_{scale})$ refers to a Gaussian distribution where its mean γ is the last value of the scale parameter, and its variance is ξ_{scale} and is truncated to lie within the interval of $(0, \gamma + 1]$ afterwards by rejecting samples outside this interval. This truncation procedure aims to satisfy the condition $\gamma > 0$ and forces candidate proposals not to lie far from the last value of γ.

Hence, the resulting acceptance ratio for life move is:

$$A_{\text{life}} = \min \left\{ 1, \frac{f(\mathbf{x}|k', \alpha', \gamma')}{f(\mathbf{x}|k, \alpha, \gamma)} \frac{f(\gamma')}{f(\gamma)} \frac{q(\gamma|\gamma')}{q(\gamma'|\gamma)} \right\} \tag{5.7}$$

5.3.5.2. FLOM Based Proposals for γ Transitions

As mentioned earlier in this chapter, using a common feature among the candidate model spaces for the transition to be made will provide efficient proposals and is important in order to link the subspaces of different classes. Assume we have two candidate families parameter vectors of which belong to Borel sets, $\mathcal{A}$ and $\mathcal{B}$, respectively. Providing fixed order norm for both of the Borel sets, the transition (e.g. $h : \mathcal{A} \mapsto \mathcal{B}$) from one set to another carries the information in the same direction which has been already learned at the most recent Borel set. Considering the convergence and mixing of the algorithm, such an approach is very important to determine the transition process between generic distribution models, whether within the family or between families.

When dealing with distribution estimation problems, moments with various orders, p, have been defined for all distribution families. Moments of t and GG families have been defined at any orders for $p > 0$ and there are no restrictions on values of p. However, moments of the SαS family have been defined subject to the constraint of $p < \alpha$. This constraint makes it possible to use the absolute *fractional lower order moments (FLOMs)* which has been also used in the parameter estimation methods of the SαS family. Since absolute FLOM expressions are defined for all impulsive families, and their success in estimating the parameters of SαS distributions is well documented in the literature (Nikias and Shao, 1995), using an absolute FLOM based approach is employed to construct a reversible jump sampler between different impulsive families, by linking the candidate distributions through absolute FLOM.

In impulsive data modelling study in this thesis, absolute FLOM-based approach will be used for the proposals of the γ parameter. In particular, to perform sampling be-

tween related subspaces and generate efficient proposals on scale parameter γ, an absolute FLOM-based method has been used. The newly proposed scale parameter, γ', is calculated via a reversible function, $g(\cdot)$ (or $w(\cdot)$), which provides equal absolute FLOMs with order p for both the most recent and candidate distribution spaces. Thus, proposals on γ carry the learned information to the candidate space via absolute FLOMs.
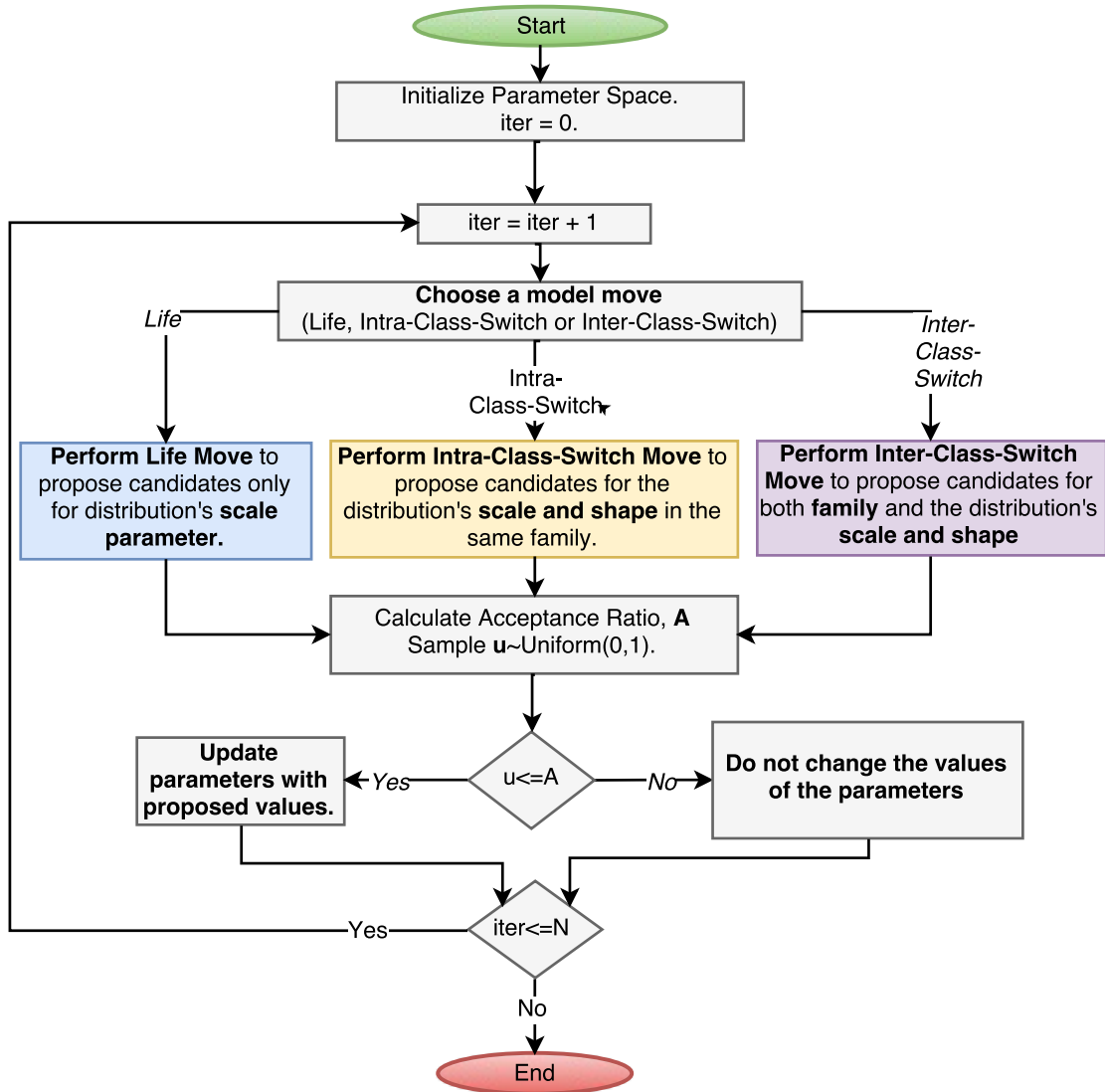

Figure 5.1. Flow Diagram for the Proposed method.

Absolute FLOMs are defined only for p values lower than α for the case of SαS distributions. Moreover, there are several studies which suggest near-optimum values for FLOM order p in order to estimate the scale parameter of SαS distributions. (Tsihrintzis

and Nikias, 1996) suggests $p = \alpha/4$ and (Ma and Nikias, 1995) suggests $p = 0.2$. However, in (Kuruoglu, 2001) it has been stated that decreasing p for a fixed value of α (i.e. increasing α/p), increases the estimation performance of γ and (Kuruoglu, 2001) suggests the choice $p = \alpha/10$. Hence, considering this limitation of the SαS distributions, we have utilized the value $p = \alpha/10$ in our simulations for all the distribution families.

For a given data, $\mathbf{x}$, in order to perform a transition from parameter space $\{k, \alpha, \gamma\}$ to $\{k', \alpha', \gamma'\}$ we assume that the absolute FLOM is the same for both the most recent and candidate distribution spaces. In particular,

$$E_k(|\mathbf{x}|^p) = E_{k'}(|\mathbf{x}|^p) \tag{5.8}$$

where absolute FLOMs for all three candidate families are given in Appendix A. The candidate proposal, γ', is calculated via reversible functions which are derived by using the relations in (A.2)-(A.9) for each transition. These functions have been derived for both of the switch moves and are shown in Tables 5.1 and 5.2.

5.3.5.3. Intra-Class-Switch Move

RJMCMC performs a transition on shape and scale parameters in the same distribution family ($k' = k$) when an intra-class-switch move is proposed. The proposed shape parameter α' is sampled from a proposal distribution $q(\alpha'|\alpha)$. In addition, the candidate scale parameter γ' is defined as a function $g(\alpha, \alpha', p, \gamma)$.

The γ transition in this move has been defined as dependent on the newly proposed α' parameter. For this reason, a step is first performed on shape parameter α to propose α', and this is used to calculate the candidate scale parameter γ'. For the shape parameter α transition, a proposal distribution such as $q(\alpha'|\alpha)$ has been used. For this distribution, we first have assumed a symmetric distribution around the most recent α value. In addition, it has been preferred that the proposal distribution has heavier tails than Gaussian in order to make it possible to sample candidates much farther than the most recent α relative to the samples from the Gaussian distribution. Since the Laplace distribution is a distribution that satisfies all these conditions, the proposal distribution is chosen as a Laplace distribution. Due to the numerical calculation problems caused when α and α' are close to each other (i.e. $|\alpha - \alpha'| \leq 0.03$), we have decided to utilize a finite number of candidate distributions (i.e. a finite number of α values) and the space on α is discretized

with increments of 0.05. That's why a discretized Laplace ($\mathcal{DL}(\alpha, \Gamma)$) distribution where the location parameter of which is equal to the most recent shape parameter α and scale parameter is Γ has been utilized.

Importantly, our choice on the proposal distribution $q(\alpha'|\alpha)$ is not restrictive; any distribution other than Laplace can be selected as the proposal distribution (e.g. Gaussian like). However, different selections will cause the algorithms to perform differently. An example figure of the proposal distribution $q(\alpha'|\alpha)$ is shown in Figure 5.2(a).

Candidate scale parameter γ' is calculated via reversible functions, $g(\cdot)$, which are derived for intra-class-switch move by using the method in Section 5.3.5.2. Functions for each family are shown in Table 5.1.

Consequently, proposals for intra-class-switch move are

$$q(\alpha'|\alpha) = \mathcal{DL}(\alpha, \Gamma) \tag{5.9}$$

$$\gamma' = g(\alpha, \alpha', p, \gamma). \tag{5.10}$$

As a result of the details explained above, acceptance ratio for RJMCMC intra-class-switch move can be expressed as

$$A_{\text{intra-cl-sw}} = \min\left\{1, \frac{f(\mathbf{x}|k', \alpha', \gamma')}{f(\mathbf{x}|k, \alpha, \gamma)} \frac{f(\gamma')}{f(\gamma)}|J|\right\} \tag{5.11}$$

where $|J|$ is the magnitude of the Jacobian (See Table 5.1).

5.3.5.4. Inter-Class-Switch Move

Different from intra-class-switch move, distribution family is also changed in inter-class-switch move ($k' \neq k$) as well as scale and shape parameters. Candidate distribution families are equiprobable for the candidate set $\{1, 2, 3\}\backslash\{k\}$, and we use functions below to propose candidate parameters of α' and γ'

$$\alpha' = \psi(\alpha, k, k') \tag{5.12}$$

$$\gamma' = w(\alpha, \alpha', p, \gamma). \tag{5.13}$$

For intra-class transitions mentioned in the section above, the knowledge (about scale γ) learned in the previous algorithm steps was carried to the next step by the proposed method via FLOM based functions. The same approach is also utilized for γ transitions in inter-class-switch move and functions $w(\cdot)$ are derived. However, this time, the sides of the transition are in different families. Details are shown in Table 5.2.

Using a similar approach in α transitions will allow the proposed method to converge faster across different distribution families. However, this process should be taken very carefully. That is to say that distribution families have common distributions. For example, Gaussian distribution is common for all families and for SαS, GG, and t, values for α are 2, 2, and ∞, respectively. During intra-class transitions, disregarding these relationships may cause the algorithm to converge to an incorrect distribution. In order to prevent such situations and to perform efficient proposals for α in inter-class-switch move, instead of using a random move, we perform a mapping, $\psi(\cdot)$ from one family to another by taking into consideration the special members which are common for both of the families. At this point, α transitions will also occur within the framework of more consistent proposals and the convergence will be accelerated. For example, to derive an invertible mapping function on α for a transition from SαS to Student's t, we utilize the information that Cauchy and Gauss distributions are common for both of the families. Cauchy refers to $\alpha = 1$ for both of the families and Gauss refers to $\alpha = 2$ for SαS and $\alpha = \infty$ for Student's t. Hence, the invertible function $f_2(\alpha)$ performs the mapping for a transition from SαS to Student's t.

Similarly, Gauss distribution is common for both SαS and GG for α value of 2. Thus, we derive another invertible function $f_1(\alpha)$ to move from SαS to GG. Both of these mapping functions are depicted in Figure 5.2(b). GG and Student's t distributions have only Gauss distribution in common for α values of 2 and ∞, respectively. Due to having only one common distribution and infinite range of α, instead of deriving an invertible mapping for transitions between these distributions, we perform a 2-stage mapping mechanism by firstly mapping α to SαS from the most recent family, then mapping this value to the candidate family by using functions $f_1(\cdot)$ or $f_2(\cdot)$. Then the mapping from GG to Student's t is derived as: $\alpha' = f_2(f_1^{-1}(\alpha))$. It is straightforward to show that the reverse transition between shape parameters from Student's t to GG results as $\alpha' = f_1(f_2^{-1}(\alpha))$. For all the transitions, mapping functions are shown in Table 5.2.

Table 5.1. FLOM-based (Deterministic) Assignments for Intra-Class-Switch

Family	Degree, p	$\gamma' = g(\alpha, \alpha', p, \gamma)$	Jacobian, $|J|$
SαS	$\alpha'/10$	$\left(\dfrac{C_\alpha(p,\alpha)}{C_\alpha(p,\alpha')}\right)^{\alpha'/p} \gamma^{\alpha'/\alpha}$	$\left(\dfrac{C_\alpha(p,\alpha)}{C_\alpha(p,\alpha')}\right)^{\alpha'/p} \dfrac{\alpha'}{\alpha} \gamma^{(\alpha'-\alpha)/\alpha}$
GG	$\alpha'/10$	$\left(\dfrac{C_{GG}(p,\alpha)}{C_{GG}(p,\alpha')}\right)^{1/p} \gamma$	$\left(\dfrac{C_{GG}(p,\alpha)}{C_{GG}(p,\alpha')}\right)^{1/p}$
t	$\alpha'/10$	$\left(\dfrac{C_t(p,\alpha)}{C_t(p,\alpha')}\right)^{1/p} \gamma$	$\left(\dfrac{C_t(p,\alpha)}{C_t(p,\alpha')}\right)^{1/p}$

Table 5.2. FLOM-based (Deterministic) Assignments for Inter-Class-Switch

$(k \rightarrow k')$	Degree, p	$\alpha' = \psi(\alpha, k, k')$	$\gamma' = w(\alpha, \alpha', p, \gamma)$
$1 \rightarrow 2$	$\alpha'/10$	$f_1(\alpha) = \dfrac{\alpha^2}{2}$	$\left(\dfrac{C_\alpha(p,\alpha)}{C_{GG}(p,\alpha')}\right)^{1/p} \gamma^{1/\alpha}$
$1 \rightarrow 3$	$\alpha'/10$	$f_2(\alpha) = logit\left(\dfrac{\alpha+2}{4}\right)$	$\left(\dfrac{C_\alpha(p,\alpha)}{C_t(p,\alpha')}\right)^{1/p} \gamma^{1/\alpha}$
$2 \rightarrow 1$	$\alpha'/10$	$f_1^{-1}(\alpha)$	$\left(\dfrac{C_{GG}(p,\alpha)}{C_\alpha(p,\alpha')}\right)^{\alpha'/p} \gamma^{\alpha'}$
$2 \rightarrow 3$	$\alpha'/10$	$f_2(f_1^{-1}(\alpha))$	$\left(\dfrac{C_{GG}(p,\alpha)}{C_t(p,\alpha')}\right)^{1/p} \gamma$
$3 \rightarrow 1$	$\alpha'/10$	$f_2^{-1}(\alpha)$	$\left(\dfrac{C_t(p,\alpha)}{C_\alpha(p,\alpha')}\right)^{\alpha'/p} \gamma^{\alpha'}$
$3 \rightarrow 2$	$\alpha'/10$	$f_1(f_2^{-1}(\alpha))$	$\left(\dfrac{C_t(p,\alpha)}{C_{GG}(p,\alpha')}\right)^{1/p} \gamma$

Thus, the acceptance ratio for inter-class-switch move can be expressed as

$$A_{\text{inter-cl-sw}} = \min\left\{1, \frac{f(\mathbf{x}|k', \alpha', \gamma')}{f(\mathbf{x}|k, \alpha, \gamma)} \frac{f(\gamma')}{f(\gamma)} \frac{f(\alpha|k)}{f(\alpha'|k')} |J|\right\} \tag{5.14}$$

where the magnitude of the Jacobian can be calculated as

$$|\mathcal{J}| = \left| \begin{matrix} \dfrac{\partial \gamma'}{\partial \gamma} & \dfrac{\partial \alpha'}{\partial \gamma} \\ \dfrac{\partial \gamma'}{\partial \alpha} & \dfrac{\partial \alpha'}{\partial \alpha} \end{matrix} \right| = \left| \begin{matrix} \dfrac{\partial \gamma'}{\partial \gamma} & 0 \\ 0 & \dfrac{\partial \alpha'}{\partial \alpha} \end{matrix} \right| = \frac{\partial \gamma'}{\partial \gamma} \frac{\partial \alpha'}{\partial \alpha}. \tag{5.15}$$

5.3.6. Experimental Study

We have studied four cases experimentally: synthetically generated noise, impulsive noise on PLC channels, 2-D DWT coefficients and seismic acceleration time series. Without loss of generality, distribution of data $\mathbf{x}$ is assumed to be symmetric around zero ($\delta = 0$). The algorithm starts with a Gaussian distribution model with initial values $k^{(0)} = 2$ and $\alpha^{(0)} = 2$. Initial value for scale parameter γ is selected as half of the interquartile range of the given data $\mathbf{x}$ and upper bounds $\alpha_{\text{max,S}\alpha\text{S}}, \alpha_{\text{max,GG}}$ and $\alpha_{\text{max},t}$ are selected as 2, 2 and 5, respectively. Some intuitive selections have been performed for the rest of the parameters. Move probabilities for intra-class-switch and inter-class-switch moves are assumed to be equally likely during the simulations. Additionally, in order to speed up the convergence of the distribution parameter estimations during the life move, which is the coefficient update move, it is chosen a bit more likely than intra-class-switch and inter-class-switch moves. Thus, the model move probabilities are selected as $P_{\text{life}} = 0.4, P_{\text{intra-cl-sw}} = 0.3$ and $P_{\text{inter-cl-sw}} = 0.3$. Hyperparameters for prior distribution of γ are set to $a = b = 1$ and variance of proposal distribution for γ in life move is set to $\xi_{scale} = 0.01$. Scale parameter Γ of the discretized Laplace distribution for intra-class-switch move is selected as 0.4.

RJMCMC performs 5000 iterations in a single RJMCMC run and half of the iterations are discarded as burn-in period when estimating the distribution parameters. Random numbers from all the families have been generated by using Matlab's Statistics and
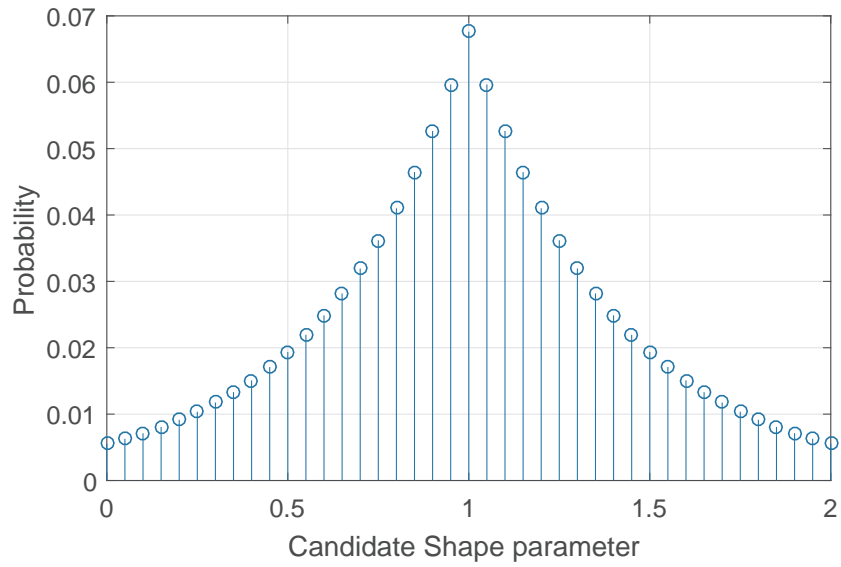
(a)
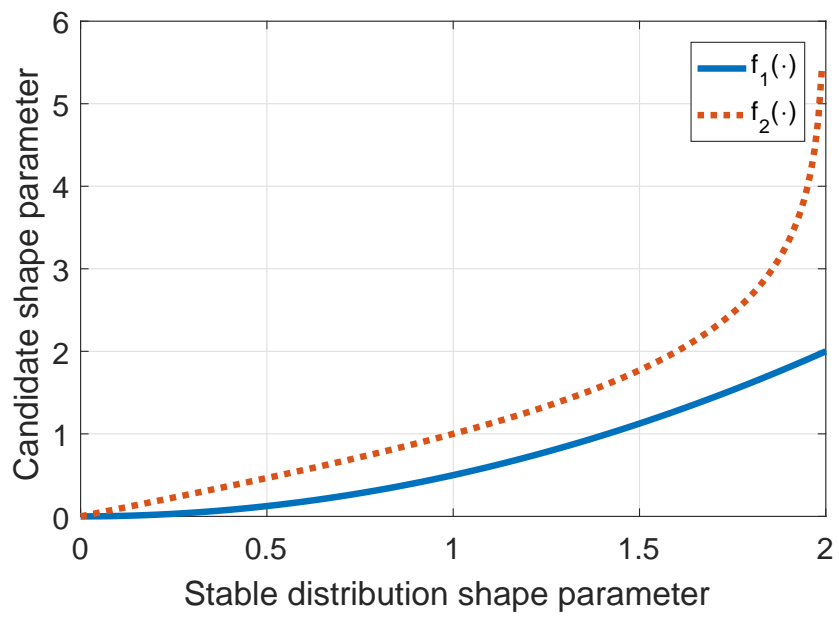
(b)

Figure 5.2. (a) - Proposal distribution, $q(\alpha'|\alpha)$ for intra-class-switch move ($\gamma = 1, \Gamma = 0.4$). (b) - Mapping functions on shape parameter for inter-class-switch move

Machine Learning Toolbox (for details please see[1]).

Performance comparison has been performed under two statistical significance tests, namely *Kullback-Leibler* (KL) divergence and *Kolmogorov-Smirnov* (KS) statistics. KL divergence has been utilized to measure fitting performance of the proposed method between estimated pdf and data histogram. Two-sample KS test compares empirical CDF of the data and the estimated CDF. It quantifies the distance between CDFs and performs an hypothesis test under a null hypothesis that two samples are drawn from the same distribution. Details about KL divergence and KS test have been discussed in Appendix C.

5.3.6.1. Case Study 1: Synthetically Generated Noise Modeling

In order to test the proposed method on modelling synthetically generated impulsive noise processes, six different distributions are chosen (2 distributions from each family). In a single RJMCMC run, data with a length of 1000 samples have been generated from one of the example distributions. The example distributions are S1S(0.75), S1.5S(2), $GG_{0.5}(0.5)$, $GG_{1.7}(1.4)$, $t_3(1)$ and $t_{0.6}(3)$.

40 RJMCMC runs have been performed for each distribution and estimated families with shape and scale parameters for each example distribution are shown in Table 5.3. Examining Table 5.3, fitting performances for all example distributions lie within KL distance of at most 0.0465. Moreover, estimated CDFs under KS statistic score are also very low and *p*-values are close to 1,0000. Please note that the estimation result in the second row of Table 5.3 is meaningful for an example Cauchy distribution, since the Cauchy distribution is a special member in both SαS and Student's *t* families.

In Figure 5.3, instantaneous estimate of shape parameter α and estimated posterior distribution of scale parameter γ are depicted for three example distributions. Results represent the estimates obtained by a randomly selected RJMCMC run out of 40 runs. Burn-in period is not removed in the sub-figures (a)-(c) in order to show the transient characteristics of the algorithm. These plots show that the proposed method converges to the correct shape parameters. In sub-figures (d)-(f), vertical dashed-lines with ∇ markers refer to $\pm\sigma$ *confidence interval* (CI). Examining these sub-figures shows that correct scale parameters lie within the $\pm\sigma$ CI of the posteriors.

Estimated pdfs and CDFs for three example distributions are depicted in Figure

---

[1]https://www.mathworks.com/help/stats/continuous-distributions.html

5.4. In addition to the statistical significance test scores in Table 5.3, fitting performance of the algorithm is presented visually. As can be seen from Figure 5.4, estimated pdfs and CDFs are very similar to the data histograms and empirical CDFs, respectively and provide evidence to the numerical results presented in Table 5.3.

Table 5.3. Modelling results for synthetically generated processes.

Distribution Distribution	**Est.** **Family**	**Est.** **Shape ($\hat{\alpha}$)**	**Est.** **Scale ($\hat{\gamma}$)**	**KL Div.**	**KS** **Score**	**KS** **p-value**
S1.5S(2)	SαS	1.4769	1.9162	0.0169	0.0125	1.0000
S1S(0.75)	t	0.9970	0.7300	0.0454	0.0489	> 0.9999
GG$_{0.5}$(0.5)	GG	0.4990	0.5199	0.0229	0.0152	1.0000
GG$_{1.7}$(1.4)	GG	1.6456	1.3374	0.0221	0.0202	1.0000
t_3(1)	t	2.9303	1.0039	0.0251	0.0203	1.0000
$t_{0.6}$(3)	t	0.6197	2.9869	0.0465	0.0452	> 0.9999

5.3.6.2. Case Study 2: Modelling Impulsive Noise on PLC Systems

PLC is an emerging technology which utilizes power-lines to carry telecommunication data. Telecommunication speeds up to 200 Mb/s with a good quality of service can be achieved on PLC systems. Apart from this, PLC offers a physical medium for indoor multimedia data traffic without additional cables (Laguna-Sanchez and Lopez-Guerrero, 2015).

A PLC system has various types of noise arising from electrical devices connected to power line and external effects via electromagnetic radiation, etc. These noise sequences are generally non-Gaussian and they are classified into three groups, namely: i) Impulsive noise, ii) Narrowband noise, iii) Background Noise (Cortes et al., 2010). Among these, impulsive noise is the most common cause of decoding (or communications) errors in PLC systems due to its high amplitudes up to 40 dBs (Andreadou and Pavlidou, 2010).

In this case study, we are going to use 3 different PLC noise measurements. First measurement (named as *PLC-1*) has been performed during a project with number PTDC/EEA-TEL/67979/2006. Details for the measurement scheme and other measurements please see (Lopes et al., 2013). Data utilized in this thesis (*PLC-1*) is an amplified impulsive noise measurement from a PLC system with a sampling rate of 200Msam-
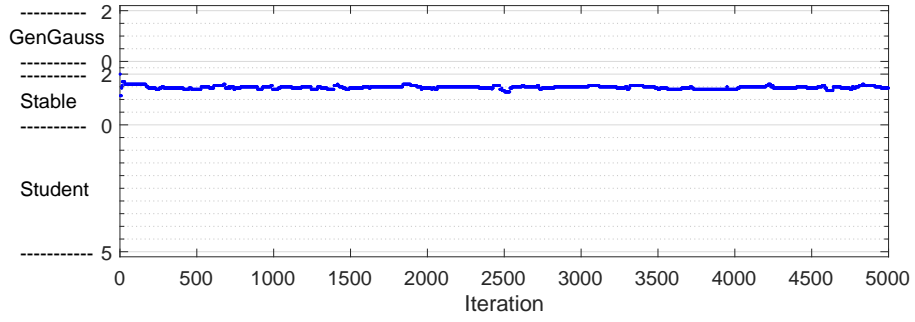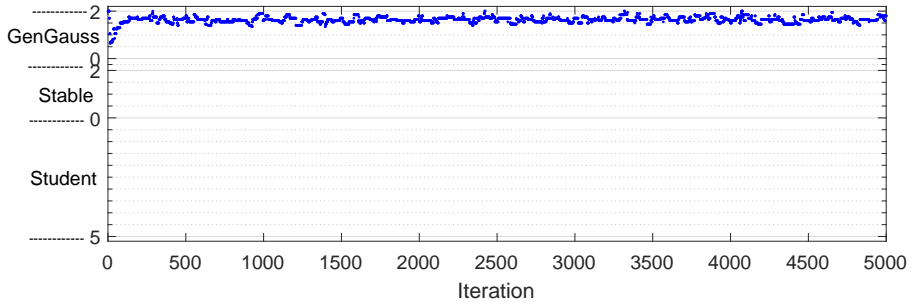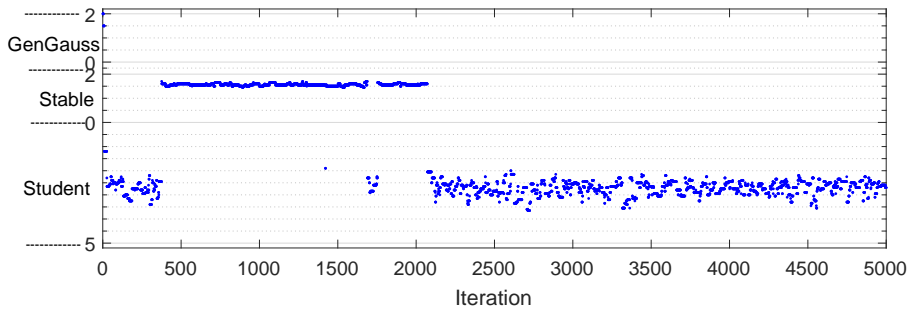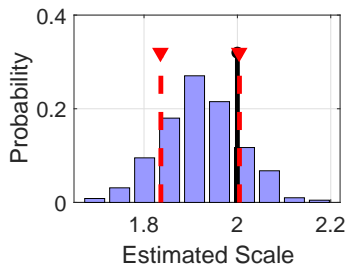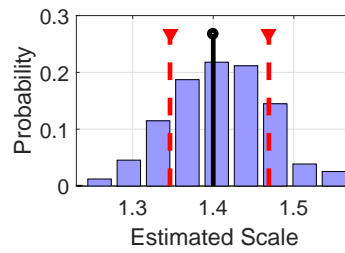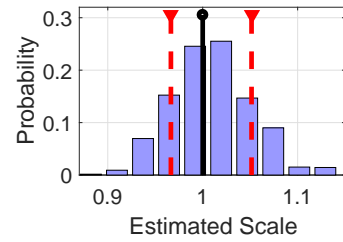
Figure 5.3. Synthetically generated noise modelling - parameter estimation results in a single RJMCMC run. (a), (b), (c): Instantaneous α estimates. (d), (e), (f): Estimated posterior distributions for γ after burn-in period. .
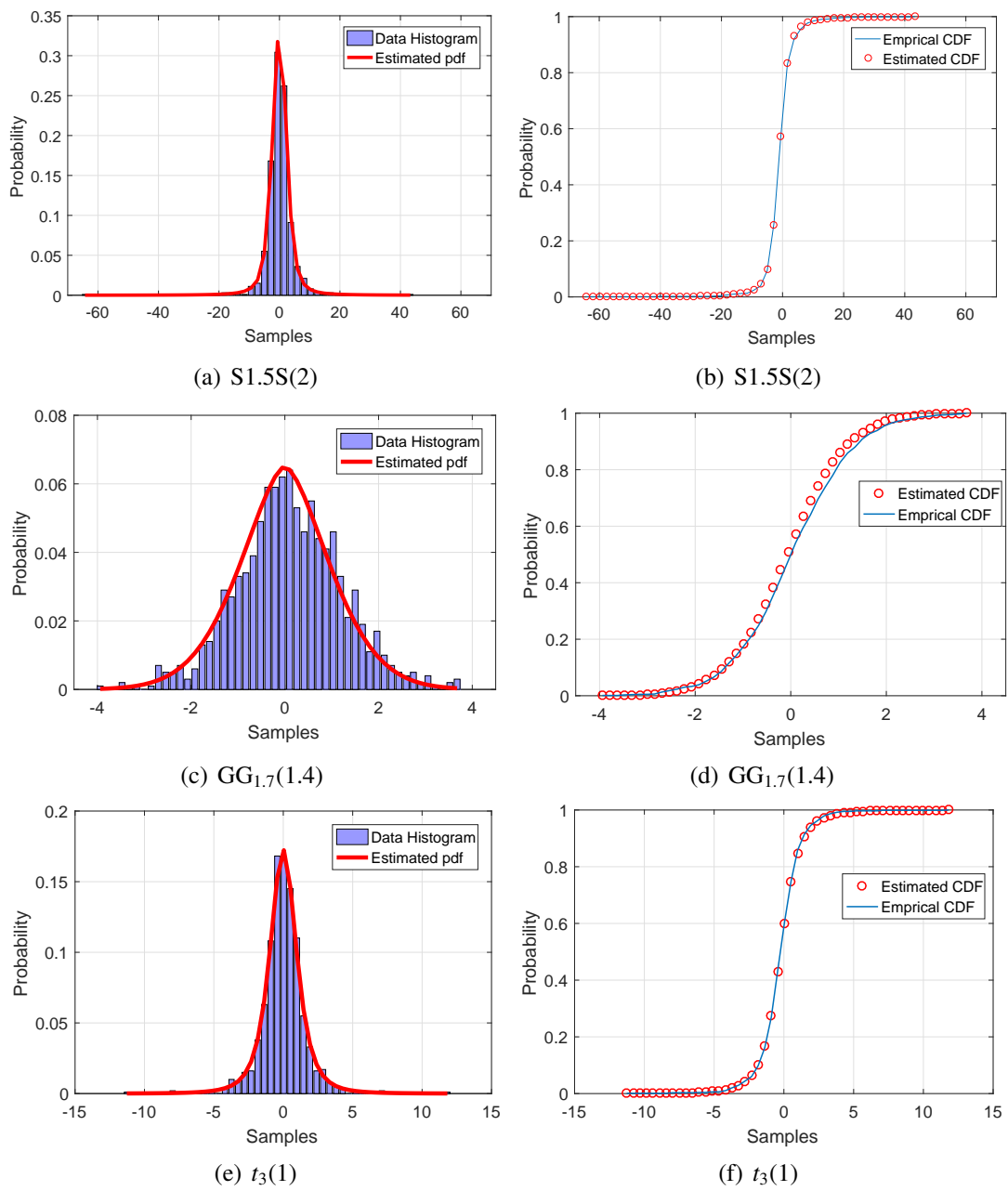
Figure 5.4. Synthetically generated noise modelling results. (a), (c), (e): Estimated pdfs, (b), (d), (f): Estimated CDFs. Estimated distributions for each row are S0.7387S(1.3213), $GG_{1.6456}(1.3374)$, and $t_{2.9303}(1.0039)$, respectively.

ples/sec. Measurements last for 5 ms and there are 100K samples in the data set. In order to reduce the computational load, the data is downsampled with a factor of 50 and the resulting 2001 samples have been used in this study. In Figure 5.5(a) a time plot of the utilized downsampled data is depicted (For detailed description of the data please see[2]).

Table 5.4. Modelling results for PLC impulsive noise.

Data	Est. Family	Est. Shape ($\hat{\alpha}$)	Est. Scale ($\hat{\gamma}$)	KL Div.	KS Score	KS p-value
PLC-1	SαS	1.2948	5.6969	0.0086	0.0112	1.0000
PLC-2	SαS	0.7042	0.1799	0.0441	0.0486	> 0.9999
PLC-3	SαS	1.3140	1.3488	0.0046	0.0132	1.0000

Remaining two data sets are periodic synchronous and asynchronous (named as *PLC-2* and *PLC-3*, respectively) impulsive noise measurements both of which have been performed during project with number TIC2003-06842 (for details please see (Cortes et al., 2010)). Periodic synchronous measurements last for 4 μs and contain 226 noise samples. Periodic asynchronous measurements contain 1901 noise samples and last for 35 μs. In Figures 5.5(b) and 5.5(c) time plots are depicted for synchronous and asynchronous noise sequences, respectively (for detailed description of the data please see[3]).

RJMCMC has been run 40 times for all three data sets. In Table 5.4, estimated distribution families and resulting scale and shape parameters are depicted with significance test results. Estimated scale and shape parameters correspond to the average values after 40 repetitions. Examining the results in Table 5.4, we can state that all three considered PLC noise processes follow SαS distribution characteristics. In the literature, there are studies (Laguna-Sanchez and Lopez-Guerrero, 2015; Tran et al., 2013) which model the impulsive noise in PLC systems by using stable distributions. Particularly, these studies provide a direct modelling scheme via stable distribution, whereas the proposed method has estimated the distribution among three impulsive distribution families. Thus, our estimation results for impulsive noise in PLC systems provide experimental verification of these studies. According to the results of KL and KS tests shown in Table 5.4 on estimated pdfs and CDFs and Figures through 5.6(a) and 5.6(f), RJMCMC fits real data within %5 significance scores. KS p-values are all approximately 1 (> 0.9999) and provide strong evidence that the estimated and the correct distributions are the same kind.

---

[2]http://sips.inesc-id.pt/~pacl/PLCNoise/index.html

[3]http://www.plc.uma.es/channels.htm
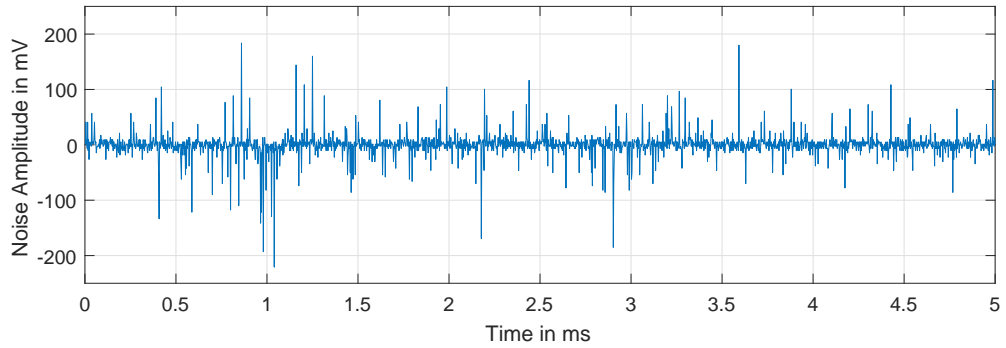
(a) *PLC-1*
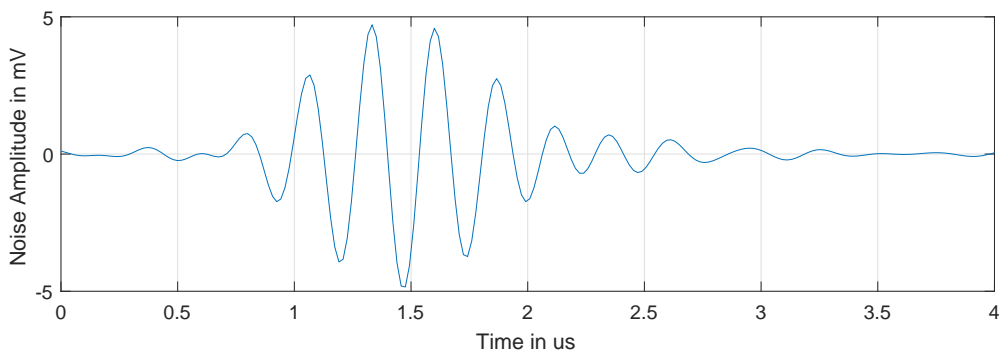
(b) *PLC-2*
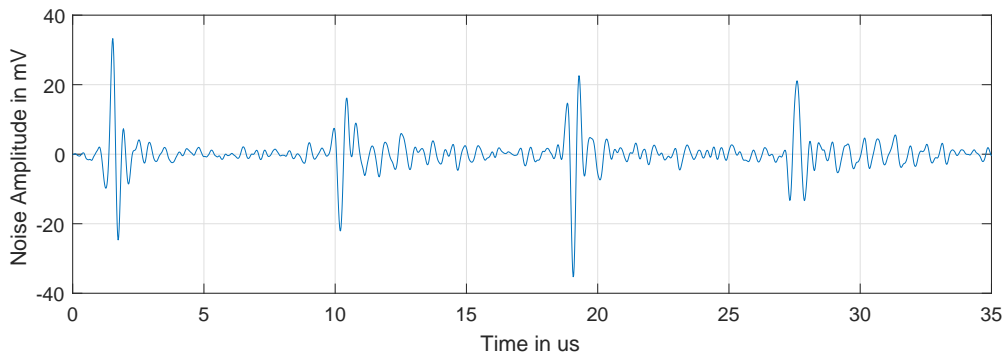
(c) *PLC-3*

Figure 5.5. PLC impulsive noise time plots. (a): an amplified impulsive noise measurement from a PLC system, (b): periodic synchronous impulsive noise measurements, (c): periodic asynchronous impulsive noise measurements.
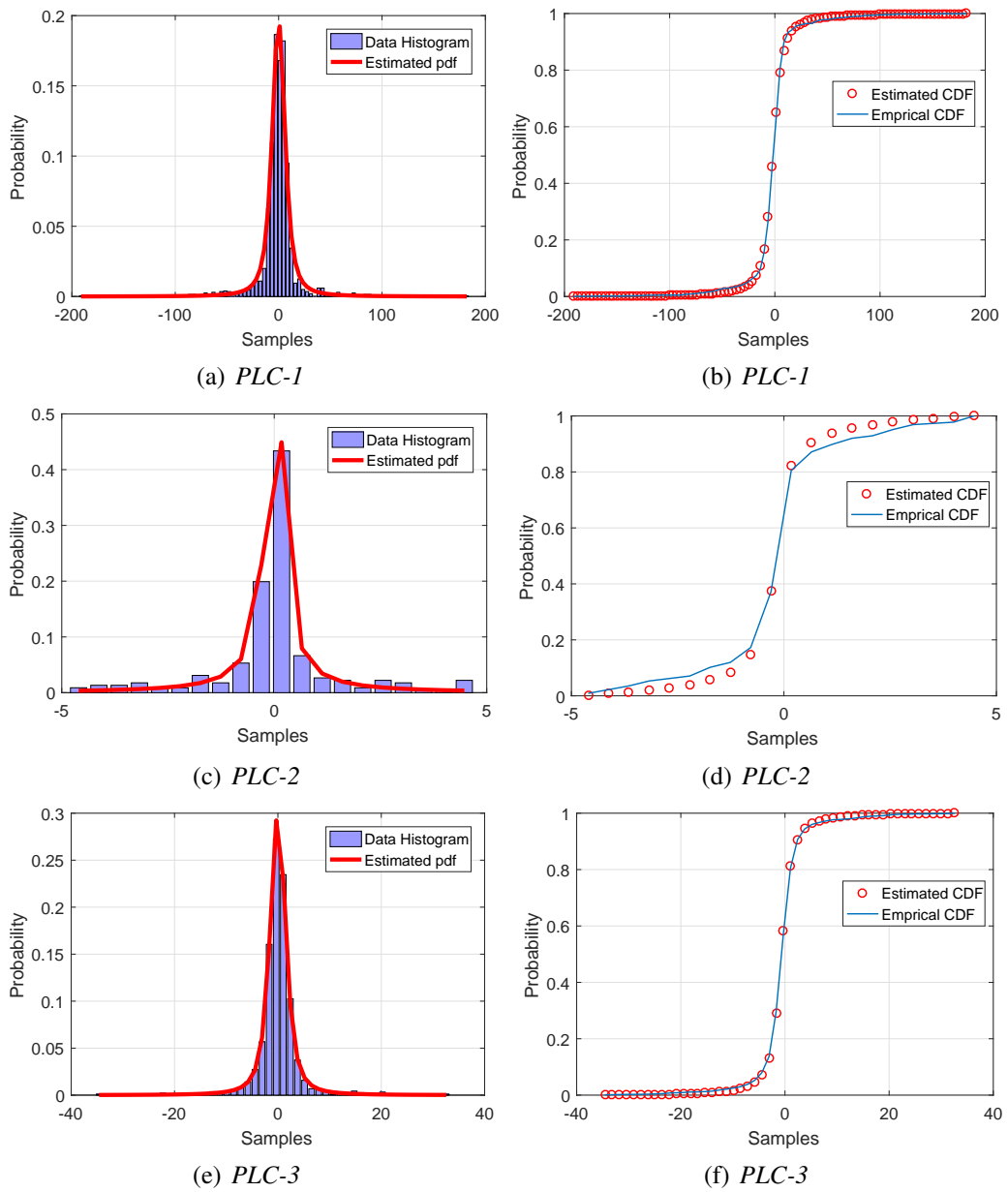
Figure 5.6. PLC impulsive noise modelling results. (a), (c), (e): Estimated pdfs, (b), (d), (f): Estimated CDFs. Estimated distributions for each row are S1.2948S(5.6969), S0.7042S(0.1799), and S1.3140S(1.3488), respectively

In order to compare the estimation results of the proposed method with a classical approach, maximum likelihood (ML) estimation method has been used for PLC noise estimation case study. For each PLC noise measurement and each candidate distribution family, k, parameters $\Theta^{(k)} = [\alpha, \quad \gamma]$ have been estimated via the expression

$$\hat{\Theta}_{\mathrm{ML}}^{(k)} = \arg\min_{\Theta^{(k)}} \left\{ -\log\left(f\left(\mathbf{x}|k, \Theta^{(k)} \right) \right) \right\} \tag{5.16}$$

where $f(\mathbf{x}|\cdot)$ refers to the likelihood function and $\mathbf{x}$ is the PLC noise measurement. A constrained *interior-point* optimization method has been employed with a maximum of 100 iterations to find a solution to the expression in (5.16). Constraints are defined for each of the shape and the scale parameters, which are upper and lower bounds given in the previous sections for each distribution family. Optimization operation has been repeated 100 times for randomly selected initial estimates ($\Theta_0^{(k)}$), and the results which minimize the expression in (5.16) are selected as the optimum solutions.

The optimum distribution from each family has been compared with others in terms of the KL and KS values, then the best family and the corresponding distribution has been decided to be the best fit. In Table 5.5, ML estimated families and parameters are shown for each PLC noise measurement. Examining the results shows us that ML estimated families and parameters are approximately the same as the results given for RJMCMC in Table 5.4. This demonstrates the success of the proposed method in estimating the best fit and being an automatic approach both searching the family and estimating the corresponding parameters.

Table 5.5. ML estimation results for PLC impulsive noise.

PLC-1			PLC-2			PLC-3		
Est. Family	Est. Shape ($\hat{\alpha}$)	Est. Scale ($\hat{\gamma}$)	Est. Family	Est. Shape ($\hat{\alpha}$)	Est. Scale ($\hat{\gamma}$)	Est. Family	Est. Shape ($\hat{\alpha}$)	Est. Scale ($\hat{\gamma}$)
SαS	1.2990	5.6530	SαS	0.6942	0.1725	SαS	1.3147	1.3459

5.3.6.3. Case Study 3: Statistical Modelling for Discrete Wavelet Transform (DWT) Coefficients

DWT which provides a multiscale representation of an image is a very important tool for recovering local and non-stationary features in an image. The resulting representation is closely related to the processing of the human visual system (Lee, 1996). DWT obtains this multiscale representation by performing a decomposition of the image into a low resolution approximation and three detail images capturing horizontal, vertical and diagonal details. It has been observed by several researchers that they have heavier tails and sharper peaks than Gaussian distribution (Simoncelli, 1997; Achim et al., 2003).

In this case study, the proposed method has been utilized to model the coefficients (e.g. subbands) of 2D-DWT, namely vertical (V), horizontal (H) and diagonal (D). Four different images have been used to test the performance of the algorithm under statistical significance tests: Lena, *synthetic aperture radar* (SAR) (Artemis Inc., 2017), *magnetic resonance imaging* (MRI) (MRI Scan Images Info., 2017) and mammogram (Martinez Lara et al., 2012) which are shown in Figure 5.7.

The proposed method has been performed for 40 RJMCMC runs. Estimated results for distribution families and their parameters (α and γ) are depicted in Table 5.6 as averages of 40 runs.

Table 5.6. Modelling results for 2D-DWT coefficients.

Image	Est. Family	Est. Shape ($\hat{\alpha}$)	Est. Scale ($\hat{\gamma}$)	KL Div.	KS Score	KS *p*-value
Lena (V)	GG	0.5002	1.7415	0.0271	0.0465	> 0.9999
Lena (H)	*t*	1.0958	2.2422	0.0094	0.0349	> 0.9999
Lena (D)	*t*	1.1628	1.7735	0.0145	0.0271	1.0000
SAR(V)	SαS	1.5381	7.7395	0.0025	0.0123	1.0000
SAR(H)	SαS	1.4500	8.6249	0.0043	0.0221	1.0000
SAR(D)	SαS	1.7500	6.3710	0.0062	0.0125	1.0000
MRI(V)	GG	0.3913	0.2693	0.0365	0.1152	0.8744
MRI(H)	GG	0.3527	0.1039	0.0305	0.0548	> 0.9999
MRI(D)	SαS	0.8504	0.5184	0.0245	0.0659	0.9998
Mammog.(V)	*t*	1.6325	1.6411	0.0363	0.0907	0.9816
Mammog.(H)	GG	0.7501	1.5154	0.0121	0.0555	> 0.9999
Mammog.(D)	*t*	1.6430	0.4851	0.0073	0.0117	1.0000

Estimated distributions for wavelet coefficients of images in Table 5.6 show dif-

ferent characteristics. SAR and MRI images follow generally SαS characteristics while results for Lena and mammogram images are generally GG or Student's t. Moreover, despite modelling with different distribution families, all the coefficients for all the images have been modelled successfully according to the KL and KS test scores and p-values. The estimated pdfs and CDFs in Figure 5.8 show remarkably good fitting and provide support to the results which are obtained numerically in Table 5.6.

(a) Lena

(b) SAR (Artemis Inc., 2017)

(c) MRI (MRI Scan Images Info., 2017)
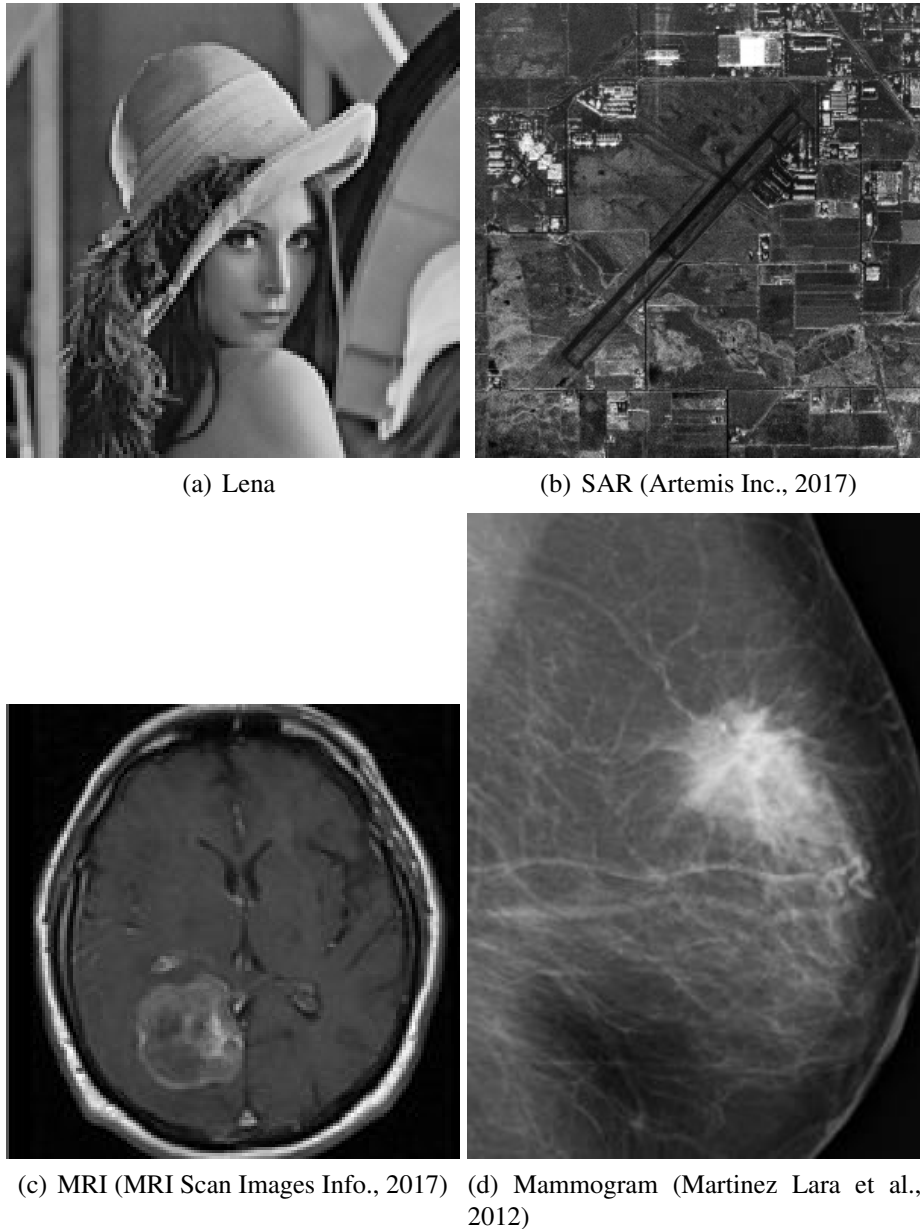
(d) Mammogram (Martinez Lara et al., 2012)
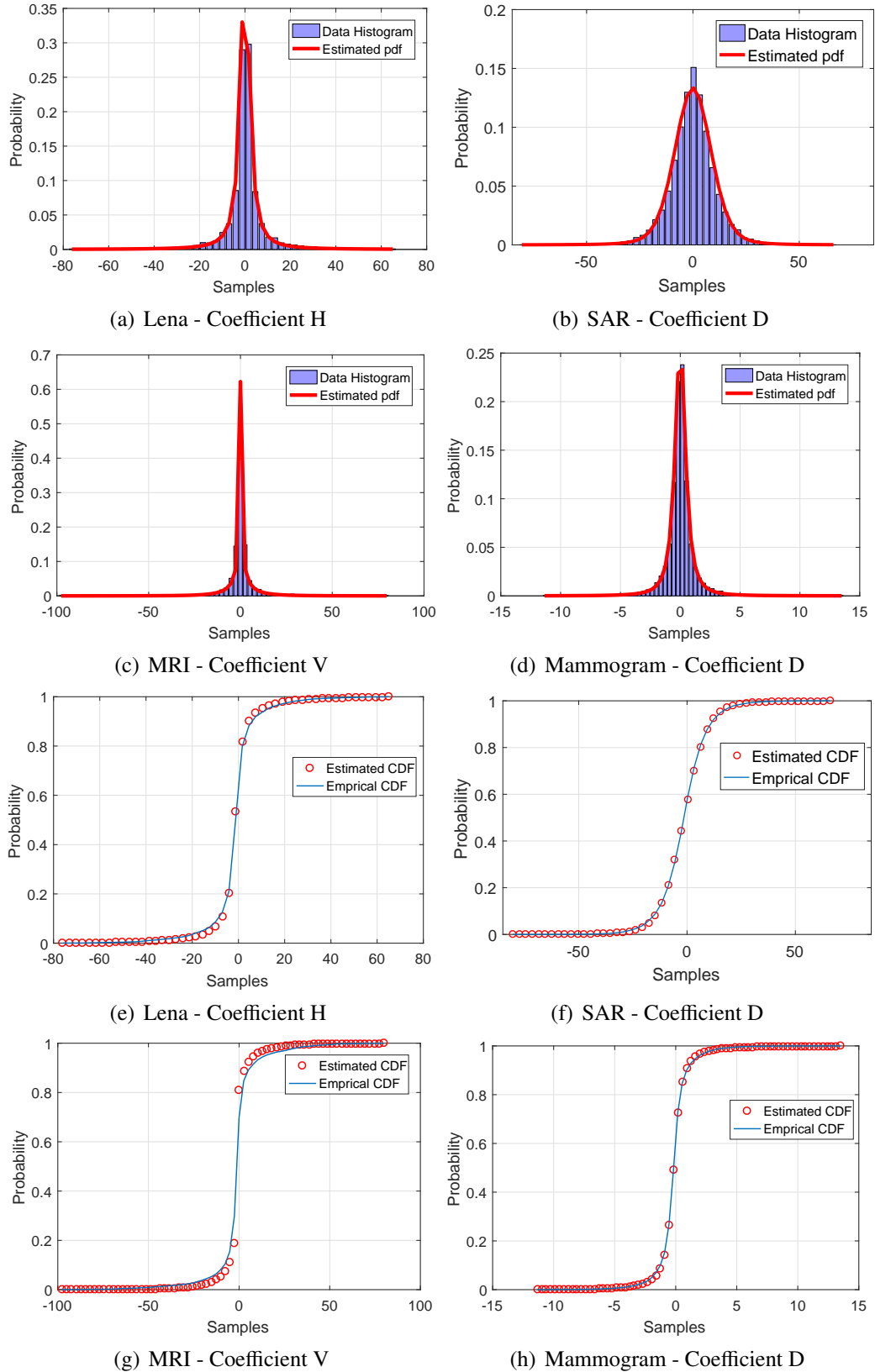
Figure 5.7. Images used for 2D-DWT coefficients modelling.

Figure 5.8. 2D-DWT coefficients modelling results. Estimated pdfs and CDFs. Estimated distributions are (a)-(e): $t_{1.0958}(2.2422)$, (b)-(f): S1.75S(6.3710), (c)-(h): $GG_{0.3913}(0.2693)$, (d)-(g): $t_{1.6430}(0.4851)$.

5.3.6.4. Case Study 4: Seismic Acceleration Time Series Modelling

Signal processing applications in seismology have been very popular in order to interpret the complex nature of the earthquakes. General practice is to assume Gaussianity for seismic signals and algorithms have been derived for denoising, seismic inversion, etc. on the basis of the Gaussian statistical models. However, in seismology and geophysics conditions are generally non-Gaussian. Hence, algorithms derived with the assumption of Gaussianity face performance degradations under non-Gaussian conditions (Yue and Peng, 2015).

In the literature, impulsive distributions are generally utilized for statistical modelling of the seismic measurements. Lévy distribution, which is a special member of α-Stable distributions, has been used to model slip and strike spatial distributions of the well-known past earthquakes in (Lavallée et al., 2006, 2011). In addition, in (Yue and Peng, 2015), seismic amplitude measurements obtained in earthquake sites in China have been modelled with impulsive distributions.

In this case study, impulsive modelling procedure introduced in this chapter has been performed to model seismic acceleration time series measurements from well-known earthquakes which are El Centro 1979, Kobe 1995, Northridge 1994 and Kozani 1995 (COSMOS, 2017). All these data sets include three distinct measurements all of which represent the directions x, y and z of the movement. In Figure 5.10, time series for all the earthquakes in terms of x, y and z directions are plotted.

El Centro and Kobe earthquakes are characterized as strike slips whereas Northridge and Kozani are dip slips. Images to explain the strike and the dip slip surface failures are shown in Figure 5.9. Both of strike and dip slips define the failure of the earth surface during an earthquake, particularly in a strike slip horizontal motion (left and right lateral) occurs with a very little vertical motion whereas in a dip slip, motion is generally vertical (normal and reverse).

20 RJMCMC runs have been performed for each movement direction and earthquake. Estimated results for distribution families and their parameters (α and γ) are shown in Table 5.7 as averages of 20 runs. Examining the results in the table, acceleration time series measurements for El Centro and Kobe earthquakes, failures of which are strike slips, follow similar characteristics. Measurements for horizontal components which are x and y, are modelled with GG distributions whereas the vertical component z is modelled as an SαS. Results for Northridge and Kozani failures, which are dip slips, are modelled as GG for all of their directions. For all the measurements, statistical significance test

scores are low and the calculated *p*-values are very high (close to > 0.9999) providing evidence to the remarkable performance of the proposed method.
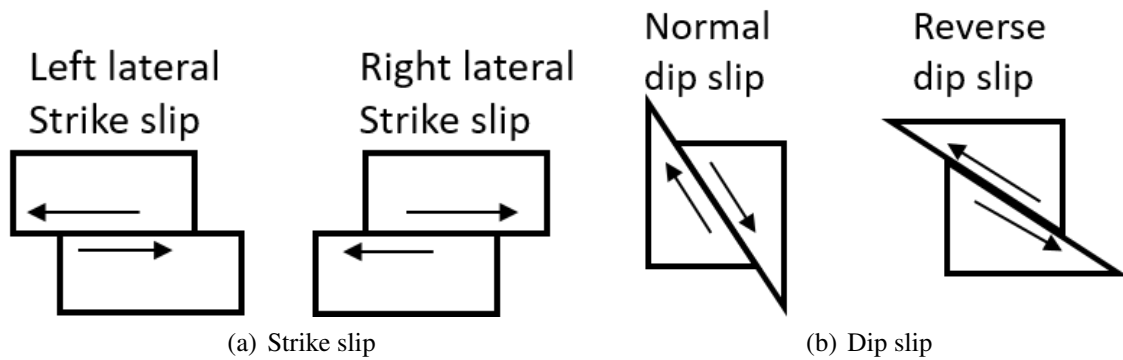

(a) Strike slip   (b) Dip slip

Figure 5.9. Earth surface faults.

Table 5.7. Modelling results for acceleration time series.

Earthquake	Component	Est. Family	Est. Shape ($\hat{\alpha}$)	Est. Scale ($\hat{\gamma}$)	KL Div.	KS Score	KS *p*-value
El Centro	x	GG	0.6816	11.8085	0.0286	0.0381	> 0.9999
	y	GG	0.5640	7.0624	0.0100	0.0250	1.0000
	z	SαS	0.9991	4.7199	0.0219	0.0232	1.0000
Kobe	x	GG	0.7355	9.6300	0.0263	0.0230	1.0000
	y	GG	0.5262	3.9587	0.0191	0.0212	1.0000
	z	SαS	1.2600	5.8229	0.0141	0.0163	1.0000
Northridge	x	GG	0.5188	3.8000	0.0114	0.0218	1.0000
	y	GG	0.5119	5.0762	0.0162	0.0216	1.0000
	z	GG	0.4500	1.8477	0.0388	0.0436	> 0.9999
Kozani	x	GG	0.4393	1.0550	0.0298	0.0136	1.0000
	y	GG	0.3930	0.5932	0.0169	0.0165	1.0000
	z	GG	0.4443	0.6658	0.0354	0.0491	> 0.9999

In Figure 5.11, 256-point FFT results for all data sets and for each direction components are plotted. Examining the figure, we can state that for El Centro and Kobe, *z* component follows different characteristics than *x* and *y* components, whereas all components follow similar characteristics for Northridge and Kozani. Frequency analysis results in Figure 5.11 are parallel with the modelling results and verify the distribution differences within components of a single earthquake.

In Figures between 5.12 and 5.15, the modelling performance of the proposed method is demonstrated visually in terms of pdfs and CDFs. Based on both of visual and statistical results, one can easily conclude that the proposed method model the seismic acceleration time series with a remarkable performance.

5.3.6.5. Graphical Evaluation by Q-Q Plots for Data Estimated as SαS

Figures between 5.16 and 5.18 show Q-Q plots for data sets estimated to be SαS. Q-Q plots provide graphical performance results and their details have been discussed in Appendix D. Examining the figures clearly shows a remarkable match between estimated distribution and the data samples. Q-Q plots for PLC-2 and MRI-D results in Figure 5.17-(a) and Figure 5.18-(d), respectively, show relatively lower performance than the others. However, this result is expected because the numerical estimation results in terms of KL and KS scores obtained for these two data sets are already higher than the others (KS scores of 0.0486 and 0.0659, respectively), but still acceptable due to p-values of 0.9999 and 0.9998, respectively.

5.4. Envelope Distribution Modelling

As another group of distributions, we are going to consider the envelope distributions in order to model the distributions of hourly average wind speed measurements and SAR images in terms of envelope distributions which are discussed in Section 5.2.2 and Appendix B.

5.4.1. Likelihood and Priors

Envelope distribution modelling study follows partially the same procedure as in impulsive data modelling section. The parameter space is formed as: $\theta = \{k, \alpha, \gamma\}$ where the family identifier k refers to Nakagami, $\mathcal{K}$, Weibull, generalized Rayleigh and Gamma for values between 1 and 5, respectively.
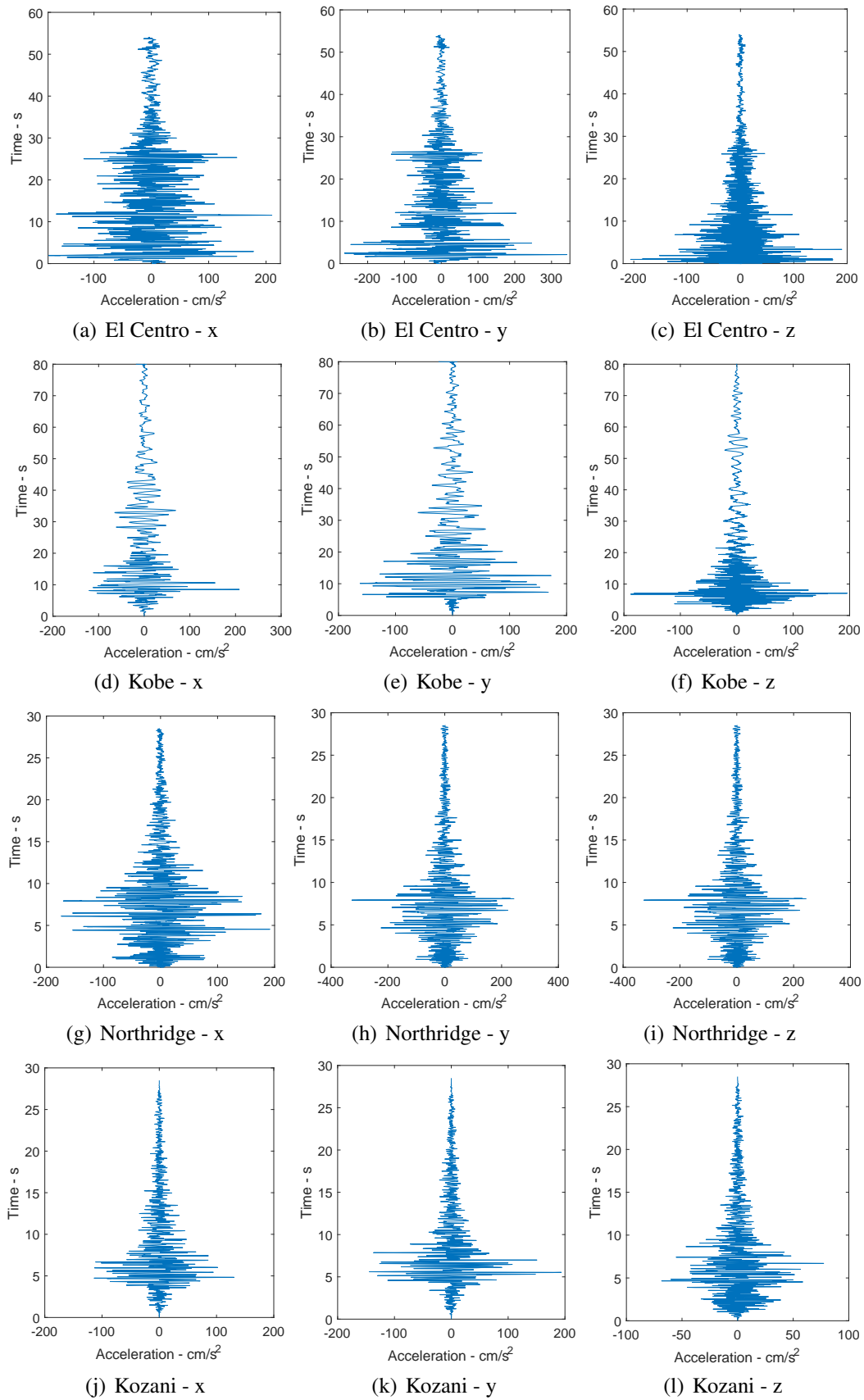
Figure 5.10. Acceleration time series for four earthquakes.
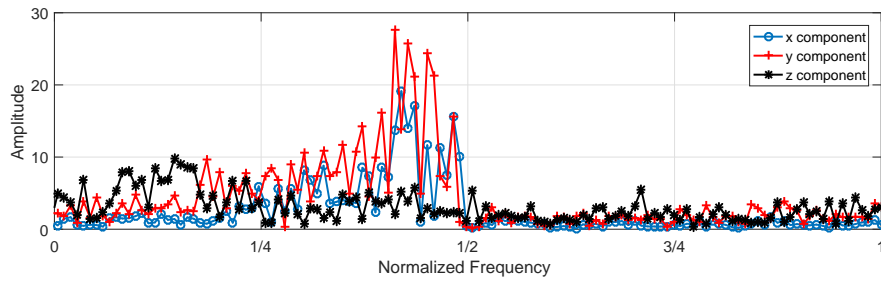
(a) El Centro
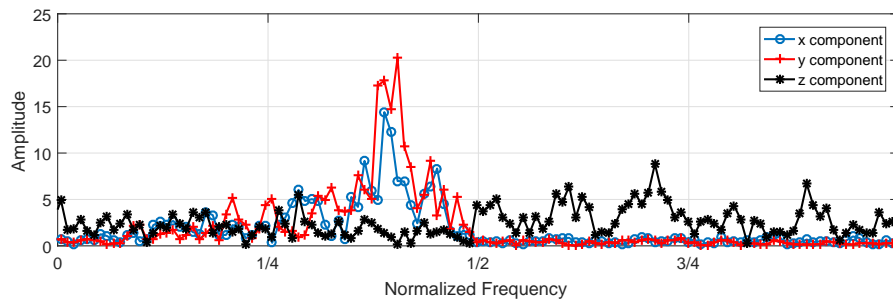
(b) Kobe
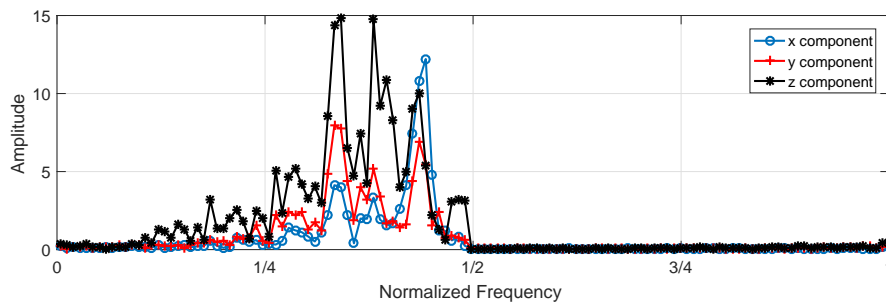
(c) Northridge
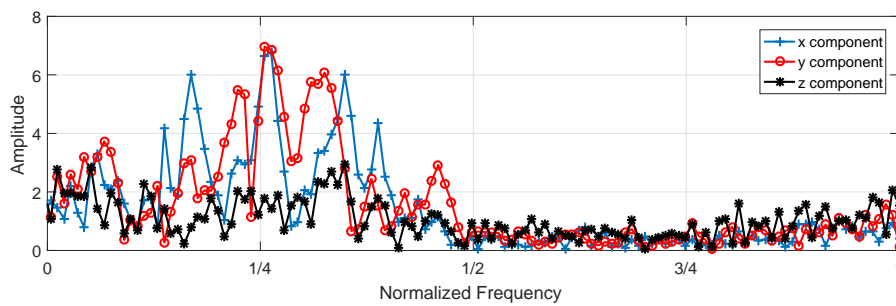
(d) Kozani

Figure 5.11. Frequency analysis of the datasets. For each component, 256 point FFT has been computed and only positive frequency results are plotted.

(a) x - pdf

(b) x - CDF

(c) y - pdf

(d) y - CDF

(e) z - pdf

(f) z - CDF
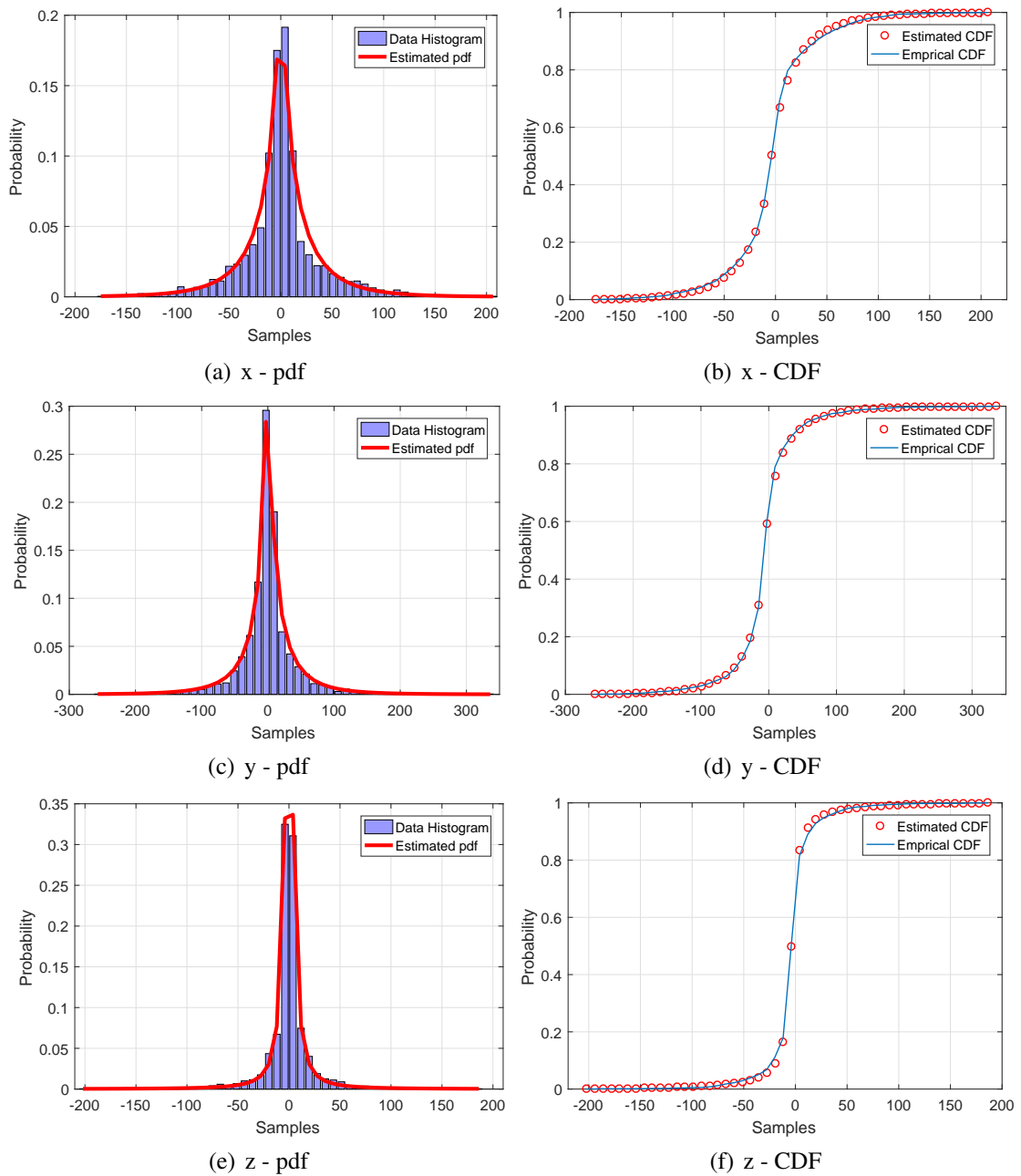
Figure 5.12. El Centro earthquake modelling results. Estimated distribution for each row are $GG_{0.6816}(11.8085)$, $GG_{0.5640}(7.0624)$, $S0.9991S(4.7199)$, respectively.
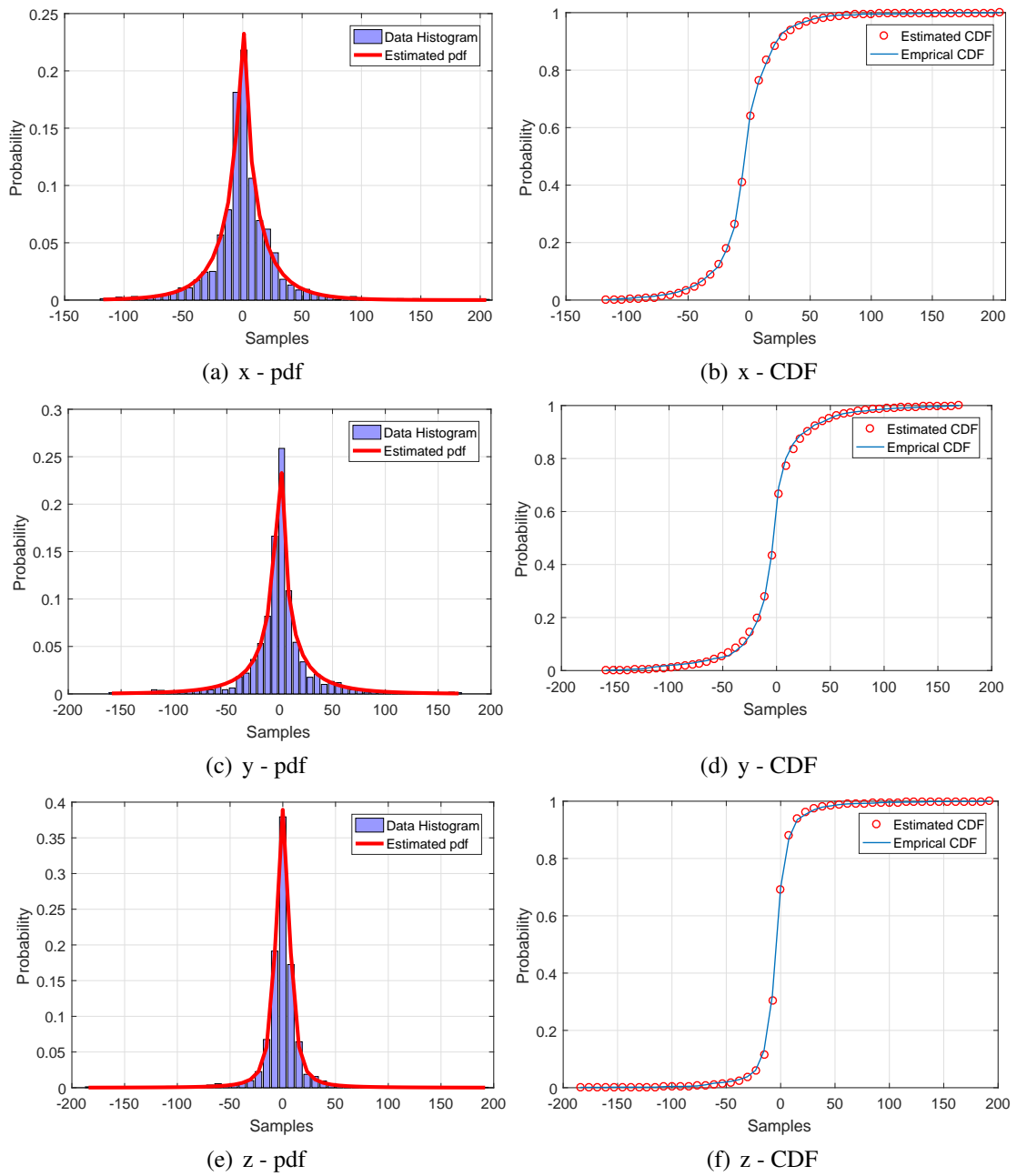
Figure 5.13. Kobe earthquake modelling results. Estimated distribution for each row are $GG_{0.7355}(9.6300)$, $GG_{0.5262}(3.9587)$, $S1.26S(5.8229)$, respectively.
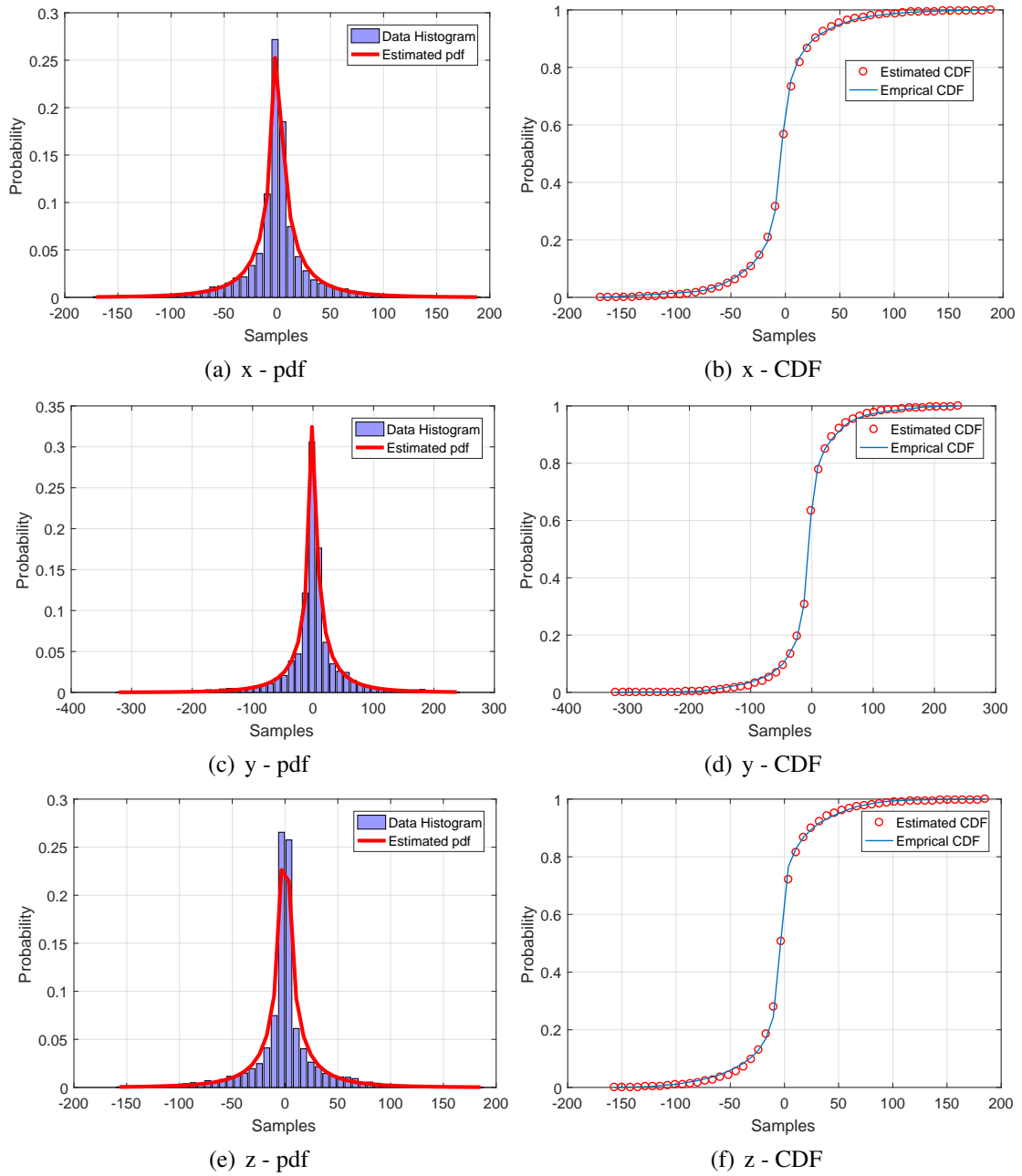
Figure 5.14. Northridge earthquake modelling results. Estimated distribution for each row are $GG_{0.5188}(3.8)$, $GG_{0.5119}(5.0762)$, $GG_{0.4500}(1.8477)$, respectively.
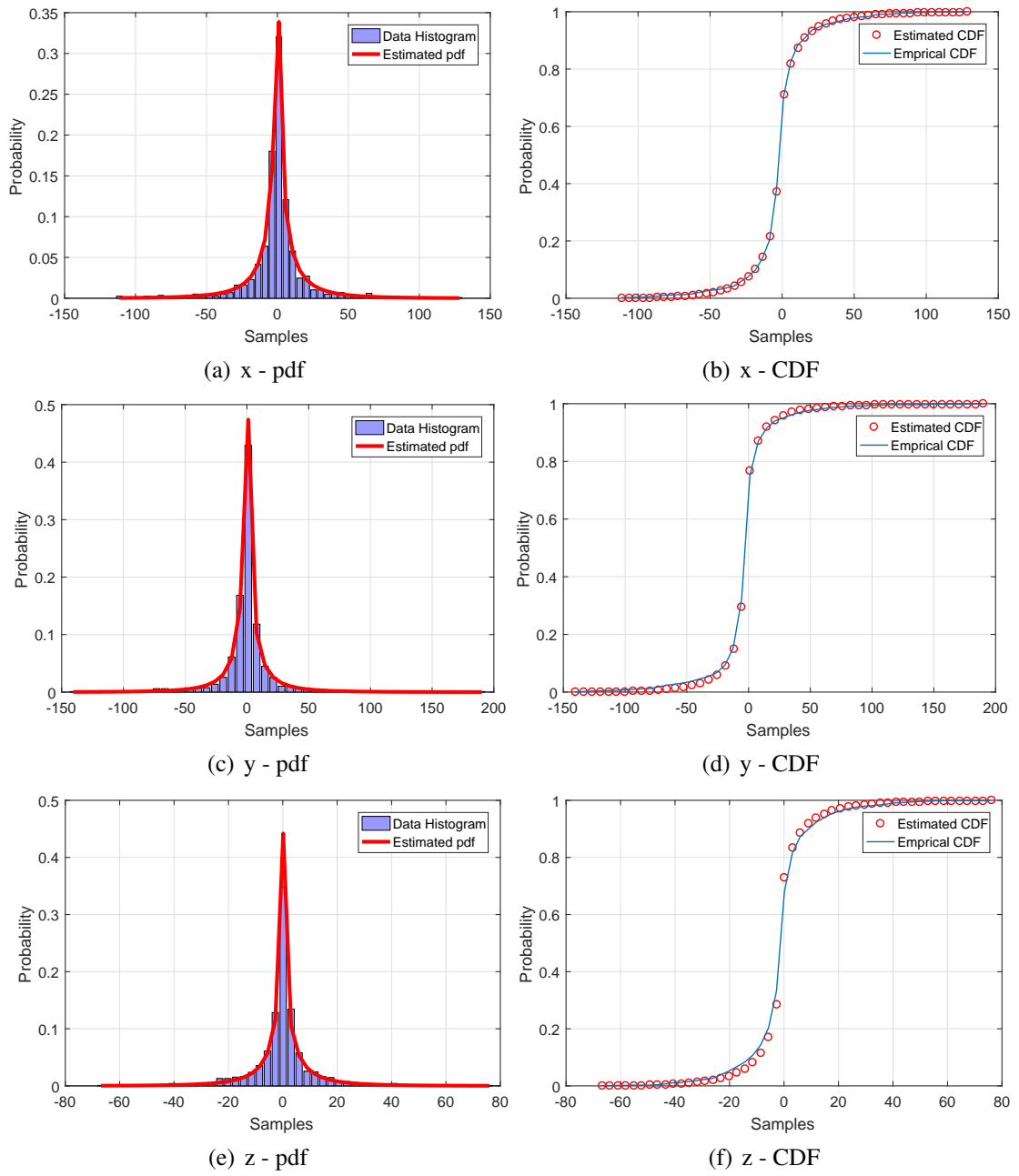
Figure 5.15. Kozani earthquake modelling results. Estimated distribution for each row are $GG_{0.4393}(1.0550)$, $GG_{0.3930}(0.5932)$, $GG_{0.4443}(0.6658)$, respectively.

(a) S1.5S(2)

(b) *PLC-1*
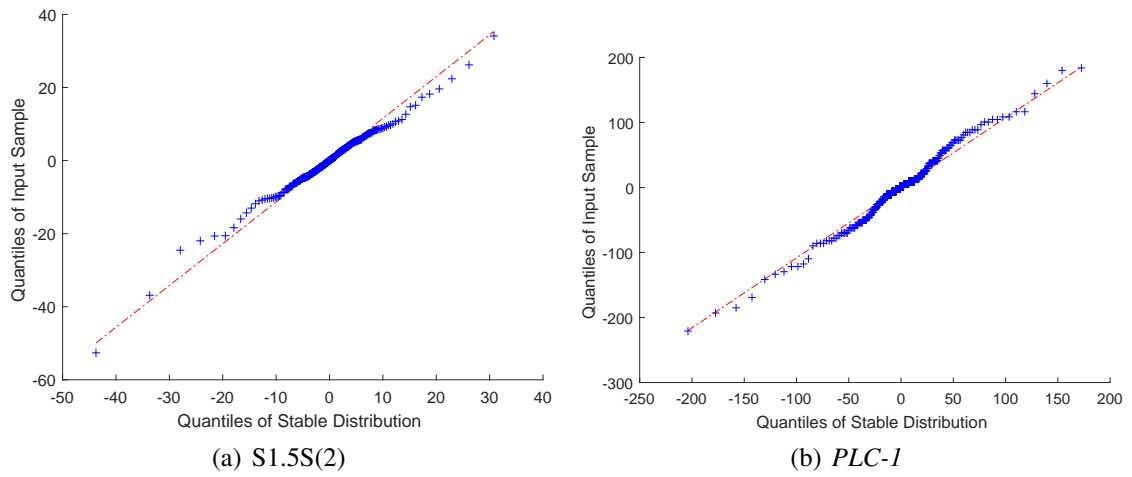
Figure 5.16. Q-Q plots for SαS estimated data sets - 1.

(a) *PLC-2*

(b) *PLC-3*

(c) SAR-V

(d) SAR-H
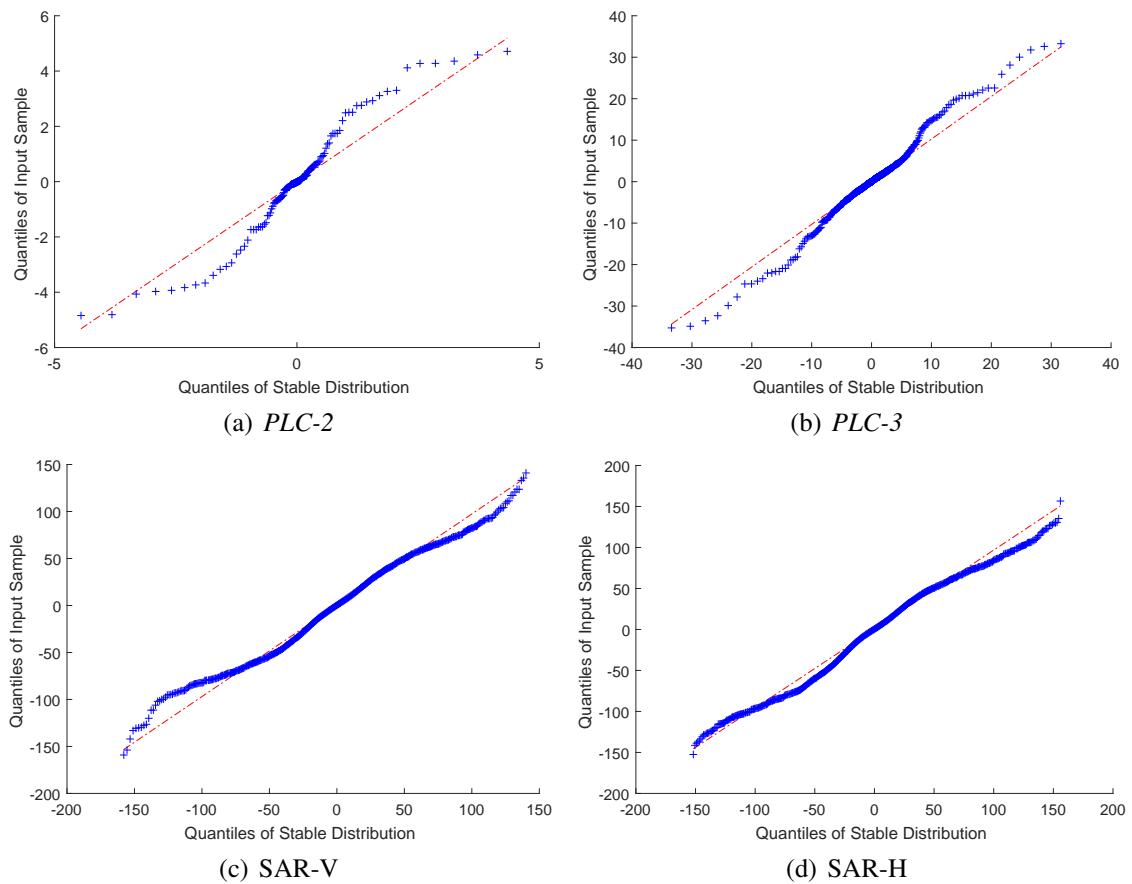
Figure 5.17. Q-Q plots for SαS estimated data sets - 2.
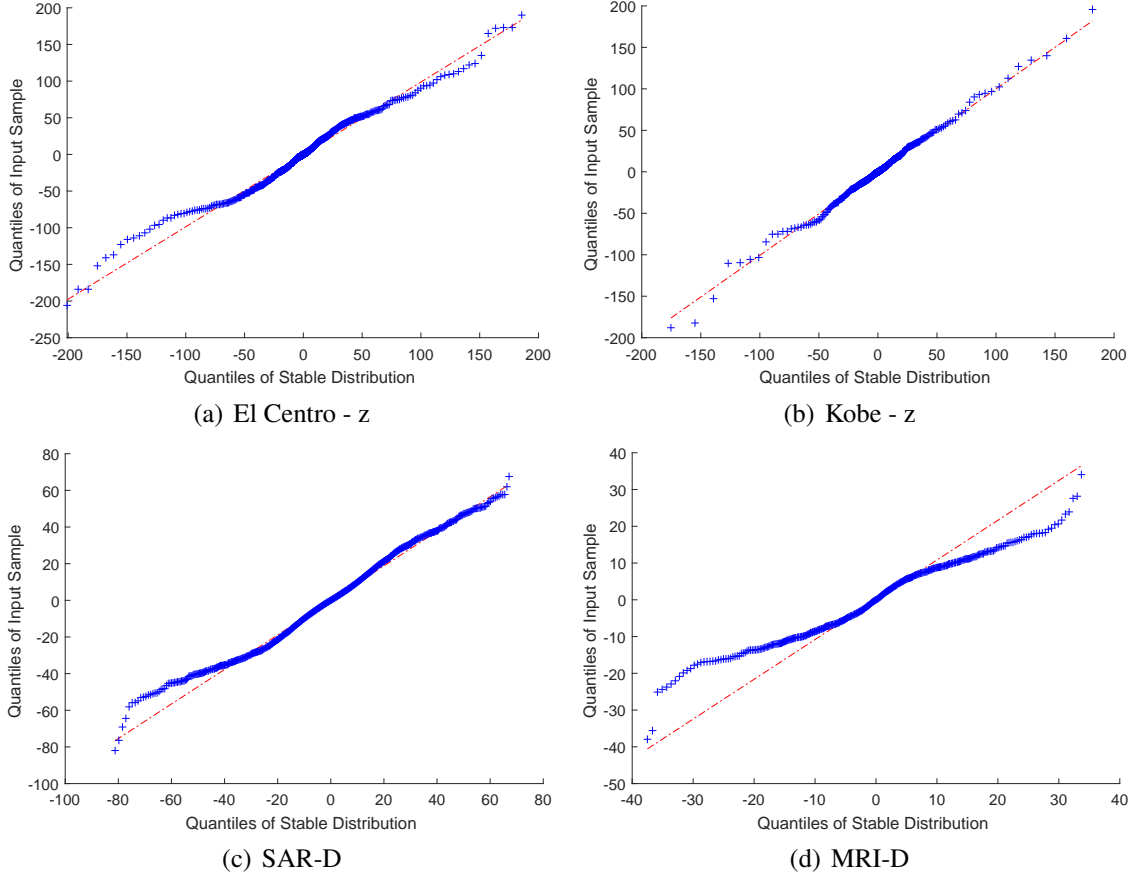
Figure 5.18. Q-Q plots for SαS estimated data sets - 3.

The target distribution, $f(\boldsymbol{\theta}|\mathbf{x})$, can be written as

$$f(\boldsymbol{\theta}|\mathbf{x}) \propto f(\mathbf{x}|k,\alpha,\gamma)f(\alpha|k)f(k)f(\gamma) \qquad (5.17)$$

where $f(\mathbf{x}|k,\alpha,\gamma)$ is the likelihood and corresponds to pdfs according to the value of k

$$f(\mathbf{x}|k,\alpha,\gamma) = \begin{cases} \prod_{i=1}^{n} \text{Nakagami}(\alpha,\gamma), & k = 1 \\ \prod_{i=1}^{n} \mathcal{K}(\alpha,\gamma), & k = 2 \\ \prod_{i=1}^{n} \text{Weibull}(\alpha,\gamma), & k = 3 \\ \prod_{i=1}^{n} \text{GenRayl}(\alpha,\gamma), & k = 4 \\ \prod_{i=1}^{n} \text{Gamma}(\alpha,\gamma), & k = 5 \end{cases} \qquad (5.18)$$

Priors are selected the same as in Section 5.4.1. The upper and lower bounds for the shape parameter α are selected as

$$f(\alpha|k) = \begin{cases} \mathcal{U}(0.5, 2) & k = 1, \\ \mathcal{U}(0, 10) & k = 2, \\ \mathcal{U}(0.5, 5) & k = 3, \\ \mathcal{U}(0, 2) & k = 4, \\ \mathcal{U}(0.5, 6) & k = 5 \end{cases} \qquad (5.19)$$

5.4.1.1. Model Moves

RJMCMC model moves for this case study are the same with the impulsive data modelling part. The algorithm in Figure 5.1 is applied in order to find a distribution model to wind speed measurements and SAR images. Life move performs the same procedure as in 5.3.5.1. Intra and inter class switch moves are the same however this time with different sampling operations for γ and α.

In both intra and inter class switch moves, instead of FLOM based proposals for γ transitions, p^{th} order moment based γ transitions are defined. For a given data, $\mathbf{x}$, when a transition from $\{k, \alpha, \gamma\}$ to $\{k', \alpha', \gamma'\}$ occurs, the learned information in the previous iterations is transferred to the candidate distribution space by keeping the pth order moment constant to the value at the most recent distribution space

$$E_k(\mathbf{x}^p) = E_{k'}(\mathbf{x}^p) \qquad (5.20)$$

where pth order moments for all the families are given in Appendix B.

For both of the switch moves, candidate values for the shape parameter α are proposed by performing the same sampling strategy as in Section 5.3.5.3. This sampling strategy includes sampling from a discretized Laplace ($\mathcal{DL}(\cdot)$) distribution the location parameter of which is equal to the most recent shape parameter α and the scale parameter is Γ.

5.4.2. Case Study 5: Wind Speed Distribution Modelling

In this case study, we have experimentally modelled the wind speed distribution by applying the proposed method. For this purpose, hourly average wind speed measurements have been used from five distinct places in the world which are Karaburun (Turkey), Basel (Switzerland), Bandon (Oregon, USA), Manson (Washington, USA) and Vredenburg (South Africa). All data sets have a length of 1 year (8760 samples) and in order to reduce the computational load, only half of these samples have been used in the simulations by performing a downsampling operation by 2. In Figure 5.19, time plots for all the locations are plotted.

As candidate distributions, we are going to deal with four distributions which are Nakagami, $\mathcal{K}$, Weibull and Gamma distributions (details are given in Appendix B). Due to their popularity in wind speed modelling applications, Weibull and Nakagami are natural choices. We have added the $\mathcal{K}$ and Gamma distributions to have heavier tailed distributions in the model space than Weibull and Nakagami. A selection like this is not unique, and interested users can vary the model space by adding distributions like, χ^2, Rice, log-normal, etc.

The algorithm starts with a Rayleigh distribution with initial values $k^{(0)} = 3$ and $\alpha^{(0)} = 2$. RJMCMC moves have been proposed with probabilities $P_{\text{life}} = 0.4$, $P_{\text{intra-cl-sw}} = 0.3$ and $P_{\text{inter-cl-sw}} = 0.3$. Hyperparameters for γ prior distributions are set to $a = b = 1$ and variance of proposal distribution for γ in life move is set to $\xi_{scale} = 0.01$. Scale parameter Γ of the discretized Laplace distribution for intra-class-switch move is selected as 0.5. For switch moves, moment order, p, is selected as 1 in order to use *mean value* as the common property between candidate distributions. For all data sets the proposed method has been run 40 times each of which lasts 5000 iterations. In Table 5.8, estimated

Table 5.8. Modelling results for acceleration time series.

| Location	Est. Family	Est. Shape ($\hat{\alpha}$)	Est. Scale ($\hat{\gamma}$)	KL Div.	KS Score	KS p-value
Karaburun	$\mathcal{K}$	0.6392	2.2369	0.0409	0.0416	> 0.9999
Bandon	Weibull	1.3599	2.1810	0.0228	0.0246	1.0000
Manson	Weibull	1.2414	0.8412	0.0237	0.0354	> 0.9999
Basel	$\mathcal{K}$	0.6899	1.2688	0.0242	0.0488	> 0.9999
Vredenburg	Nakagami	1.0973	25.8843	0.0117	0.0229	1.0000

distribution families and resulting scale and shape parameters are shown together with KL and KS test results. Examining the results in Table 5.8, we can state that locations are divided into three distribution groups. First group which consists of data from Bandon and Manson has been estimated to be Weibull distributed. Karaburun and Basel constitute the second group and wind speed measurements from these locations are $\mathcal{K}$ distributed. These results are more meaningful if we examine the histograms of the data from each location in Figure 5.20. Bandon and Manson have much lighter tails than Karaburun and Basel. Thus, the proposed method chose Weibull for Bandon and Manson whereas it chose $\mathcal{K}$ for heavy tailed measurements. Examining the histograms in Figure 5.20, Vredenburg measurements have a different characteristics from the others and Nakagami distribution is the best match to this type of measurements. The CDF plots in Figure 5.21 and results of the significance tests in Table 5.8 demonstrate the remarkable modelling performance of the proposed method both visually and numerically.

5.4.3. Case Study 6: Distribution Modelling of SAR Images

SAR imaging and statistical modelling of it have been favorable in the literature. General practice is to model the distribution of SAR images with Weibull, $\mathcal{K}$, Rayleigh, Gamma, etc. In a previous study (Kuruoglu and Zerubia, 2004), authors have stated that a generalization of Rayleigh distribution, which is the heavy-tailed Rayleigh distribution, performs modelling of SAR images of urban areas better than any other envelope distributions. Thus, we have chosen five candidate distributions for this purpose, which are Weibull, $\mathcal{K}$, Nakagami, Gamma, and generalized Rayleigh distributions.

Six SAR images have been used in this case study in order to investigate the performance of the proposed method. All of the images are downloaded from (Artemis Inc., 2017) and categorized into three groups according to their scenes which are urban (2 images), forrest (2 images) and agricultural (2 images). Additionally, there are different frequency bands used in the acquisition of SAR images. The images used in this case study were chosen to be of different frequency bands within each group. These bands are X, L and UHF bands. In Figure 5.22, SAR images utilized in this case study are depicted.

The simulation environment for SAR image distribution modelling has been selected as the same with wind speed modeling simulations with one exception. Since positive order moments of generalized Rayleigh distribution are not defined, moment degree, p, for the proposal distribution used in the scale parameter estimation step is chosen
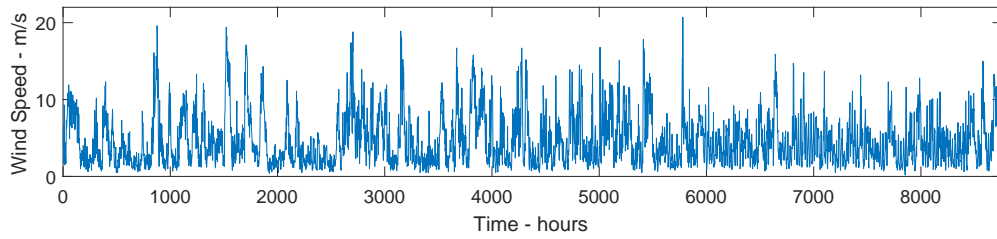
(a) Karaburun
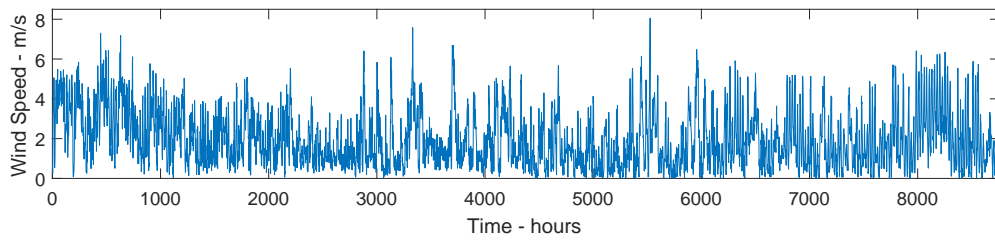
(b) Bandon
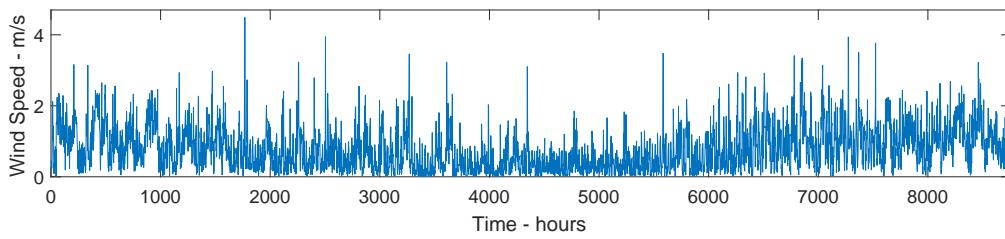
(c) Manson
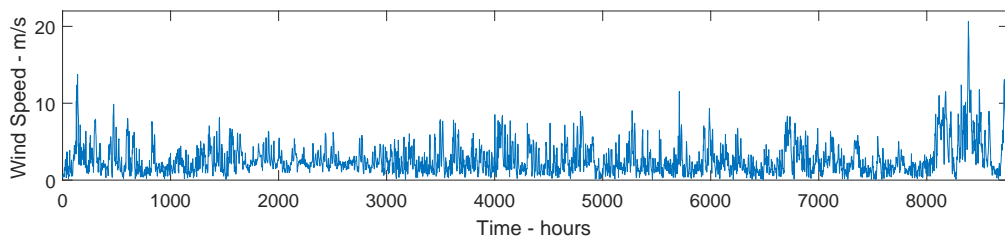
(d) Basel
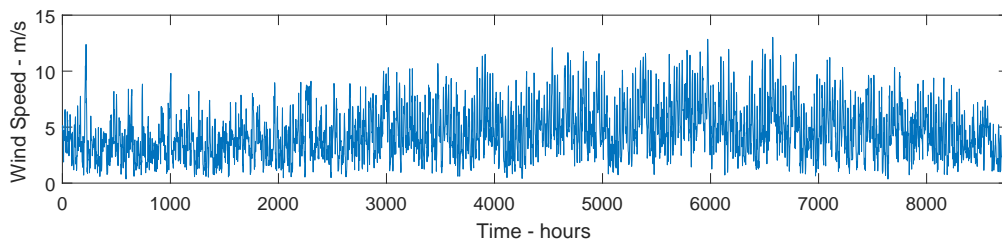
(e) Vredenburg

Figure 5.19. Hourly average wind speed measurements.

(a) Karaburun

(b) Bandon

(c) Manson

(d) Basel

(e) Vredenburg
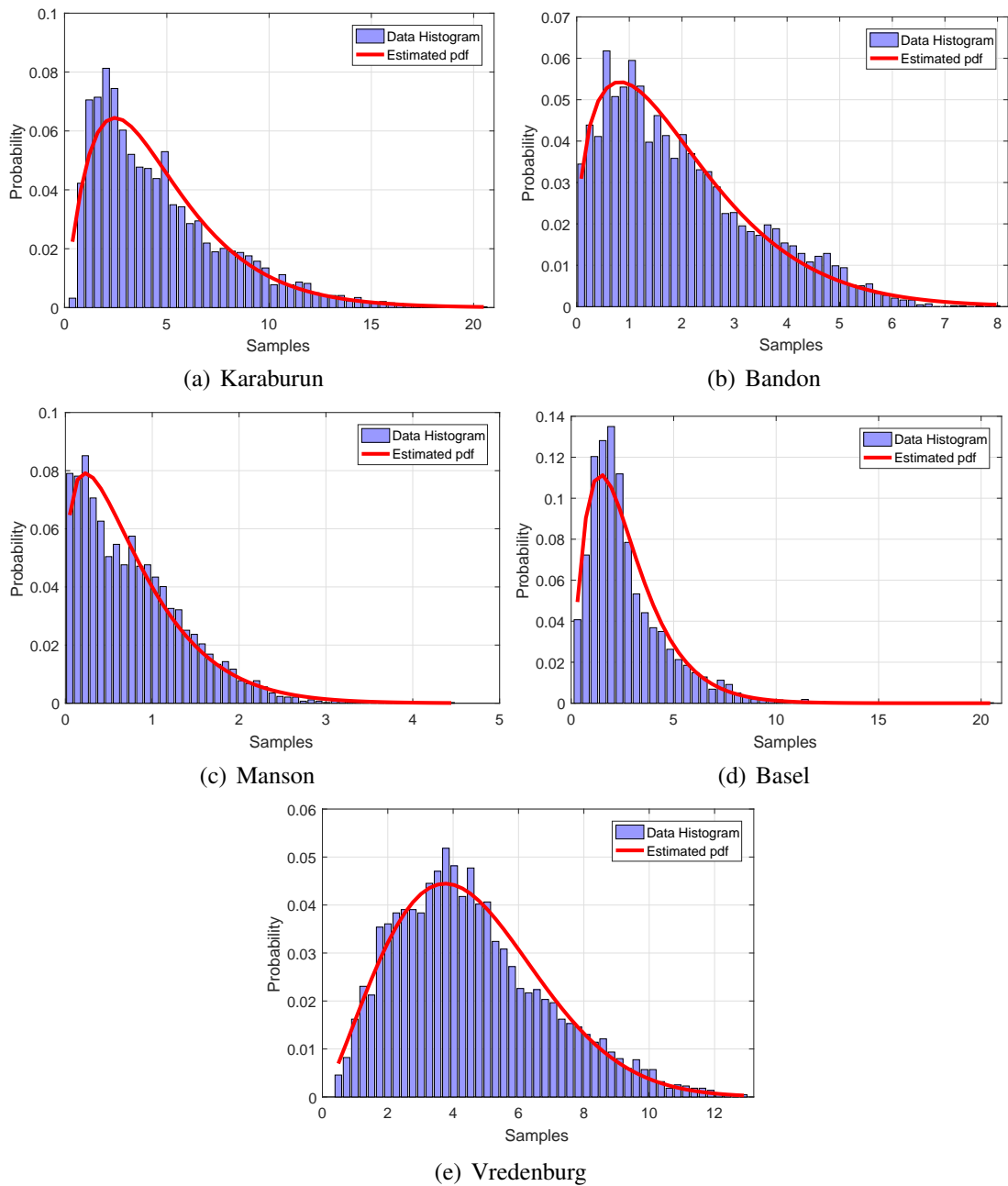
Figure 5.20. Wind speed distribution modelling results 1 - Estimated pdfs. Estimated distributions are (a): $\mathcal{K}(0.6392, 2.2369)$, (b): Weibull(1.3599, 2.1810), (c): Weibull(1.2414, 0.8412), (d): $\mathcal{K}(0.6899, 1.2688)$, (e): Nakagami(1.0973, 25.8843).

(a) Karaburun

(b) Bandon

(c) Manson

(d) Basel

(e) Vredenburg
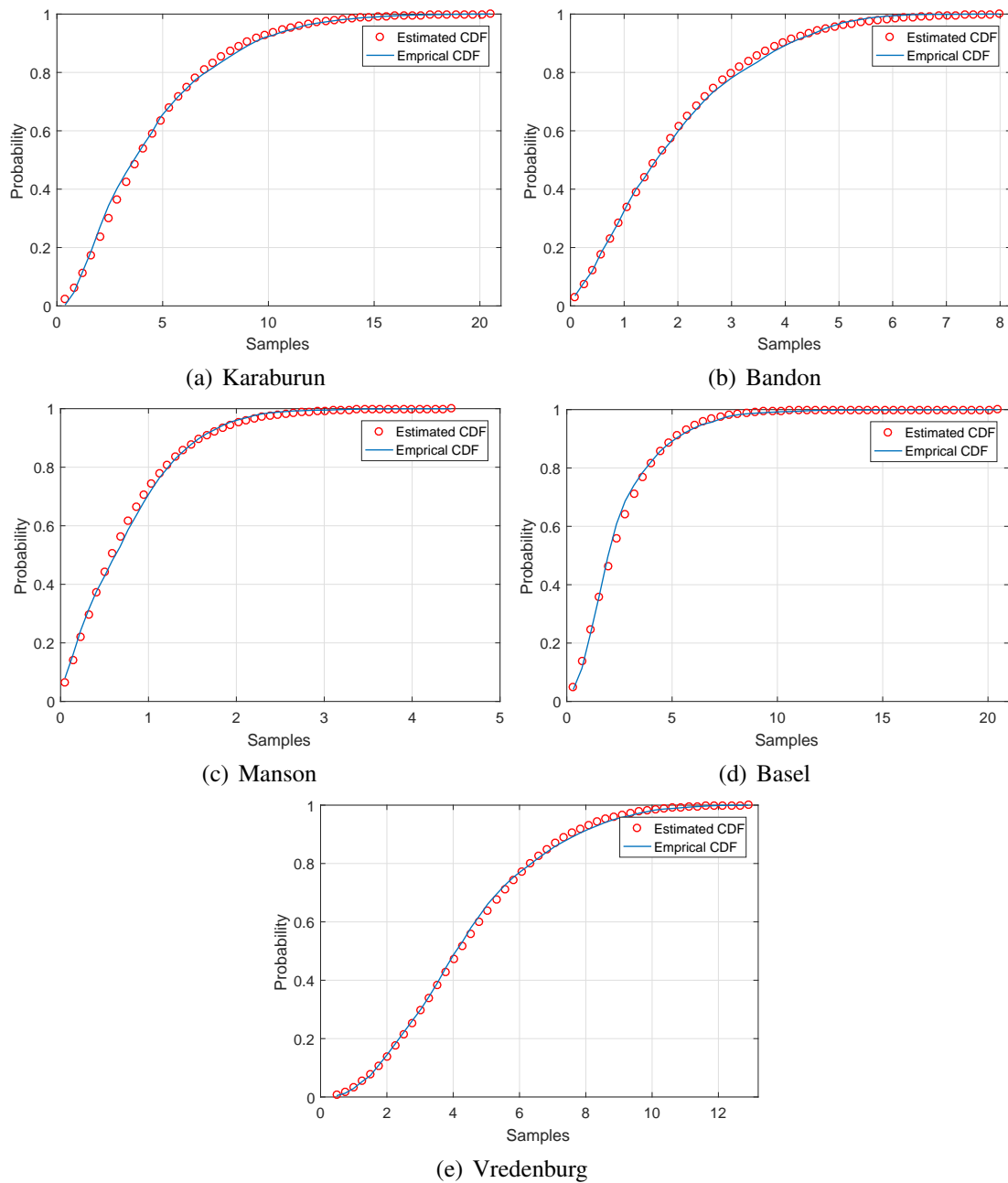
Figure 5.21. Wind speed distribution modelling results 2 - Estimated CDFs. Estimated distributions are (a): $\mathcal{K}(0.6392, 2.2369)$, (b): Weibull(1.3599, 2.1810), (c): Weibull(1.2414, 0.8412), (d): $\mathcal{K}(0.6899, 1.2688)$, (e): Nakagami(1.0973, 25.8843).

(a) SAR1

(b) SAR2

(c) SAR3
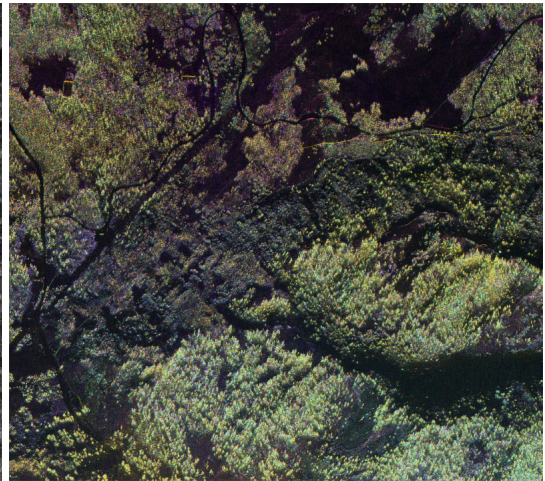
(d) SAR4

(e) SAR5

(f) SAR6

Figure 5.22. Images used for SAR distribution modelling.
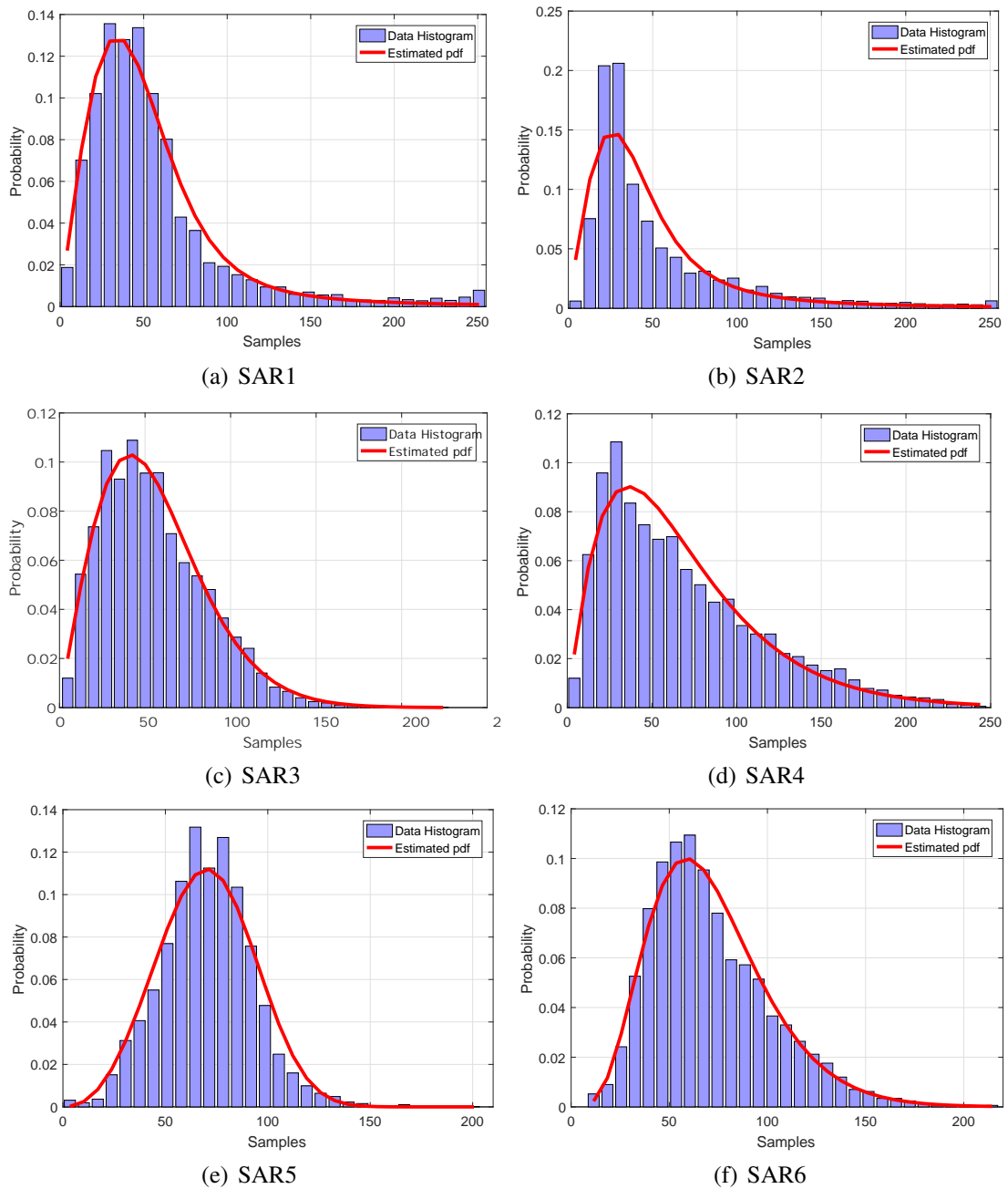
Figure 5.23. SAR image distribution modelling results 1 - Estimated pdfs. Estimated distributions are (a): GenRayl(1.6489, 234.820), (b): GenRayl(1.4479, 96.425), (c): $\mathcal{K}$(7.933, 10.665), (d): Gamma(2.115, 0.0308), (e): Weibull(3.333, 78.222), (f): Gamma(5.577, 0.078).

(a) SAR1

(b) SAR2

(c) SAR3

(d) SAR4

(e) SAR5

(f) SAR6
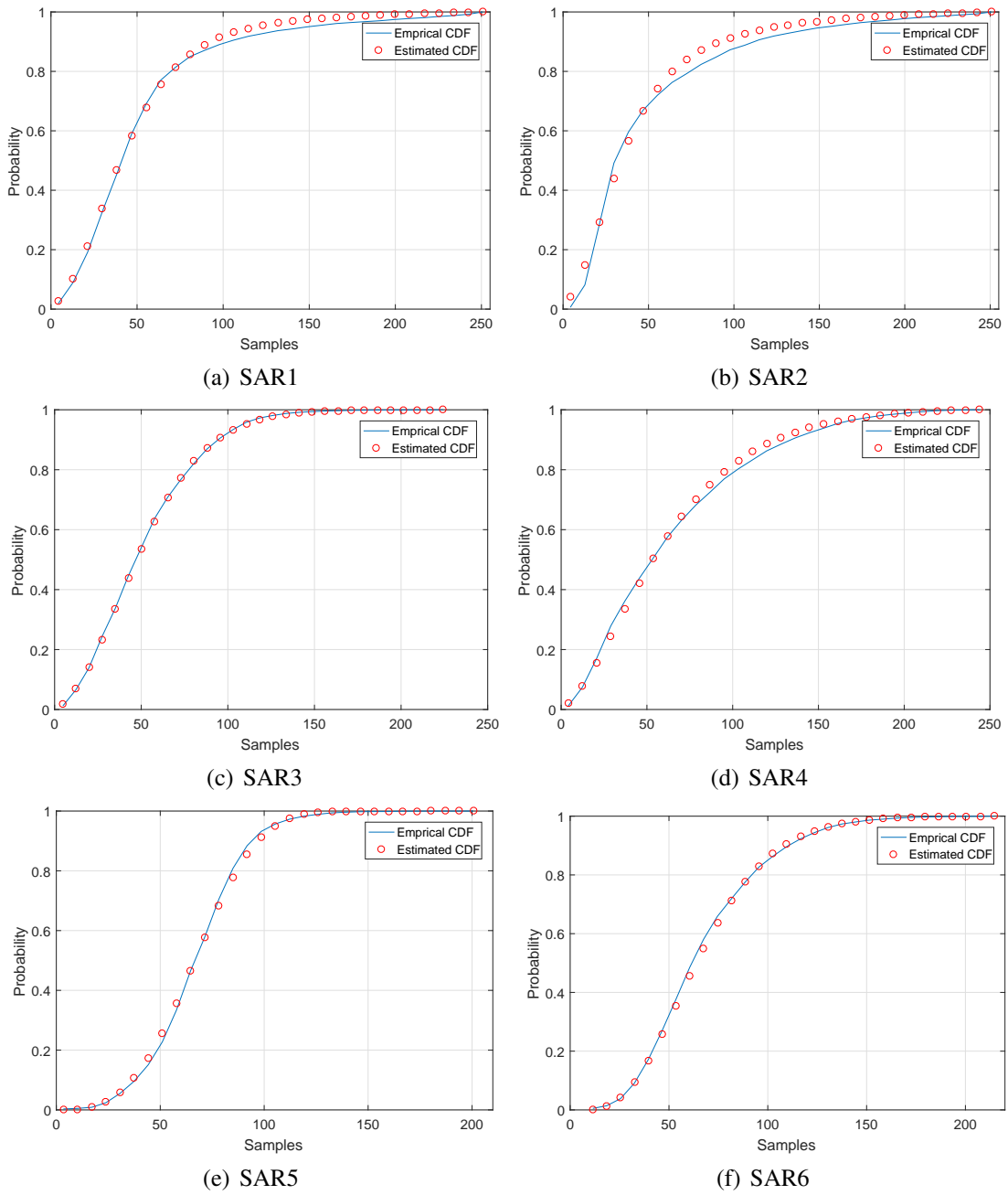
Figure 5.24. SAR image distribution modelling results 2 - Estimated CDFs. Estimated distributions are (a): GenRayl(1.6489, 234.820), (b): GenRayl(1.4479, 96.425), (c): $\mathcal{K}$(7.933, 10.665), (d): Gamma(2.115, 0.0308), (e): Weibull(3.333, 78.222), (f): Gamma(5.577, 0.078).

as -1. In Table 5.9, estimated distribution parameters for six example images are shown together with numerical performance metrics KL divergence and KS score. SAR image distribution modelling performance of the proposed method is remarkable according to statistical significance results for all six SAR images. In particular, modelling results of urban scene SAR images provide empirical evidence to the previous studies (Kuruoglu and Zerubia, 2004) and (Achim et al., 2006) by estimating generalized Rayleigh distribution for these images. Moreover, results for forrest and agricultural images are also remarkable and provide p-values of 1.0000 for all four images. Estimated pdfs and CDFs for all the images are plotted in Figures 5.23 and 5.24, respectively. Both pdf and CDF estimates shows the estimation performance of the algorithm visually.

Table 5.9. Modelling results for SAR Images.

| Image	Frequency Band	Scene	Est. Family	Est. Shape ($\hat{\alpha}$)	Est. Scale ($\hat{\gamma}$)	KL Div.	KS Score	KS p-value
SAR1	X	Urban	GenRayl	1.6489	234.8200	0.0242	0.0269	1.0000
SAR2	L	Urban	GenRayl	1.4479	96.2450	0.1024	0.0718	0.9992
SAR3	X	Forrest	K	7.9330	10.6650	0.0166	0.0163	1.0000
SAR4	UHF	Forrest	Gamma	2.1150	0.0310	0.0222	0.0271	1.0000
SAR5	X	Agricultural	Weibull	3.3330	78.2220	0.0234	0.0296	1.0000
SAR6	L	Agricultural	Gamma	5.5770	0.0780	0.0151	0.0307	1.0000

5.5. Conclusion

In this chapter, the proposed method, trans-space RJMCMC, has been used in estimating the distribution of given sets of data for both synthetically generated and real measurements. Simulation studies verify the remarkable performance in modelling the distributions in terms of both visual and numerical tests. KL and KS tests show the numerical results are statistically significant in terms of p-values which are higher than 0.85 for all the example data sets. In order to demonstrate the success of the proposed method, PLC impulsive noise measurements were also modeled by the ML method and the ML estimation results were compared to that of RJMCMC. Examining this comparison, we can clearly state that estimated families and the corresponding distributions via RJMCMC are the same as the results from the ML method. This shows the success of the proposed method and confirms the accuracy of the estimated results.

The RJMCMC approach, developed in the previous chapters of this thesis, have

reached its final form in this chapter and as case studies, probability distributions are estimated including all their parameters. The most important contribution of this application is that the transitions between different distribution families are realized by matching some common features among the families. Thus, transitions between the model spaces that are irrelevant of each other are performed more efficiently by increasing the mixing of the algorithm and accelerating the convergence to the correct family.

One can expect higher benefits from the trans-space RJMCMC compared to considering different model classes separately in the cases when the different model class spaces have intersections to exploit. The intersections for the trans-distributional RJM-CMC considered in this chapter have been the common distributions in the impulsive noise families. They made it possible to use the mapping functions benefitting from the moments (norms) of the observed data. These functions in turn have enabled to transfer the information learned while searching in one family to the subsequent search after an inter-class-switch move.

This new perspective provided for RJMCMC also lends itself to much more complex selection of model classes, such as kernel selection and classification with support vector machines (SVM), stable process selection, automatic model selection.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

Expressing an observed real life signal with a mathematical model is of great importance in solving the problems and improves the analysis of these problems. Although various approaches have been developed in the literature to select "the best model", the procedures pose serious problems especially when the number of candidate models is high. A Bayesian numerical method, RJMCMC, offers an outstanding performance in solving model uncertainty because it allows efficient exploration of spaces with varying dimensions without requiring an exhaustive search. Despite its potential, the use of RJMCMC has remained limited to the trans-dimensional case in the literature.

In this thesis, we have shown that the formal presentation of RJMCMC (Green, 1995) can be used to explore more generic model spaces. We have shown that RJMCMC is a general Bayesian model selection method that offers more general transitions than just between different dimensions of the models of the same type.

Firstly, we have created an RJMCMC based general framework for linear-in-the-parameters nonlinear models PAR, PMA and PARMA. Motivated by AR model estimation study in the methodology chapter, we have designed RJMCMC to explore different structural models and particularly, switch between linear and nonlinear models using the trans-structural approach in an anomalous way. This approach provides a means to estimated polynomial nonlinearity degree of the models. Numerical results demonstrated promising performance of the proposed method in estimating model orders and the corresponding model coefficients concurrently for all three nonlinear time series models. Our first application of RJMCMC to nonlinear model estimation has made it clear that RJMCMC is a method that can be used for model selection among different structural models instead of focusing only on different dimensional space transitions. This utilization leads us to demonstrate model selection applications involving different classes of models than those used in this first application, PAR, PMA and PARMA.

Extending the nonlinear model estimation studies for PAR/PMA/PARMA models, we suggested a trans-structural RJMCMC framework, which enables RJMCMC transitions across multiple parameter dimensions concurrently regardless of the dimension sizes. Transitions of this kind are performed with a new type of RJMCMC move, called switch. This new type of move also includes classical trans-dimensional transitions and

was described as a more general move. This approach was presented compactly for its use in VSI problems and, unlike the general approach, allowed for the estimation of the non-linearity degree of the unknown system. The approach we proposed in this thesis, offers a very significant advantage especially when the unknown systems have variable degrees of nonlinearity. In experimental analysis, apart from the synthetically generated data case, the proposed method was used in the estimation of a nonlinear communication channel of a synthetically generated OFDM communication system. Despite not having known the correct model orders, our proposed method showed approximately the same performance in coefficient estimation as the informed NLS method which perfectly knows the model orders beforehand.

In various application areas, information such as what are statistical properties of the observed data and what kind of distribution family it comes from is of great importance especially in the development of parameter estimation methods. The Gaussian assumption has seriously affected various application areas in the literature such as communication channel models, wind speed, finance, and made possible the use of methods such as least squares and mean square error. However, in those cases where the data are more impulsive and skewed, Gaussian based methods cause serious problems. In such cases, methods that can suggest us the choice of "the best (or suitable) pdf model" avoiding searching all possible models, are very useful. In this thesis, we proposed an approach on RJMCMC by using it beyond trans-dimensional sampling framework, so that it can provide transitions between different probability distribution families.

The approach suggested in Chapter 3, demonstrated transitions between different structural models by performing a cascaded RJMCMC procedure in cases of creation of a new parameter space (e.g. linear ($p = 1$) $\rightarrow$ nonlinear ($p > 1$)) or removal of the existing parameter space. With the new move proposed in Chapter 4, the proposed approach in this thesis performed transitions across multiple parameter spaces simultaneously and had a hierarchical structure. Transitions in both of the applications could be accomplished with the combination of newly born coefficients or the death of existing ones, making it possible to explore linear and nonlinear model spaces in the same setup. Exploring irrelevant spaces of different distribution spaces is much more challenging than the previous applications and cannot be accomplished with the same approach. Extending the RJMCMC moves are suggested in Chapter 4, where the transitions between different distribution spaces are accomplished by matching common properties of the distribution families (e.g. norm). That is, the information learned at the most recent distribution was carried to a new distribution and we enabled RJMCMC to choose the best model among the different

distribution families. Simulation studies provide remarkable performance results in estimating the distribution of given data set in terms of impulsive or envelope distributions. Visual results were also supported numerically with statistical significance tests and all these results demonstrate the remarkable success of the method.

The most important contribution of this thesis is invalidating the widespread opinion that RJMCMC algorithm is a Bayesian tool which mainly enables dimension search. Although the original formulation of Green was more general, the opinion that RJMCMC is a trans-dimensional model estimation method has been strengthened by generally selected applications. The thesis reveals that RJMCMC allows us to search in an indefinitely extensible union of multiple spaces.

The multiplicity of the searched spaces does not decrease the efficiency of the algorithm, i.e. the learned information in previous RJMCMC iterations can be transferred without any loss to the search in the next space which is made possible by describing the spaces subject to jump with some common attribute. This attribute was "norms" when jumping between impulsive distribution families. Keeping this attribute constant during an inter-class move, the point to jump in the next space is determined by the point in the most recent space. The applications in the thesis give several examples to this efficient jump. Additionally, we have seen in some experiments in which "blind" model transitions were performed, the algorithm did not estimate the correct family in most of the cases without a common feature-based model transition (trans-space moves), and got stuck in the neighborhood of a distribution far from the correct one. This result emphasizes the importance of using the approach of transferring the learned knowledge to the candidate model space, that is, the trans-space approach proposed in this thesis, when it is desired to use RJMCMC to switch between different classes of models.

6.1. Current and Future Research Directions

With this thesis, we contributed to the literature with a generalization on RJMCMC beyond its trans-dimensional usage. The applications are not limited to the ones given in this thesis and we opened paths to many future research directions which will be explained in this section.

6.1.1. Multivariate RJMCMC

In this thesis, all models used in model selection and estimation studies are univariate models. In the literature, the use of multivariate approaches, especially in the most recent machine learning algorithms, has increased significantly. Additionally, multivariate models such as vector autoregressive models also have applications in economics, meteorology, etc.

A multivariate use of the RJMCMC will also offer using it in the applications mentioned above. The trans-space approach demonstrated by this thesis brings to light the potential that RJMCMC can make transitions among multivariate models. Such an approach would reveal the feasibility of RJMCMC in spatio-temporal models. This will also reveal that RJMCMC can be utilized in estimating spatio-temporal models e.g. for more complex prediction tasks or for gene expression studies.

6.1.2. Time Varying RJMCMC

Time-varying systems have an important place in signal processing especially in economics, channel modelling, tracking, etc. As the Bayesian approach, sequential Monte Carlo algorithms or particle filtering operations are used in time-varying model estimation, especially in applications such as tracking, financial modelling, machine learning, etc (De Freitas et al., 2001).

The feasibility of the RJMCMC in time-varying modelling has also become an important research direction after this thesis. The ability of the RJMCMC to switch between different models shown in this thesis motivates a potential for adaptation to time-varying models. An RJMCMC approach combined with methods like Sequential Monte Carlo will make a very significant contribution to the Bayesian learning literature.

6.1.3. Bayesian Prediction with Model Selection of Real Measurements

The approach presented in this thesis on nonlinear (polynomial) time series models PAR, PMA and PARMA were implemented for only synthetically generated data sets. These models can be favourable for various meteorological prediction studies such as

wind, solar radiation, etc. In a previous study (Karakuş et al., 2017c), we have shown that PAR models achieve remarkable performance among different nonlinear statistical models such as *artificial neural networks* (ANN) and *adaptive neuro fuzzy inference system* (ANFIS) models for wind speed/power predictions up to 24 hours ahead of time.

A Bayesian prediction scheme based on RJMCMC can be constructed by using the procedure presented in this thesis. This scheme offers a complete Bayesian prediction framework which performs model estimation and prediction at the same time. In addition, the proposed approach in this sub-section is not limited to the application of wind speed prediction. Other meteorological events such as solar radiation, rainfall or prediction studies in finance are also object of interest for RJMCMC based prediction studies.

6.1.4. Non-Gaussian PAR Modeling of Real Data

RJMCMC based estimation procedures for PAR, PMA and PARMA models in this thesis were constructed using Gaussian excitation sequences. Non-Gaussian polynomial nonlinear model like Stable-PAR are also applicable to various studies such as wind speed prediction, etc. especially to ones possessing highly nonlinear nature. We believe that expressing the nonlinear nature of these types of real life measurements with both a nonlinear model and a non-Gaussian excitation provides advantage especially in prediction studies. In the continuation of our research, we plan to investigate the non-Gaussian PAR models and their applications on prediction studies.

6.1.5. Nonlinear Channel Equalization

In this thesis, we provided nonlinear channel estimation study for an OFDM communication system with various modulation schemes. Our study was limited to only identification of the unknown system. However, by taking into consideration the properties of the Volterra systems, nonlinear channel estimation approach in this thesis can be easily extended to a Bayesian channel equalization study and the presented procedure turns into a fully Bayesian equalizer for Volterra based nonlinear channels.

6.1.6. Kernel Selection and Classification with SVM

Support vector machines (SVMs) are supervised learning procedures in machine learning which analyse the data for classification or regression. In classification problems, a training data set including features and real class labels for each sample are given initially and SVM creates a decision criterion to classify the data according to this criterion. In SVM, a procedure named as Kernel Trick is used to perform nonlinear classification. There are various kernels such as polynomial, radial basis, multi-layer perceptron, splines, etc. Generally, Gaussian radial basis kernel is assumed to be the best kernel for classification, however kernel performance changes according to data sets. Thus, selecting the appropriate kernel is crucial in SVM based classification operations.

The procedure which was presented in this thesis, can be used to select an appropriate kernel for a given training data set. Thus, we can use RJMCMC in SVM training procedure in a complete manner which selects the best kernel and estimates its required parameters. We also plan to apply our RJMCMC procedure in this thesis to SVM classification problems.

6.1.7. Selection of Stable Processes with RJMCMC

A Stable process is a type of stochastic process whose finite-dimensional probability distributions are α-stable. There are various types of stable processes such as linear, harmonizable, sub-Gaussian, etc. Linear stable processes are AR, MA or ARMA processes where the driving noise sequence is i.i.d. α-stable (Swami and Sadler, 1998). An α-sub-Gaussian process is obtained multiplying a Gaussian process with a positive $\alpha/2$ stable random variable (Tsihrintzis and Nikias, 1997). Harmonizable stable processes are an important class of stable processes which do have a spectral representation (Tsihrintzis et al., 1998) (for details see (Nikias and Shao, 1995).) In the continuation of our research, it is also of great interest to use RJMCMC in identifying which stable-process a given stochastic process belongs to.

REFERENCES

Abbas, H. M. and M. M. Bayoumi (2006). Volterra-system identification using adaptive real-coded genetic algorithm. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 36*(4), 671–684.

Achim, A., E. E. Kuruoglu, and J. Zerubia (2006). SAR image filtering based on the heavy-tailed Rayleigh model. *IEEE Transactions on Image Processing 15*(9), 2686–2693.

Achim, A., P. Tsakalides, and A. Bezerianos (2003). SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Transactions on Geoscience and Remote Sensing 41*(8), 1773–1784.

Aksoy, H. (2000). Use of gamma distribution in hydrological analysis. *Turkish Journal of Engineering and Environmental Sciences 24*(6), 419–428.

Al-Ahmadi, S. and H. Yanikomeroglu (2010). On the approximation of the generalized-K distribution by a gamma distribution for modeling composite fading channels. *IEEE Transactions on Wireless Communications 9*(2), 706–713.

Al-Naffouri, T. Y., A. A. Quadeer, and G. Caire (2011). Impulsive noise estimation and cancellation in DSL using orthogonal clustering. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 2841–2845. IEEE.

Alavi, O., K. Mohammadi, and A. Mostafaeipour (2016). Evaluating the suitability of wind speed probability distribution models: A case of study of east and southeast parts of Iran. *Energy Conversion and Management 119*, 101–108.

Alper, P. (1965). A consideration of the discrete Volterra series. *IEEE Transactions on Automatic Control 10*(3), 322–327.

Alsusa, E. and K. M. Rabie (2013). Dynamic peak-based threshold estimation method for mitigating impulsive noise in power-line communication systems. *IEEE Transactions on Power Delivery 28*(4), 2201–2208.

Andreadou, N. and F.-N. Pavlidou (2010). Modeling the noise on the OFDM power-line communications system. *IEEE Transactions on Power Delivery 25*(1), 150–157.

Andrieu, C., N. De Freitas, A. Doucet, and M. I. Jordan (2003). An introduction to MCMC for machine learning. *Machine Learning 50*(1-2), 5–43.

Aravkin, A., T. Van Leeuwen, and F. Herrmann (2011). Robust full-waveform inversion using the Student's t-distribution. In *SEG Technical Program Expanded Abstracts 2011*, pp. 2669–2673. Society of Exploration Geophysicists.

Artemis Inc. (2017). Synthetic aperture radar solutions. (Accessed: September 2017), http://artemisinc.net/media.php.

Baïle, R., J.-F. Muzy, and P. Poggi (2011). An M-Rice wind speed frequency distribution. *Wind Energy 14*(6), 735–748.

Barker, R. J. and W. A. Link (2013). Bayesian multimodel inference by RJMCMC: A Gibbs sampling approach. *The American Statistician 67*(3), 150–156.

Batselier, K., Z. Chen, and N. Wong (2016). A tensor network Kalman filter with an application in recursive MIMO Volterra system identification. *arXiv preprint arXiv:1610.05434*.

Beck, J. L. (2010). Bayesian system identification based on probability logic. *Structural Control and Health Monitoring 17*(7), 825–847.

Bekleric, S. (2008). Nonlinear prediction via Volterra series and applications to geophysical data. In *Masters Abstracts International*, Volume 47.

Belloni, A. and V. Chernozhukov (2009). On the computational complexity of MCMC-based estimators in large samples. *The Annals of Statistics*, 2011–2055.

Benedetto, S. and E. Biglieri (1999). *Principles of Digital Transmission: With Wireless Applications*. Springer Science & Business Media.

Benmouiza, K. and A. Cheknane (2016). Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models. *Theoretical and Applied Climatology 124*(3-4), 945–958.

Besag, J., P. Green, D. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems. *Statistical Science*, 3–41.

Bhatti, S. A., Q. Shan, I. A. Glover, R. Atkinson, I. E. Portugues, P. J. Moore, and R. Rutherford (2009). Impulsive noise modelling and prediction of its impact on the performance of WLAN receiver. In *Signal Processing Conference, 2009 17th European*, pp. 1680–1684. IEEE.

Bian, Y. and B. Mercer (2015). SAR probability density function estimation using a generalized form of K-distribution. *IEEE Transactions on Aerospace and Electronic Sys-*

tems 51(2), 1136–1146.

Blackard, K. L., T. S. Rappaport, and C. W. Bostian (1993). Measurements and models of radio frequency impulsive noise for indoor wireless communications. *IEEE Journal on Selected Areas in Communications 11*(7), 991–1001.

Bouilloc, T. and G. Favier (2012). Nonlinear channel modeling and identification using baseband Volterra–Parafac models. *Signal Processing 92*(6), 1492–1498.

Bouman, C. and K. Sauer (1993). A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Transactions on Image Processing 2*(3), 296–310.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (2011). *Time Series Analysis: Forecasting and Control*, Volume 734. John Wiley & Sons.

Brooks, S. and P. Giudici (2000). Markov chain Monte Carlo convergence assessment via two-way analysis of variance. *Journal of Computational and Graphical Statistics 9*(2), 266–285.

Brown, L., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao (2005). Statistical analysis of a telephone call center. *Journal of the American Statistical Association 100*(469), 36–50.

Burnham, K. P. and D. R. Anderson (2003). *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer Science & Business Media.

Cadzow, J. A., D. M. Wilkes, R. A. Peters, and X. Li (1993). Image texture synthesis-by-analysis using moving-average models. *IEEE Transactions on Aerospace and Electronic Systems 29*(4), 1110–1122.

Castelloe, J. M. and D. L. Zimmerman (2002). Convergence assessment for reversible jump MCMC samplers. *Department of Statistics and Actuarial Science, University of Iowa, Technical Report 313*.

Chaudhary, N. I., M. A. Z. Raja, M. S. Aslam, and N. Ahmed (2016). Novel generalization of Volterra LMS algorithm to fractional order with application to system identification. *Neural Computing and Applications*, 1–18.

Chen, S., K. Jeong, and W. Härdle (2015). Recurrent support vector regression for a nonlinear ARMA model with applications to forecasting financial returns. *Computational Statistics 30*(3), 821–843.

Chikkula, Y. and J. H. Lee (2000). Robust adaptive predictive control of nonlinear pro-

cesses using nonlinear moving average system models. *Industrial & Engineering Chemistry Research 39*(6), 2010–2023.

Chipman, H., E. I. George, and R. E. Mcculloch (2001). The practical implementation of Bayesian model selection. In *Institute of Mathematical Statistics*, pp. 65–134.

Chou, K. C. and R. B. Corotis (1983). Generalized wind speed probability distribution. *Journal of Engineering Mechanics 109*(1), 14–29.

Choudhary, D. and A. L. Robinson (2014). A new approach to wireless channel modeling using finite mixture models. *International Journal of Digital Information and Wireless Communications (IJDIWC) 4*(2), 169–183.

Claser, R., V. H. Nascimento, and Y. V. Zakharov (2016). A low-complexity RLS-DCD algorithm for Volterra system identification. In *Signal Processing Conference (EU-SIPCO), 2016 24th European*, pp. 6–10. IEEE.

Connor, J., L. Atlas, and D. Martin (1991). Recurrent networks and NARMA modeling. In *NIPS*, pp. 301–308.

Contan, C., B. S. Kirei, et al. (2013). Modified NLMF adaptation of Volterra filters used for nonlinear acoustic echo cancellation. *Signal Processing 93*(5), 1152–1161.

Cordeiro, G. M., E. M. Ortega, and S. Nadarajah (2010). The Kumaraswamy Weibull distribution with application to failure data. *Journal of the Franklin Institute 347*(8), 1399–1429.

Cortes, J. A., L. Diez, F. J. Canete, and J. J. Sanchez-Martinez (2010). Analysis of the indoor broadband power-line noise scenario. *IEEE Transactions on Electromagnetic Compatibility 52*(4), 849–858.

COSMOS (2017). COSMOS, Strong Motion Virtual Data Center (VDC). http://www.strongmotioncenter.org/vdc/scripts/earthquakes.plx.

De Freitas, N., C. Andrieu, P. Højen-Sørensen, M. Niranjan, and A. Gee (2001). Sequential Monte Carlo methods for neural networks. In *Sequential Monte Carlo Methods in Practice*, pp. 359–379. Springer.

Delignon, Y. and W. Pieczynski (2002). Modeling non-Rayleigh speckle distribution in SAR images. *IEEE Transactions on Geoscience and Remote Sensing 40*(6), 1430–1435.

Dellaportas, P. and J. J. Forster (1999). Markov chain Monte Carlo model determination

for hierarchical and graphical log-linear models. *Biometrika 86*(3), 615–633.

Do, M. N. and M. Vetterli (2002). Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing 11*(2), 146–158.

Drobinski, P., C. Coulais, and B. Jourdier (2015). Surface wind-speed statistics modelling: Alternatives to the Weibull distribution and performance evaluation. *Boundary-Layer Meteorology 157*(1), 97–123.

Eğri, E. et al. (2010). Bayesian model selection in ARFIMA models. *Expert Systems with Applications 37*(12), 8359–8364.

Ehlers, R. and S. Brooks (2003). Efficient construction of reversible jump MCMC proposal distributions (with discussion). In *Journal of the Royal Statistical Society, Series B*. Citeseer.

Ehlers, R. S. and S. P. Brooks (2004). Bayesian analysis of order uncertainty in ARIMA models. *Federal University of Parana, Tech. Rep.*

Engle, R. F. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric Reviews 5*(1), 1–50.

Erdem, E. and J. Shi (2011). ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy 88*(4), 1405–1414.

Eun, C. and E. J. Powers (1997). A new Volterra predistorter based on the indirect learning architecture. *IEEE Transactions on Signal Processing 45*(1), 223–227.

Fantacci, R., A. Tani, and D. Tarchi (2010). Impulse noise mitigation techniques for xDSL systems in a real environment. *IEEE Transactions on Consumer Electronics 56*(4), 2106–2114.

Farsad, N., W. Guo, C.-B. Chae, and A. Eckford (2015). Stable distributions as noise models for molecular communication. In *Global Communications Conference (GLOBECOM), 2015 IEEE*, pp. 1–6.

Fernandes, C. A. R., J. C. M. Mota, G. Favier, et al. (2010). MIMO Volterra modeling for nonlinear communication channels. *Learning and Nonlinear Models 2*(8), 71–92.

Flores, J., M. Graff, and H. Rodriguez (2012). Evolutive design of ARMA and ANN models for time series forecasting. *Renewable Energy 44*, 225–230.

Fortuna, L., A. Rizzo, M. Sinatra, and M. Xibilia (2003). Soft analyzers for a sulfur recovery unit. *Control Engineering Practice 11*(12), 1491–1500.

Friedman, N., L. Cai, and X. S. Xie (2006). Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Physical Review Letters 97*(16), 168302.

Ganapathy, S. (2015). Robust speech processing using ARMA spectrogram models. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pp. 5029–5033.

Ganapathy, S., S. Thomas, P. Motlicek, and H. Hermansky (2009). Applications of signal analysis using autoregressive models for amplitude modulation. In *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*, pp. 341–344.

Gao, G. (2010). Statistical modeling of SAR images: A survey. *Sensors 10*(1), 775–795.

Gao, G., K. Ouyang, Y. Luo, S. Liang, and S. Zhou (2017). Scheme of parameter estimation for generalized gamma distribution and its application to ship detection in SAR images. *IEEE Transactions on Geoscience and Remote Sensing 55*(3), 1812–1832.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003, July). *Bayesian Data Analysis, Second Edition (Chapman & Hall/CRC Texts in Statistical Science)* (2 ed.). Chapman and Hall/CRC.

Gelman, A. and D. B. Rubin (1992a). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457–472.

Gelman, A. and D. B. Rubin (1992b). A single series from the Gibbs sampler provides a false sense of security. *Bayesian Statistics 4*, 625–631.

Geman, S. and D. Geman (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*(6), 721–741.

Ghofrani, S., M. Jahed-Motlagh, and A. Ayatollahi (2001). An adaptive speckle suppression filter based on Nakagami distribution. In *EUROCON'2001, Trends in Communications, International Conference on*, Volume 1, pp. 84–87. IEEE.

Glentis, G.-O., K. Georgoulakis, and K. Angellopoulos (2014). Adaptive channel identification in optical fiber communication systems. In *Transparent Optical Networks (ICTON), 2014 16th International Conference on*, pp. 1–4. IEEE.

Goodman, L. A. (1954). Kolmogorov-Smirnov tests for psychological research. *Psychological Bulletin 51*(2), 160.

Goulet, J.-A. and I. F. Smith (2013). Structural identification with systematic errors and unknown uncertainty dependencies. *Computers & Structures 128*, 251–258.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*(4), 711–732.

Green, P. J. and S. Richardson (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics*, 355–375.

Green, P. L. (2015). Bayesian system identification of a nonlinear dynamical system using a novel variant of simulated annealing. *Mechanical Systems and Signal Processing 52*, 133–146.

Green, P. L. and K. Worden (2015). Bayesian and Markov chain Monte Carlo methods for identifying nonlinear systems in the presence of uncertainty. *Phil. Trans. R. Soc. A 373*(2051), 20140405.

Hadjidoukas, P., P. Angelikopoulos, D. Rossinelli, D. Alexeev, C. Papadimitriou, and P. Koumoutsakos (2014). Bayesian uncertainty quantification and propagation for discrete element simulations of granular materials. *Computer Methods in Applied Mechanics and Engineering 282*, 218–238.

Harris, R. I. and N. J. Cook (2014). The parent wind speed distribution: Why Weibull? *Journal of Wind Engineering and Industrial Aerodynamics 131*, 72–87.

Hastie, D. I. and P. J. Green (2012). Model choice using reversible jump Markov chain Monte Carlo. *Statistica Neerlandica 66*(3), 309–338.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

Hershey, J. R. and P. A. Olsen (2007). Approximating the Kullback Leibler divergence between Gaussian mixture models. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, Volume 4, pp. IV–317.

Hubbard, J. and B. West (1995). *Differential Equations: A Dynamical Systems Approach. Part II: Higher Dimensional Systems*. Applications of Mathematics. Springer.

Isaksson, M., D. Wisell, and D. Rönnow (2006). A comparative analysis of behavioral models for RF power amplifiers. *IEEE Transactions on Microwave Theory and Tech-*

niques 54(1), 348–359.

Iskander, D. R. and A. M. Zoubir (1999). Estimation of the parameters of the K-distribution using higher order and fractional moments [radar clutter]. *IEEE Transactions on Aerospace and Electronic Systems 35*(4), 1453–1457.

Ivanov, M. A. (1987). Distribution of the output reponse of Volterra-type narrowband nonlinear-systems. *Telecommunications and Radio Engineering 41*(12), 99–102.

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Ji, W. and K. Chee (2011). Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. *Solar Energy 85*(5), 808–817.

Ji, W. and W.-S. Gan (2012). Identification of a parametric loudspeaker system using an adaptive Volterra filter. *Applied Acoustics 73*(12), 1251–1262.

Kadane, J. B. and N. A. Lazar (2004). Methods and criteria for model selection. *Journal of the American Statistical Association 99*(465), 279–290.

Kalouptsidis, N., G. Mileounis, B. Babadi, and V. Tarokh (2011). Adaptive algorithms for sparse system identification. *Signal Processing 91*(8), 1910–1919.

Kang, S.-Y. (2004). Volterra type integral equation method for the radial Schrödinger equation: Single channel case. *Computers & Mathematics with Applications 48*(10), 1425–1440.

Karakuş, O., E. E. Kuruoğlu, and M. A. Altınkaya (2015). Estimation of the nonlinearity degree for polynomial autoregressive processes with RJMCMC. In *23rd European Signal Processing Conference (EUSIPCO)*, pp. 953–957. IEEE.

Karakuş, O., E. E. Kuruoğlu, and M. A. Altınkaya (2016). Bayesian estimation of polynomial moving average models with unknown degree of nonlinearity. In *24th European Signal Processing Conference (EUSIPCO)*, pp. 1543–1547. IEEE.

Karakuş, O., E. E. Kuruoğlu, and M. A. Altınkaya (2017a). Bayesian Volterra system identification using reversible jump MCMC algorithm. *Signal Processing 141*, 125–136.

Karakuş, O., E. E. Kuruoğlu, and M. A. Altınkaya (2017b). Nonlinear model selection for PARMA processes using RJMCMC. In *25th European Signal Processing Conference (EUSIPCO)*, pp. 2110–2114. IEEE.

Karakuş, O., E. E. Kuruoğlu, and M. A. Altınkaya (2017c). One-day ahead wind speed/power prediction based on polynomial autoregressive model. *IET Renewable Power Generation 11*(11), 1430–1439.

Kareem, A. and T. Wu (2014). Changing dynamic of bridge aerodynamics. *Proceedings of the ICE-Structures and Buildings 168*(2), 94–106.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Khan, Z., T. Balch, and F. Dellaert (2005). MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27*(11), 1805–1819.

Knorr-Held, L. and G. Raßer (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics 56*(1), 13–21.

Koh, T. and E. J. Powers (1985). Second-order Volterra filtering and its application to nonlinear system identification. *IEEE Transactions on Acoustics, Speech and Signal Processing 33*(6), 1445–1455.

Kotoulas, D., P. Koukoulas, and G.-O. Glentis (2011). Subspace based blind identification of LTI FIR MIMO systems and equalization of finite memory SIMO Volterra systems. *Signal Processing 91*(8), 1941–1950.

Kotz, S. and S. Nadarajah (2004). *Multivariate t-distributions and their Applications.* Cambridge University Press.

Kozin, F. (1988). Autoregressive moving average models of earthquake records. *Probabilistic Engineering Mechanics 3*(2), 58–63.

Kullback, S. (1997). *Information Theory and Statistics.* Courier Corporation.

Kuruoglu, E. E. (2001). Density parameter estimation of skewed α-stable distributions. *IEEE Transactions on Signal Processing 49*(10), 2192–2201.

Kuruoglu, E. E., W. J. Fitzgerald, and P. J. Rayner (1998). Near optimal detection of signals in impulsive noise modeled with a symmetric α-stable distribution. *IEEE Communications Letters 2*(10), 282–284.

Kuruoglu, E. E. and J. Zerubia (2004). Modeling SAR images with a generalization of the Rayleigh distribution. *IEEE Transactions on Image Processing 13*(4), 527–533.

Kuruoğlu, E. E. (2002). Nonlinear least l_p-norm filters for nonlinear autoregressive α-stable processes. *Digital Signal Processing 12*(1), 119–142.

Laguna-Sanchez, G. and M. Lopez-Guerrero (2015). On the use of alpha-stable distributions in noise modeling for PLC. *IEEE Transactions on Power Delivery 30*(4), 1863–1870.

Lahaye, P.-J., J.-B. Poline, G. Flandin, S. Dodel, and L. Garnero (2003). Functional connectivity: Studying nonlinear, delayed interactions between BOLD signals. *Neuroimage 20*(2), 962–974.

Lavallée, D., P. Liu, and R. J. Archuleta (2006). Stochastic model of heterogeneity in earthquake slip spatial distributions. *Geophysical Journal International 165*(2), 622–640.

Lavallée, D., H. Miyake, and K. Koketsu (2011). Stochastic model of a subduction-zone earthquake: Sources and ground motions for the 2003 Tokachi-oki, Japan, earthquake. *Bulletin of the Seismological Society of America 101*(4), 1807–1821.

Le Caillec, J.-M. (2011). Spectral inversion of second order Volterra models based on the blind identification of Wiener models. *Signal Processing 91*(11), 2541–2555.

Le Cam, S., A. Belghith, C. Collet, and F. Salzenstein (2009). Wheezing sounds detection using multivariate generalized Gaussian distributions. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 541–544.

Lee, D. (2011). Short-term prediction of wind farm output using the recurrent quadratic Volterra model. In *Power and Energy Society General Meeting, 2011 IEEE*, pp. 1–8.

Lee, T. S. (1996). Image representation using 2D Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*(10), 959–971.

Leon-Garcia, A. (2008). *Probability, Statistics, and Random Processes for Electrical Engineering*. Pearson/Prentice Hall.

Li, H.-C., W. Hong, Y.-R. Wu, and P.-Z. Fan (2011). On the empirical-statistical modeling of SAR images with generalized gamma distribution. *IEEE Journal of Selected Topics in Signal Processing 5*(3), 386–397.

Li, Y. (2016). Generalized-K (GK) distribution: An important general channel model for mobile fading channels. In *Advanced Materials and Structural Engineering: Proceedings of the International Conference on Advanced Materials and Engineering*

Structural Technology (ICAMEST 2015), April 25-26, 2015, Qingdao, China, pp. 473. CRC Press.

Liang, R., Q. Xia, J. Pan, and J. Liu (2017). Testing a linear ARMA model against threshold-ARMA models: A Bayesian approach. *Communications in Statistics-Simulation and Computation 46*(2), 1302–1317.

Liang, Y., G. Chen, S. Naqvi, and J. A. Chambers (2013). Independent vector analysis with multivariate Student's t-distribution source prior for speech separation. *Electronics Letters 49*(16), 1035–1036.

Liew, V. K.-S., T. T.-L. Chong, and K.-P. Lim (2003). The inadequacy of linear autoregressive model for real exchange rates: Empirical evidence from Asian economies. *Applied Economics 35*(12), 1387–1392.

Lin, J., M. Nassar, and B. L. Evans (2013). Impulsive noise mitigation in powerline communications using sparse Bayesian learning. *IEEE Journal on Selected Areas in Communications 31*(7), 1172–1183.

Lopes, P. A., J. M. Pinto, and J. B. Gerald (2013). Dealing with unknown impedance and impulsive noise in the power-line communications channel. *IEEE Transactions on Power Delivery 28*(1), 58–66.

Loussifi, H., K. Nouri, and N. B. Braiek (2016). A new efficient hybrid intelligent method for nonlinear dynamical systems identification: The wavelet kernel fuzzy neural network. *Communications in Nonlinear Science and Numerical Simulation 32*, 10–30.

Lu, L., H. Zhao, and B. Chen (2016). Improved-variable-forgetting-factor recursive algorithm based on the logarithmic cost for Volterra system identification. *IEEE Transactions on Circuits and Systems II: Express Briefs 63*(6), 588–592.

Lunn, D. J., N. Best, and J. C. Whittaker (2009). Generic reversible jump MCMC using graphical models. *Statistics and Computing 19*(4), 395–408.

Lydia, M., S. Kumar, A. Selvakumar, and G. Kumar (2016). Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy Conversion and Management 112*, 115–124.

Ma, X. and C. L. Nikias (1995). Parameter estimation and blind channel identification in impulsive signal environments. *IEEE Transactions on Signal Processing 43*(12), 2884–2897.

Martinez Lara, C., M. Martin Perez, I. Martin Garcia, R. Blanco Hernández,

B. Sánchez Sánchez, and J. Sevillano Sánchez (2012). Radiological findings invasive lobular carcinoma. In *European Congress of Radiology (ECR)*, pp. C–1062.

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association 46*(253), 68–78.

Mckenzie, E. (1982). Product autoregression: A time-series characterization of the gamma distribution. *Journal of Applied Probability 19*(2), 463–468.

Medhi, J. (2002). *Stochastic Models in Queueing Theory*. Academic Press.

Merabti, H. and D. Massicotte (2014). Nonlinear adaptive channel equalization using genetic algorithms. In *New Circuits and Systems Conference (NEWCAS), 2014 IEEE 12th International*, pp. 209–212. IEEE.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys. 21*, 1087.

Meyer-Gohde, A. et al. (2014). Generalized exogenous processes in DSGE: A Bayesian approach. Technical report, Humboldt University.

Mhatli, S., B. Nsiri, M. A. Jarajreh, M. Channoufi, and R. Attia (2015). Adaptive Volterra equalizer for optical OFDM modem. In *Photonics Prague 2014*, pp. 94500R–94500R. International Society for Optics and Photonics.

Mileounis, G. and N. Kalouptsidis (2009). Blind identification of second order Volterra systems with complex random inputs using higher order cumulants. *IEEE Transactions on Signal Processing 57*(10), 4129–4135.

Mileounis, G., P. Koukoulas, and N. Kalouptsidis (2009). Input–output identification of nonlinear channels using PSK, QAM and OFDM inputs. *Signal Processing 89*(7), 1359–1369.

MRI Scan Images Info. (2017). MRI scan images info., MRI image of brain with gadolinium contrast showing enhancing mass in the right. (Accessed: 2017-09-01), http://mri-scan-img.info/mri-image-of-brain-with-gadolinium-contrast-showing-enhancing-mass-in-the-right/mri-image-of-brain-with-gadolinium-contrast-showing-enhancing-mass-in-the-right-jpg/.

Mudholkar, G. S., D. K. Srivastava, and G. D. Kollia (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association 91*(436), 1575–1583.

Nakagami, M. (1960). The m-distribution-a general formula of intensity distribution of rapid fading. *Statistical Method of Radio Propagation*, 3–34.

Nguyen, T. M. and Q. J. Wu (2012). Robust Student's-t mixture model with spatial constraints and its application in medical image segmentation. *IEEE Transactions on Medical Imaging 31*(1), 103–116.

Nikias, C. L. and M. Shao (1995). *Signal Processing with Alpha-stable Distributions and Applications*. New York, NY, USA: Wiley-Interscience.

Ninness, B. and S. Henriksen (2010). Bayesian system identification via Markov chain Monte Carlo techniques. *Automatica 46*(1), 40–51.

Nolan, J. (2010). Bibliography on stable distributions, processes and related topics. Technical report.

Novey, M., T. Adali, and A. Roy (2010). A complex generalized Gaussian distribution—Characterization, generation, and estimation. *IEEE Transactions on Signal Processing 58*(3), 1427–1433.

Nowak, R. D. (2002). Nonlinear system identification. *Circuits, Systems, and Signal Processing 21*(1), 109–122.

Oedekoven, C. S., R. King, S. T. Buckland, M. L. Mackenzie, K. Evans, and L. Burger (2016). Using hierarchical centering to facilitate a reversible jump MCMC algorithm for random effects models. *Computational Statistics & Data Analysis 98*, 79–90.

Ohtsuka, Y., T. Oga, and K. Kakamu (2010). Forecasting electricity demand in Japan: A Bayesian spatial autoregressive ARMA approach. *Computational Statistics & Data Analysis 54*(11), 2721–2735.

Palomares-Salas, J. C., J. J. G. De la Rosa, J. G. Ramiro, J. Melgar, A. Agüera, and A. Moreno (2009). Comparison of models for wind speed forecasting. In *Proc. International Conference on Computational Science*.

Pappas, S., L. Ekonomou, P. Karampelas, D. Karamousantas, S. Katsikas, G. Chatzarakis, and P. Skafidas (2010). Electricity demand load forecasting of the Hellenic power system using an ARMA model. *Electric Power Systems Research 80*(3), 256–264.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review 47*(2), 527–556.

Patzold, M., U. Killat, F. Laue, and Y. Li (1998). On the statistical properties of determin-

istic simulation models for mobile fading channels. *IEEE Transactions on Vehicular Technology 47*(1), 254–269.

Peng, S.-H., C.-W. Li, and Y.-W. Liu (2015). Joint estimation of vocal tract and nasal tract area functions from speech waveforms via auto-regression moving-average modeling and a pole assignment method. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pp. 4644–4648.

Peskun, P. H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika 60*(3), 607–612.

Potter, S. (1999). Nonlinear time series modelling: An introduction. *Journal of Economic Surveys 13*(5), 505–528.

Press, W. H. (2007). *Numerical Recipes (3rd edition): The Art of Scientific Computing*. Cambridge University Press.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163.

Rannala, B. and Z. Yang (2013). Improved reversible jump algorithms for Bayesian species delimitation. *Genetics 194*(1), 245–253.

Reig, J. and L. Rubio (2013). Estimation of the composite fast fading and shadowing distribution using the log-moments in wireless communications. *IEEE Transactions on Wireless Communications 12*(8), 3672–3681.

Richardson, S. and P. J. Green (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59*(4), 731–792.

Robinson, P. M. (1977). The estimation of a nonlinear moving average model. *Stochastic Processes and their Applications 5*(1), 81–90.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics*, 163–168.

Rodrigo-Peñarrocha, V. M., J. Reig, L. Rubio, H. Fernández, and S. Loredo (2016). Analysis of small-scale fading distributions in vehicle-to-vehicle communications. *Mobile Information Systems 2016*.

Roeth, O., D. Zaum, and C. Brenner (2017). Extracting lane geometry and topology information from vehicle fleet trajectories in complex urban scenarios using a reversible

jump MCMC method. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences 4*.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association 90*(430), 558–566.

Sadreazami, H., M. O. Ahmad, and M. S. Swamy (2014). A study of multiplicative watermark detection in the contourlet domain using alpha-stable distributions. *IEEE Transactions on Image Processing 23*(10), 4348–4360.

Salas-Gonzalez, D., E. E. Kuruoglu, and D. P. Ruiz (2010). Modelling with mixture of symmetric stable distributions using Gibbs sampling. *Signal Processing 90*(3), 774–783.

Samorodnitsky, G. and M. S. Taqqu (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Volume 1. CRC press.

Sawyer, S. (2006). The Metropolitan-Hastings algorithm and extensions. *Washington University*.

Schetzen, M. (1980). *The Volterra and Wiener theories of nonlinear systems*. John Wiley & Sons.

Sekiguchi, M. and T. Nakajima (2008). A K-distribution-based radiation code and its computational optimization for an atmospheric general circulation model. *Journal of Quantitative Spectroscopy and Radiative Transfer 109*(17), 2779–2793.

Shankar, P. M. (2001). Ultrasonic tissue characterization using a generalized Nakagami model. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 48*(6), 1716–1720.

Shankar, P. M. (2004). Error rates in generalized shadowed fading channels. *Wireless Personal Communications 28*(3), 233–238.

Shankar, P. M., V. Dumane, J. M. Reid, V. Genis, F. Forsberg, C. W. Piccoli, and B. B. Goldberg (2001). Classification of ultrasonic B-mode images of breast masses using Nakagami distribution. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 48*(2), 569–580.

Shi, K. and P. Shi (2011). Adaptive sparse Volterra system identification with l_0-norm penalty. *Signal Processing 91*(10), 2432–2436.

Shoaib, M., S. Werner, and J. A. Apolinário (2010). Multichannel fast QR-decomposition

algorithms: Weight extraction method and its applications. *IEEE Transactions on Signal Processing 58*(1), 175–188.

Simoen, E., C. Papadimitriou, and G. Lombaert (2013). On prediction error correlation in Bayesian model updating. *Journal of Sound and Vibration 332*(18), 4136–4152.

Simoncelli, E. P. (1997). Statistical models for images: Compression, restoration and synthesis. In *Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on*, Volume 1, pp. 673–678. IEEE Computer Society.

Sisson, S. A. and Y. Fan (2007). A distance-based diagnostic for trans-dimensional Markov chains. *Statistics and Computing 17*(4), 357–367.

Smith, K. C. (2007). *Bayesian Methods for Visual Multi-object Tracking with Applications to Human Activity Recognition*. Ph. D. thesis, Ecole Polytechnique Federale de Lausanne, Lausanne , Switzerland.

So, M. K., C. W. Chen, and M.-T. Chen (2005). A Bayesian threshold nonlinearity test for financial time series. *Journal of Forecasting 24*(1), 61–75.

Stark, H. and J. Woods (2002). *Probability and Random Processes with Applications to Signal Processing*. Prentice Hall PTR.

Stephens, M. A. (1970). Use of the Kolmogorov-Smirnov, Cramér-Von Mises and related statistics without extensive tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 115–122.

Swami, A. and B. Sadler (1998). Parameter estimation for linear alpha-stable processes. *IEEE Signal Processing Letters 5*(2), 48–50.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics 22*, 1701–1762.

Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability 8*, 1–9.

Tong, H. (2002). Nonlinear time series analysis since 1990: Some personal reflections. *Acta Mathematicae Applicatae Sinica (English Series) 18*(2), 177–184.

Tong, L., J. Yang, and R. S. Cooper (2010). Efficient calculation of p-value and power for quadratic form statistics in multilocus association testing. *Annals of Human Genetics 74*(3), 275–285.

Tran, T. H., D. D. Do, and T. H. Huynh (2013). PLC impulsive noise in industrial zone: measurement and characterization. *International Journal of Computer and Electrical Engineering 5*(1), 48.

Troughton, P. T. and S. J. Godsill (1998). A reversible jump sampler for autoregressive time series. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, Volume 4, pp. 2257–2260.

Troughton, P. T. and S. J. Godsill (2001). MCMC methods for restoration of nonlinearly distorted autoregressive signals. *Signal Processing 81*(1), 83–97.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association 84*(405), 231–240.

Tsihrintzis, G. A. and C. L. Nikias (1996). Fast estimation of the parameters of alpha-stable impulsive interference. *IEEE Transactions on Signal Processing 44*(6), 1492–1503.

Tsihrintzis, G. A. and C. L. Nikias (1997). Data-adaptive algorithms for signal detection in sub-Gaussian impulsive interference. *IEEE Transactions on Signal Processing 45*(7), 1873–1878.

Tsihrintzis, G. A., P. Tsakalides, and C. L. Nikias (1998). Spectral methods for stationary harmonizable alpha-stable processes. In *Signal Processing Conference (EUSIPCO 1998), 9th European*, pp. 1–4. IEEE.

Tuli, A., N. Kumar, and K. Sharma (2014). Image transmission using M-QAM OFDM system over composite fading channel. *9*, 69–77.

Tzagkarakis, G. and P. Tsakalides (2010). Greedy sparse reconstruction of non-negative signals using symmetric alpha-stable distributions. In *Signal Processing Conference, 2010 18th European*, pp. 417–421. IEEE.

Van Der Meulen, F., M. Schauer, and H. Van Zanten (2014). Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Computational Statistics & Data Analysis 71*, 615–632.

Ver Hoef, J. M. and P. L. Boveng (2007). Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology 88*(11), 2766–2772.

Verdoolaege, G. and P. Scheunders (2011). Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination. *International Journal of Computer Vision 95*(3), 265–286.

Vermaak, J., C. Andrieu, A. Doucet, and S. J. Godsill (2004). Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes. *Journal of Time Series Analysis 25*(6), 785–809.

Viallefont, V., S. Richardson, and P. J. Green (2002). Bayesian analysis of Poisson mixtures. *Journal of Nonparametric Statistics 14*(1-2), 181–202.

Vomisescu, R. (2003). Probabilites and Lebesgue measure. *General Mathematics 11*(1-2), 87–91.

Walck, C. (1996). Hand-book on statistical distributions for experimentalists. Technical report.

Wang, J., E. E. Kuruoglu, and T. Zhou (2011). Alpha-stable channel capacity. *IEEE Communications Letters 15*(10), 1107–1109.

Wang, J., W. W. Tsang, and G. Marsaglia (2003). Evaluating Kolmogorov's distribution. *Journal of Statistical Software 8*(18).

Wilcox, R. (2005). Kolmogorov–Smirnov test. *Encyclopedia of Biostatistics 4*.

Wolf, R. (1997). *Proof, Logic, and Conjecture: The Mathematician's Toolbox*. W. H. Freeman.

Wolff, R. W. (1982). Poisson arrivals see time averages. *Operations Research 30*(2), 223–231.

Woods, J. W. (2013). *Subband Image Coding*, Volume 115. Springer Science & Business Media.

Xia, Q., J. Pan, Z. Zhang, and J. Liu (2010). A Bayesian nonlinearity test for threshold moving average models. *Journal of Time Series Analysis 31*(5), 329–336.

Yacoub, M. D., J. V. Bautista, and L. G. de Rezende Guedes (1999). On higher order statistics of the Nakagami-m distribution. *IEEE Transactions on Vehicular Technology 48*(3), 790–794.

Yue, B. and Z. Peng (2015). A validation study of α-stable distribution characteristic for seismic data. *Signal Processing 106*, 1–9.

Yueh, S., J. A. Kong, J. Jao, R. Shin, and L. Novak (1989). K-distribution and polarimetric terrain radar clutter. *Journal of Electromagnetic Waves and Applications 3*(8), 747–768.

Zaffaroni, P. and B. d'Italia (2003). Gaussian inference on certain long-range dependent volatility models. *Journal of Econometrics 115*(2), 199–258.

Zhang, J. and H. Wang (2008). Minimum entropy control of nonlinear ARMA systems over a communication network. *Neural Computing and Applications 17*(4), 385–390.

Zhang, Z., K. Lai, Z. Lu, and X. Tong (2013). Bayesian inference and application of robust growth curve models using Student's t distribution. *Structural Equation Modeling: A Multidisciplinary Journal 20*(1), 47–78.

Zhong, J. et al. (2003). Segmenting foreground objects from a dynamic textured background via a robust Kalman filter. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 44–50.

Zolotarev, V. M. (1986). *One-dimensional Stable Distributions*, Volume 65. American Mathematical Soc.

APPENDIX A

IMPULSIVE DISTRIBUTION FAMILIES

A.1. Symmetric α-Stable Distribution Family

There is no closed form expression for probability density function (pdf) of SαS distributions except for the special cases of Cauchy and Gaussian. However, their characteristic function, $\varphi(x)$, can be expressed explicitly as

$$\varphi(x) = \exp(j\delta x - \gamma|x|^{\alpha}) \tag{A.1}$$

where $0 < \alpha \leq 2$ is the characteristic exponent, *a.k.a. shape parameter*, which controls the impulsiveness of the distribution. Special cases Cauchy and Gaussian distributions occur when $\alpha = 1$ and $\alpha = 2$, respectively. $-\infty < \delta < \infty$ represents the *location parameter*. $\gamma > 0$ provides a measure of the dispersion which is the *scale parameter* expressing the spread of the distribution around δ.

Absolute FLOM definition of SαS distributions is

$$E(|\mathbf{x}|^p) = C_\alpha(p, \alpha)\gamma^{p/\alpha} \tag{A.2}$$

where

$$C_\alpha(p, \alpha) = \frac{\Gamma\left(\dfrac{p+1}{2}\right)\Gamma\left(\dfrac{-p}{\alpha}\right)}{\alpha\,\sqrt{\pi}\Gamma\left(\dfrac{-p}{2}\right)} 2^{p+1}. \tag{A.3}$$

In Figure A.1, the standard SαS density functions and their tail characteristics are shown in terms of different shape parameter α. Note that the plot for $\alpha = 2$ refers to the standard normal density, and this figure gives us the opportunity to compare Gaussian and

non-Gaussian stable densities. Stable densities are more peaked than normal for the values of samples around location parameter δ. Due to having algebraic tails unlike the Gaussian density tail which is exponential, stable densities have heavier tails than the Gaussian. Moreover, the smaller α a stable density has, the heavier are the tails (Zolotarev, 1986; Samorodnitsky and Taqqu, 1994; Nikias and Shao, 1995).

A.2. Generalized Gaussian Distribution Family

The univariate GG pdf can be defined as

$$ f(x) = \frac{\alpha}{2\gamma\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x - \delta|}{\gamma}\right)^{\alpha}\right) \tag{A.4} $$

where $\Gamma(\cdot)$ refers to the gamma function, $\alpha > 0$ is the shape parameter, $-\infty < \delta < \infty$ represents the location parameter and $\gamma > 0$ is the scale parameter. GG family has well-known members such as Laplace, Gauss and uniform distributions for α values of 1, 2 and ∞, respectively.

Absolute FLOM definition of GG distributions is

$$ E(|\mathbf{x}|^p) = C_{\text{GG}}(p, \alpha)\gamma^p, \tag{A.5} $$

where

$$ C_{\text{GG}}(p, \alpha) = \frac{\Gamma\left(\dfrac{p+1}{\alpha}\right)}{\Gamma(1/\alpha)}. \tag{A.6} $$

In Figure A.2, the standard GG density functions and tail characteristics for different shape parameter α are depicted. Similarly, note that the density with $\alpha = 2$ corresponds to a Gaussian. For samples around the location parameter δ, GG densities with higher α values are more peaky than the ones with small α, whereas GG densities with

smaller α have heavier tails (Woods, 2013).

A.3.   Student's t Distribution Family

The univariate symmetric Student's t distribution family is an impulsive distribution family with parameters, $\alpha > 0$ which is the number of degrees of freedom, *a.k.a shape parameter*, the location parameter $-\infty < \delta < \infty$ and the scale parameter $\gamma > 0$. Its pdf can be defined as

$$f(x) = \frac{\Gamma\left(\dfrac{\alpha + 1}{2}\right)}{\Gamma(\alpha/2)\gamma \sqrt{\pi\alpha}} \left(1 + \frac{1}{\alpha}\left(\frac{x-\delta}{\gamma}\right)^2\right)^{-((\alpha+1)/2)}. \tag{A.7}$$

Special members of the symmetric Student's t distribution family are Cauchy and Gauss which are obtained for shape parameter values of $\alpha = 1$ and $\alpha = \infty$, respectively.

Absolute FLOM definition of Student's t distributions is

$$E(|\mathbf{x}|^p) = C_t(p, \alpha)\gamma^p, \tag{A.8}$$

where

$$C_t(p, \alpha) = \frac{\Gamma\left(\dfrac{p+1}{2}\right)\Gamma\left(\dfrac{\alpha-p}{2}\right)}{\sqrt{\pi}\Gamma\left(\dfrac{\alpha}{2}\right)}\alpha^{p/2}. \tag{A.9}$$

In Figure A.3, the standard Student's t density functions and tail characteristics are given in order to show the differences for varying shape parameters. General characteristics both for densities and tails are similar to GG densities. However t distributions are much more heavy tailed than GG. Their members with larger α have more peaky densities around the location parameter, but ones with smaller α have heavier tails (Kotz and Nadarajah, 2004).
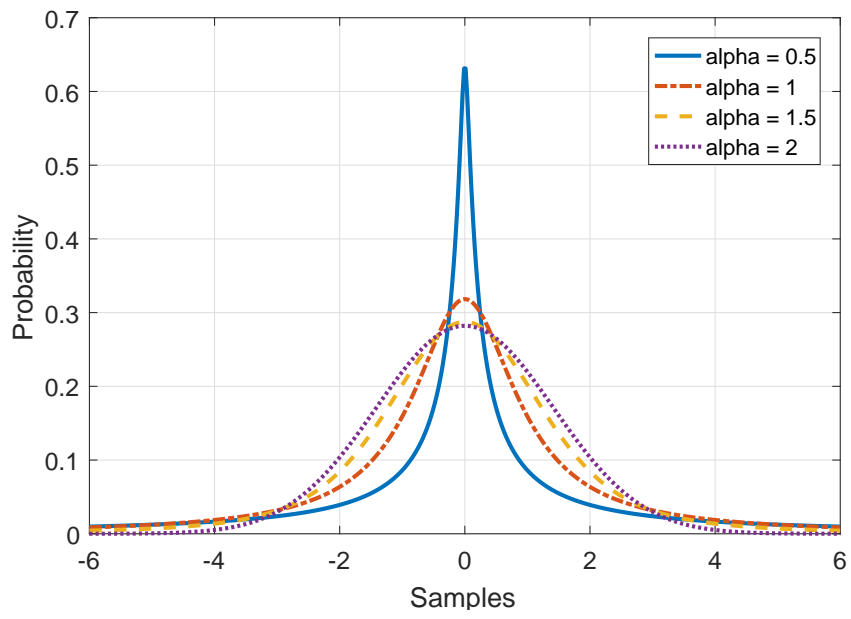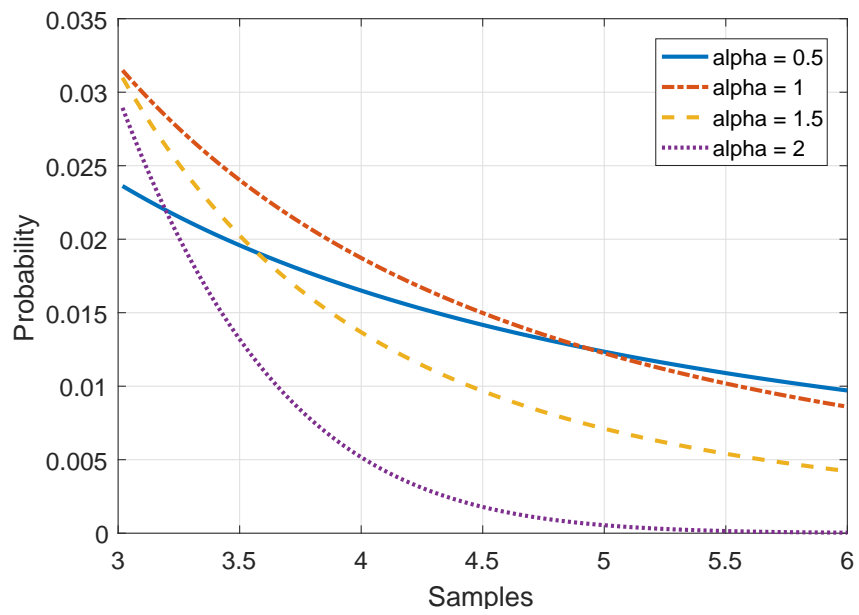
Figure A.1. Density function of SαS for different values of the shape parameter α. (a): the overall densities (b): the tails of the densities. For all the distribution families the scale parameter $\gamma = 1$ and $\delta = 0$.
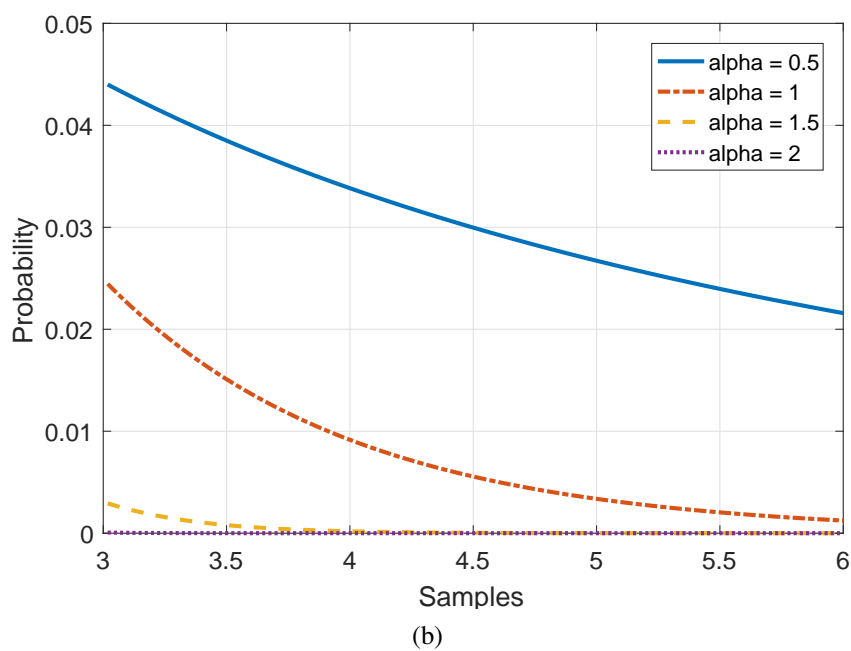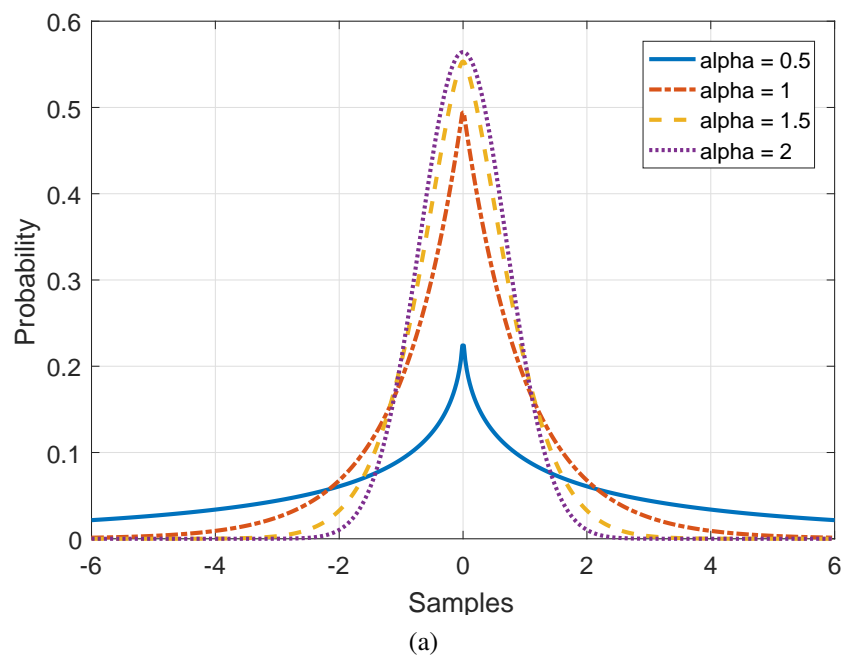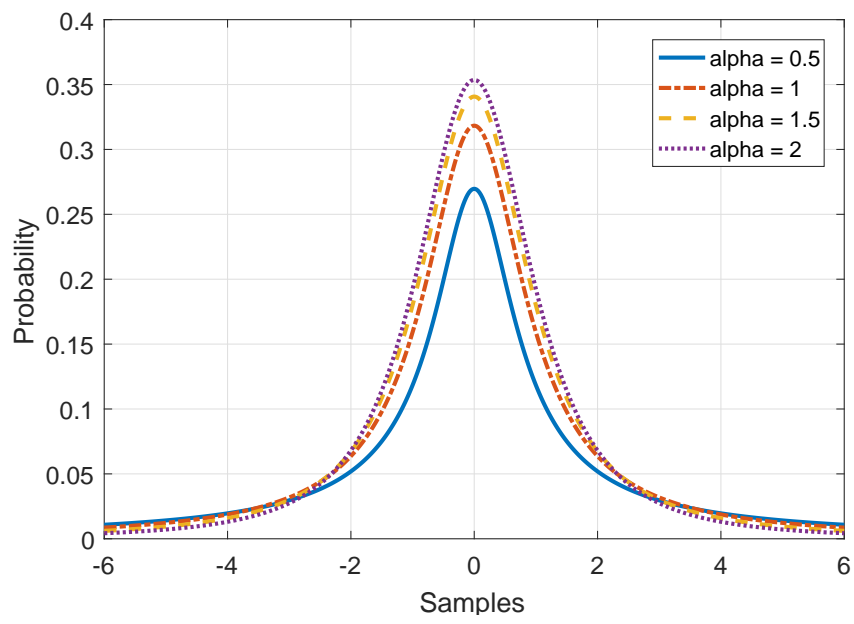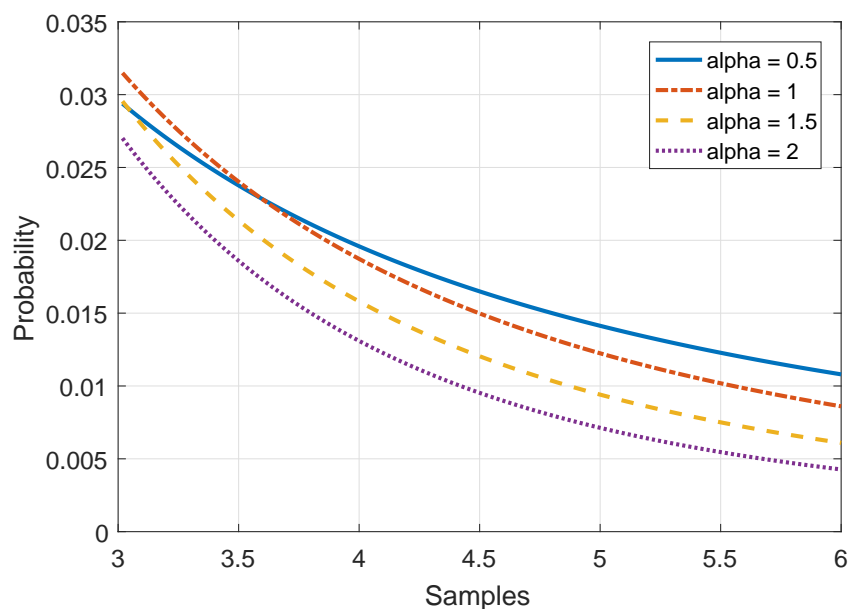
Figure A.2. Density function of GG for different values of the shape parameter α. (a): the overall densities (b): the tails of the densities. For all the distribution families the scale parameter $\gamma = 1$ and $\delta = 0$.

(a)

(b)

Figure A.3. Density function of Student's *t* for different values of the shape parameter
α. (a): the overall densities (b): the tails of the densities. For all the
distribution families the scale parameter $\gamma = 1$ and $\delta = 0$.

APPENDIX B

ENVELOPE DISTRIBUTION FAMILIES

B.1. Nakagami Distribution Family

Probability density function of the univariate Nakagami distribution can be defined as (Nakagami, 1960; Yacoub et al., 1999)

$$f(x) = \frac{2\alpha^{\alpha}}{\Gamma(\alpha)\gamma^{\alpha}} x^{2\alpha-1} \exp\left(-\frac{\alpha}{\gamma}x^2\right) \tag{B.1}$$

where $\Gamma(\cdot)$ is the gamma function, $\alpha \geq 0.5$ refers to the shape parameter and $\gamma > 0$ represents the scale parameter of the Nakagami distribution family. One-sided Gaussian distribution and Rayleigh distribution are the special members of Nakagami distribution family for shape parameters of 0.5 and 1, respectively.

The pth order moment definition of the Nakagami distribution is (Yacoub et al., 1999)

$$E(\mathbf{x}^p) = C_1(p,\alpha)\left(\frac{\gamma}{\alpha}\right)^{p/2}, \tag{B.2}$$

and

$$C_1(p,\alpha) = \frac{\Gamma\left(\alpha + \dfrac{p}{2}\right)}{\Gamma(\alpha)}. \tag{B.3}$$

B.2. $\mathcal{K}$-Distribution Family

$\mathcal{K}$-distribution is a distribution family which arises by compounding two gamma

distributions. Assume a random variable x is Gamma distributed with mean m and shape L, the mean m of which is also a Gamma distributed random variable. Thus, x is $\mathcal{K}$-distributed and $\mathcal{K}$-distribution is a compound process. Additionally, $\mathcal{K}$-distribution is also a product distribution. Particularly, it is the distribution of a random variable which is the product of two independent Gamma random variables. The pdf of a univariate $\mathcal{K}$-distribution can be defined as (Iskander and Zoubir, 1999)

$$f(x) = \frac{2}{\gamma\Gamma(\alpha + 1)} \left(\frac{x}{2\gamma}\right)^{\alpha+1} K_\alpha\left(\frac{x}{\gamma}\right) \tag{B.4}$$

where K_α refers to the modified Bessel function of order α and α and γ represent the shape and the scale parameters, respectively.

The pth order moment definition of $\mathcal{K}$-distribution is (Iskander and Zoubir, 1999)

$$E(\mathbf{x}^p) = C_2(p, \alpha)\gamma^p, \tag{B.5}$$

and

$$C_2(p, \alpha) = \frac{2^p\Gamma\left(\frac{p}{2} + 1\right)\Gamma\left(\alpha + 1 + \frac{p}{2}\right)}{\Gamma(\alpha + 1)}. \tag{B.6}$$

B.3. Weibull Distribution Family

For a univariate Weibull distribution, the pdf can be defined as (Walck, 1996)

$$f(x) = \frac{\alpha}{\gamma}\left(\frac{x}{\gamma}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\gamma}\right)^\alpha\right) \tag{B.7}$$

where α is the shape and γ is the scale parameter. Weibull distribution family has special members for $\alpha = 1$ and $\alpha = 2$ which are the well-known exponential and Rayleigh distributions, respectively.

The pth order moment definition of the Weibull distribution is (Walck, 1996)

$$E(\mathbf{x}^p) = C_3(p, \alpha)\gamma^p \tag{B.8}$$

and

$$C_3(p, \alpha) = \Gamma\left(1 + \frac{p}{\alpha}\right). \tag{B.9}$$

B.4. Gamma Distribution Family

Univariate gamma distribution pdf can be expressed as (Walck, 1996)

$$f(x) = \frac{x^{\alpha-1}}{\gamma^\alpha \Gamma(\alpha)} \exp\left(-\frac{x}{\gamma}\right) \tag{B.10}$$

where α refers to the shape parameter and γ is the scale parameter. The well-known exponential and chi-squared distributions are special members of the gamma distribution family.

The pth order moment definition of the Gamma distribution is (Walck, 1996)

$$E(\mathbf{x}^p) = C_4(p, \alpha)\gamma^p \tag{B.11}$$

and

$$C_4(p, \alpha) = \frac{\Gamma(\alpha + p)}{\Gamma(\alpha)}. \tag{B.12}$$

B.5. Generalized Rayleigh Distribution.

The classical Rayleigh distribution represents the distribution of a random variable which is the magnitude of a 2 dimensional vector, components of which are zero mean, equal variance Gaussian random variables. Generalized Rayleigh (or heavy tailed) Rayleigh distribution represents the distribution of a random variable which is again a 2-dimensional vector. However this time, the components of the vector are zero location equal dispersion SαS random variables. Generalized Rayleigh distribution has a pdf expression in integral form as (Kuruoglu and Zerubia, 2004)

$$f(x) = x \int_0^\infty s \exp\left(-\gamma s^\alpha\right) J_0(sx) ds \tag{B.13}$$

where $\gamma > 0$ is the scale parameter, $0 < \alpha \leq 2$ is the shape parameter and $J_0(\cdot)$ refers to the zeroth order Bessel function of the first kind. Rayleigh distribution is a special member of generalized Rayleigh distribution with $\alpha = 2$.

The pth order moment definition of the generalized Rayleigh distribution is (Kuruoglu and Zerubia, 2004)

$$E(\mathbf{x}^p) = C_5(p, \alpha)\gamma^{p/\alpha} \tag{B.14}$$

and

$$C_5(p, \alpha) = \frac{2^{p+1}\Gamma\left(\frac{p}{2} + 1\right)\Gamma\left(\frac{-p}{\alpha}\right)}{\alpha\Gamma\left(\frac{-p}{2}\right)}. \tag{B.15}$$

In Figures through B.1 and B.5, densities for Nakagami, $\mathcal{K}$, Weibull, Gamma and generalized Rayleigh distributions are shown, respectively. For Nakagami densities increasing the shape parameter α makes the densities peaky and heavy tailed relative to the ones with smaller α. Weibull distribution also follows very similar characteristics to Nakagami distribution. For $\mathcal{K}$ and Gamma distributions, decreasing the shape parameter makes both $\mathcal{K}$ and Gamma more peaky and heavy tailed than the ones with larger α. For generalized Rayleigh distributions, tail probabilities increase as order of α values

increases, whereas peak values are very similar for all α values. Generalized Rayleigh, $\mathcal{K}$ and Gamma distributions are different from Nakagami and Weibull due to having heavier tails. Generally, Nakagami and Weibull distributions are not heavy tailed envelope distributions and densities diminish towards zero faster than the other distributions.
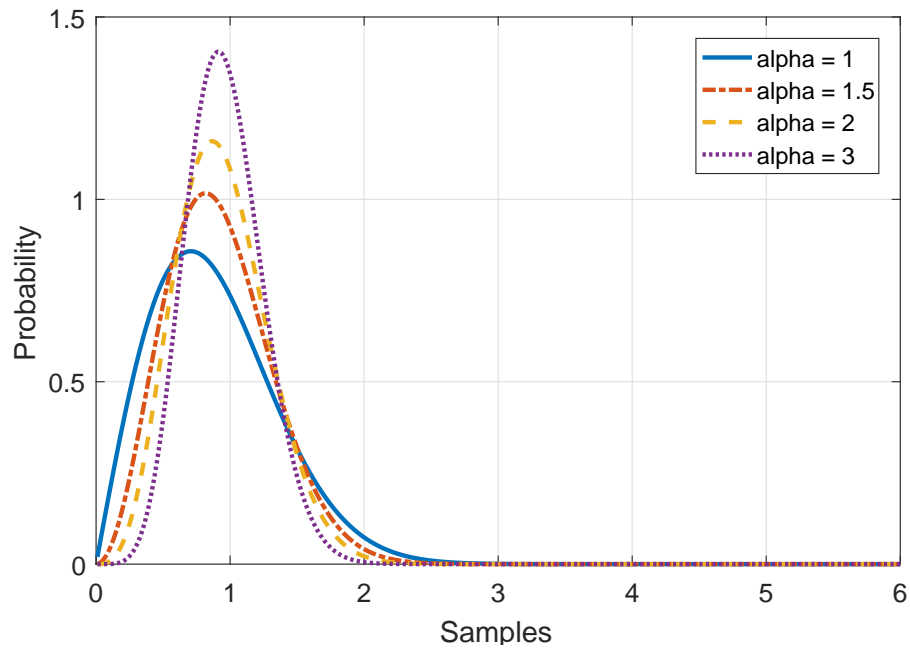

Figure B.1. Density function of Nakagami distribution for different values of the shape parameter α. For all the distribution families the scale parameter $\gamma = 1$ and $\delta = 0$.
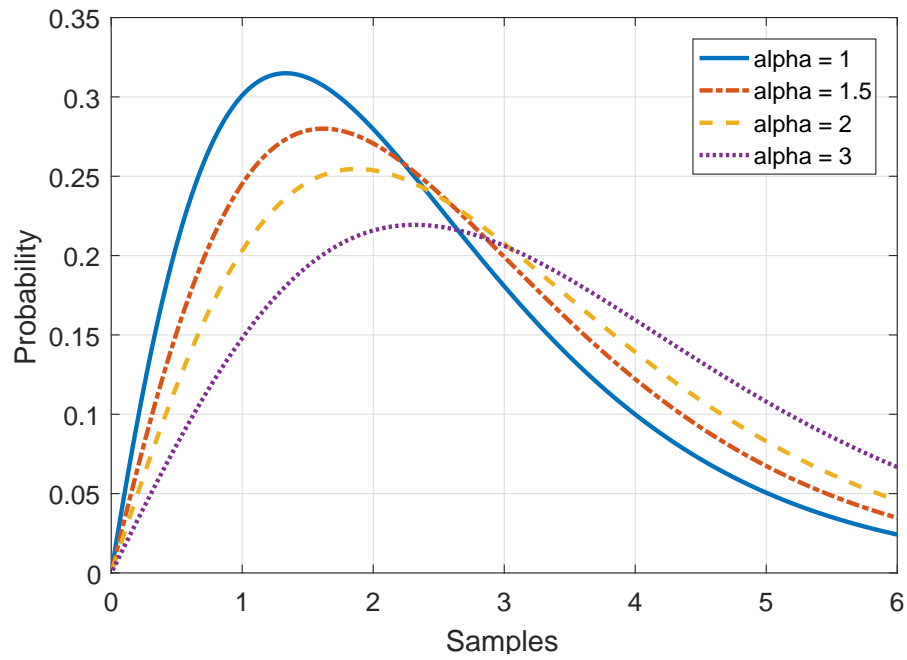
Figure B.2. Density function of $\mathcal{K}$-distribution for different values of the shape parameter α. For all the distribution families the scale parameter $\gamma = 1$ and $\delta = 0$.
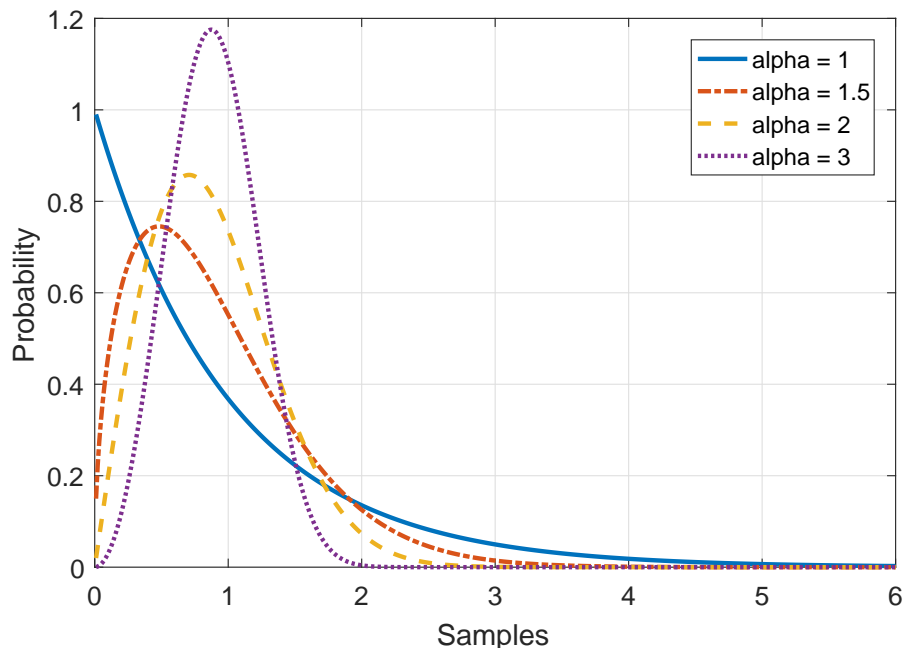

Figure B.3. Density function of Weibull distribution for different values of the shape parameter α. For all the distribution families the scale parameter $\gamma = 1$ and $\delta = 0$.
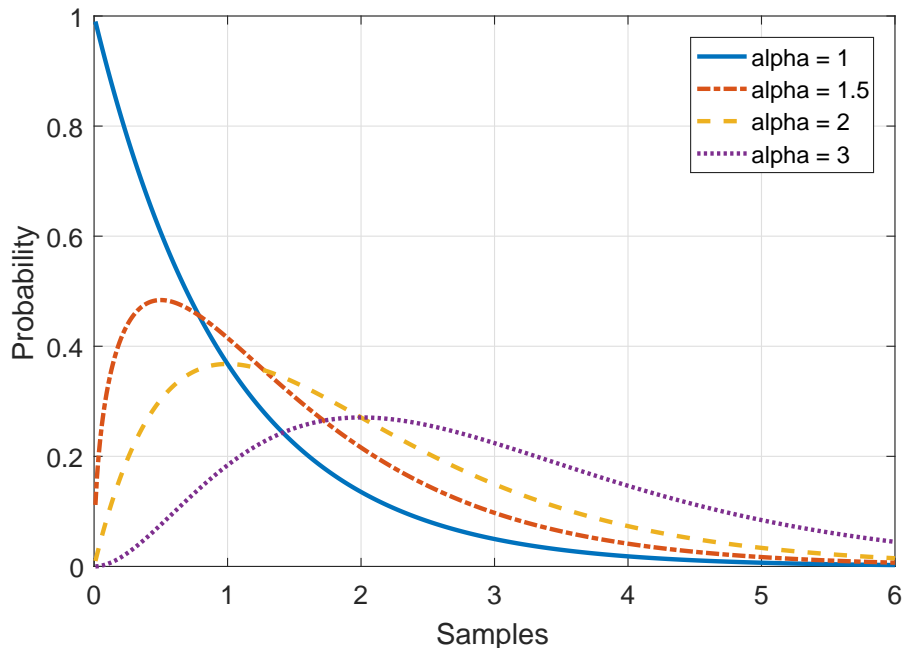
Figure B.4. Density function of Gamma distribution for different values of the shape parameter α. For all the distribution families the scale parameter $\gamma = 1$ and $\delta = 0$.
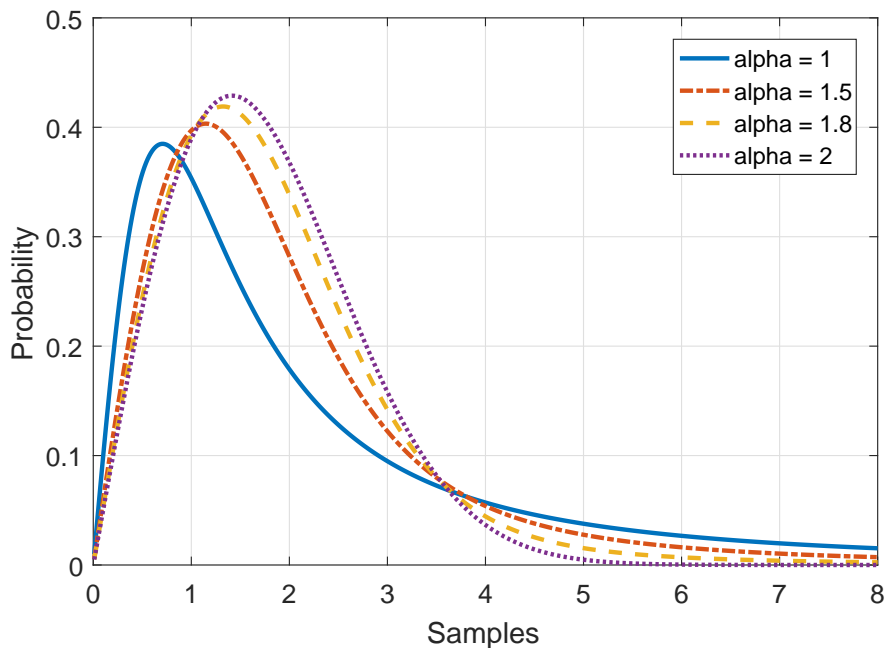

Figure B.5. Density function of generalized Rayleigh distribution for different values of the shape parameter α. For all the distribution families the scale parameter $\gamma = 1$ and $\delta = 0$.

APPENDIX C

STATISTICAL SIGNIFICANCE TESTS

C.1. Kullback-Leibler Divergence

In order to measure the difference between two probability distributions, in probability and statistics, a well-known approach named *Kullback-Leibler divergence* or shortly *KL divergence* has been commonly used. KL divergence provides a non-symmetric measure about how different two probability distributions, e.g. p and g are. It is also known as relative entropy between p and g.

KL divergence between two continuous probability distributions $p(x)$ and $g(x)$ can be defined as

$$D_{KL}(p\|q) = E\left[\log(p(x)) - \log(g(x))\right] \tag{C.1}$$

where $\log(\cdot)$ refers to the natural logarithm and $E[\cdot]$ is the expectation. The most common way to represent $D_{KL}(p\|q)$ is (Kullback, 1997; Hershey and Olsen, 2007)

$$D_{KL}(p\|q) = \int_x p(x) \log\left(\frac{p(x)}{g(x)}\right). \tag{C.2}$$

The notation $D_{KL}(p\|q)$ denotes "information lost where g is used to approximate p" (Burnham and Anderson, 2003).

The KL divergence can also be used to measure the distance between discrete distributions, such as Poisson, negative binomial, or in cases when comparing two discrete populations. Discrete KL divergence can be defined as (Burnham and Anderson, 2003)

$$D_{KL}(p\|q) = \sum_{i=1}^{k} p(x_i) \log\left(\frac{p(x_i)}{g(x_i)}\right). \tag{C.3}$$

KL divergence satisfies three properties, which are (Hershey and Olsen, 2007)

1. Self-similarity $\rightarrow D_{KL}(p\|p) = 0$,

2. Self-identification $\rightarrow D_{KL}(p\|q) = 0$ only if $p = g$,

3. Positivity $\rightarrow D_{KL}(p\|q) \geq 0$ for all p, g.

To understand the information that KL divergence provides clearly, let's create a toy example. Assume that a sequence of data, x has been observed and the distribution of these samples, $f(x)$ will be tested to be uniform or binomial distributions which refer to f_1 and f_2, respectively. The equation has been used and KL divergence values are calculated. Resulting values are, $D_{KL}(f\|f_1) = 0.22$ and $D_{KL}(f\|f2) = 0.105$. Examining the KL divergence values, we can state that the distribution of observed samples is more likely to come from uniform distribution, or conversely approximating the distribution of the observed samples with binomial distribution causes more information loss than uniform distribution.

C.2. Kolmogorov-Smirnov Test

Kolmogorov-Smirnov test, or simply KS test, can be defined as a non-parametric test which can be used to test equality of continuous distributions by comparing one sample with a reference distribution (one-sample KS test) or two samples (two-sample KS test).

Particularly, suppose that a population has a cumulative distribution function $F(x)$ (reference distribution) which is clearly specified. There is also an observed population the empirical cumulative distribution function of which is $G(x)$. One can think that a measure between these two distributions may provide means about whether the reference distribution is the correct one or not. KS test quantifies a measure for this purpose as (Massey Jr, 1951; Goodman, 1954; Wilcox, 2005)

$$D_{KS} = \max_x |F(x) - G(x)|. \tag{C.4}$$

If the calculated KS score is large, this provides evidence that the reference distribution $F(x)$ is not the correct distribution for the observed samples. The measure D_{KS} can be defined as one-sample KS score (or measure). If we deal with observations from two
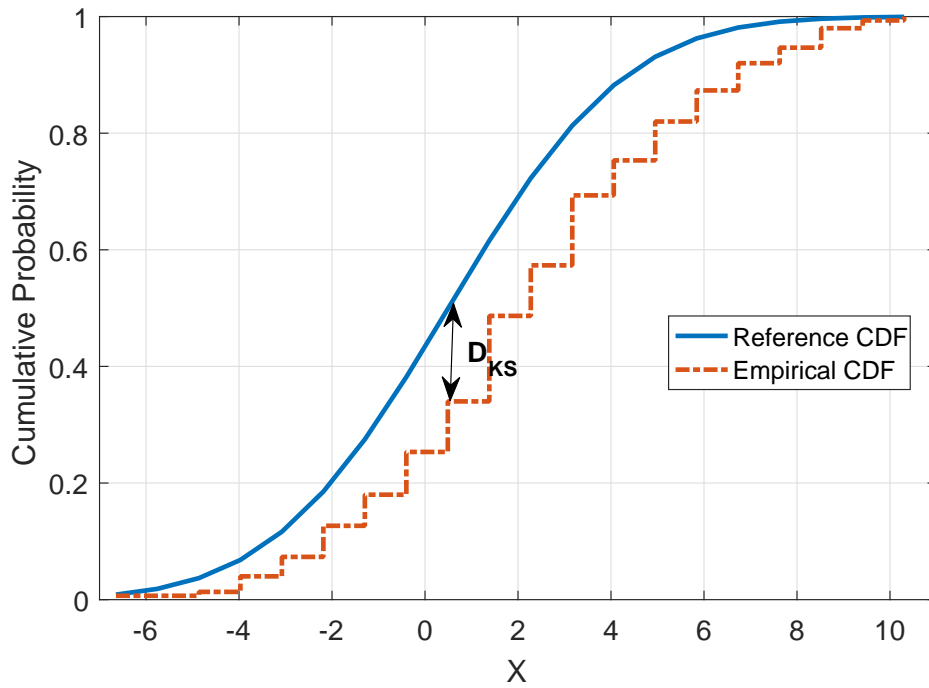
Figure C.1. KS Score calculation example.

populations the empirical cumulative distribution functions of which are $F_1(x)$ and $F_2(x)$. Here, KS test will be used to test whether two populations come from the same distribution or not. Two-sample KS test score is

$$D_{KS} = \max_x |F_1(x) - F_2(x)|. \tag{C.5}$$

KS score has also meaning from a graphical point of view. The largest vertical distance between two cumulative distribution functions can be defined as KS test score (Wilcox, 2005). In Figure C.1, an example to KS score is shown.

In addition to the KS test score explained above, KS test also performs a hypothesis testing under the null hypothesis that says "two populations are drawn from the same underlying continuous population". This hypothesis, $\mathcal{H}$, is rejected providing that any given significance value, α, is as large as or larger than p-value. In order to calculate p-value, the limiting forms of Kolmogorov's distribution should be calculated (Massey Jr,

1951; Wang et al., 2003; Press, 2007) as

$$\lim_{n \to \infty} Pr(D_{KS} \leq t) = L(t) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} \exp(-2i^2 t^2). \tag{C.6}$$

The corresponding p-value can be computed as (Press, 2007; Tong et al., 2010)

$$p\text{-value} = Pr(D_{KS} > t) = 1 - L(t) = 2 \sum_{i=1}^{\infty} (-1)^{i-1} \exp(-2i^2 t^2). \tag{C.7}$$

There is still one unknown, t, and it can be obtained approximately as (Stephens, 1970; Press, 2007)

$$t = D_{KS} \left[N_e + 0.12 + \frac{0.11}{N_e} \right], \tag{C.8}$$

where N_e refers to the sample size, N, for one-sample KS test and $\sqrt{\dfrac{N_1 N_2}{N_1 + N_2}}$ for two-sample KS test. Substituting t obtained via (C.8) in (C.7) gives p-value.

APPENDIX D

Q-Q PLOTS

Quantile-Quantile plot, or simply Q-Q plot can be described as a graphical representation of the sorted quantiles of a data set against the sorted quantiles of another data set. Suppose we have two samples with length n, $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$. In terms of Q-Q plot, these two samples are from the same distribution, as long as their ordered sequences, $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ and $Y_{(1)}, Y_{(2)}, \ldots, Y_{(2)}$, should satisfy, $X_{(i)} \approx Y_{(i)} \quad i = 1, 2, \ldots, n$.

Q-Q has been used to compare distributions of two populations, or compare distribution of one population to a reference distribution. Q-Q plot may provide information about the location, shape and scale parameters comparison between two populations. Q-Q plots provide different characteristics for erroneous cases whose details are given in Figure D.1.
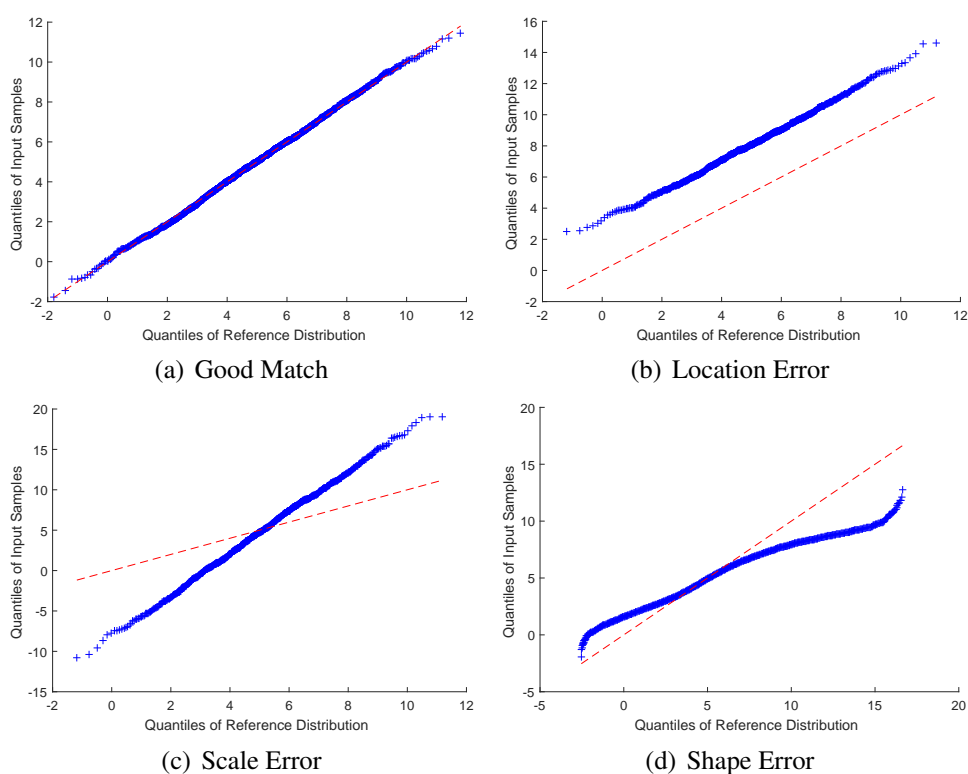

(a) Good Match (b) Location Error

(c) Scale Error (d) Shape Error

Figure D.1. Q-Q plots example figures for good match and erroneous cases.

VITA

Date and Place of Birth: 15.05.1984, İzmir-TURKEY

EDUCATION

2012 - 2017 Doctor of Philosophy in Electronics and Communication Engineering
Graduate School of Engineering and Sciences, İzmir Institute of Technology, Turkey
Thesis Title: GENERALISED BAYESIAN MODEL SELECTION USING
REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO
Supervisor: Assoc. Prof. Dr. Mustafa A. ALTINKAYA
Co-Supervisor: Assoc. Prof. Dr. Ercan E. KURUOĞLU

2009 - 2012 Master of Science in Electronics and Communication Engineering
Graduate School of Engineering and Sciences, İzmir Institute of Technology, Turkey
Thesis Title: PERFORMANCE COMPARISON OF EUROPEAN DIGITAL
VIDEO BROADCASTING STANDARDS UNDER STRONG
MULTIPATH INTERFERENCE
Supervisor: Assoc. Prof. Dr. Şevket GÜMÜŞTEKİN

2007 - 2009 Bachelor of Science in Electronics Engineering
Faculty of Engineering and Architecture, İstanbul Kültür University, Turkey

PUBLICATIONS

Karakuş, O., E. E. Kuruoğlu, and M. A. Altınkaya (2015). Estimation of the nonlinearity degree for polynomial autoregressive processes with RJMCMC. In *23rd European Signal Processing Conference (EUSIPCO)*, pp. 953–957. IEEE.

Karakuş, O., E. E. Kuruoğlu, and M. A. Altınkaya (2016). Bayesian estimation of polynomial moving average models with unknown degree of nonlinearity. In *24th European Signal Processing Conference (EUSIPCO)*, pp. 1543–1547. IEEE.

Karakuş, O., E. E. Kuruoğlu, and M. A. Altınkaya (2017). Bayesian Volterra system identification using reversible jump MCMC algorithm. *Signal Processing, 141*, 125–136.

Karakuş, O., E. E. Kuruoğlu, and M. A. Altınkaya (2017). Nonlinear model selection for PARMA processes using RJMCMC. In *25th European Signal Processing Conference (EUSIPCO)*, pp. 2110–2114. IEEE.