Beyond Trans-dimensional RJMCMC: Application to Impulsive Data Modeling

O. Karakuş[a,*], E.E. Kuruoğlu[b], M.A. Altınkaya[a]

[a] *İzmir Institute of Technology (IZTECH), Electrical-Electronics Engineering, İzmir, Turkey*
[b] *ISTI-CNR, via G. Moruzzi 1, 56124, Pisa, Italy*

Abstract

Reversible jump Markov chain Monte Carlo (RJMCMC) is a Bayesian model estimation method which has been generally used for trans-dimensional sampling and model order selection studies in the literature. In this study, we have utilized RJMCMC beyond trans-dimensional sampling and the proposed usage, which we call *trans-space RJMCMC*, reveals and draws attention to the generality and potentials of RJMCMC by exploiting the original formulation to explore spaces of different classes or structures. This provides flexibility in using different types of candidate classes in the combined model space such as spaces of linear and nonlinear models or of various distribution families. As for application, we have performed a special case of trans-space sampling, namely *trans-distributional* RJMCMC in impulsive data modeling. In many areas such as seismology, radar, image, using Gaussian models is a common practice due to analytical ease. However, many noise processes do not follow a Gaussian character and generally exhibit events too impulsive to be successfully described by the Gaussian model. We test the proposed usage of RJMCMC to choose between various impulsive distribution families to model both synthetically generated noise processes and real life measurements on *power line communications* (PLC) impulsive noises and *2-D discrete wavelet transform* (2-D DWT) coefficients.

Keywords: Reversible jump MCMC, Impulsive data modeling, PLC noise modeling, Wavelet coefficients modeling, Symmetric α-stable distribution, Generalised Gaussian distribution, Student's t distribution.

1. Introduction

Reversible jump Markov chain Monte Carlo (RJMCMC) is a Bayesian model determination method which has had success in vast areas of applications since its introduction by Peter Green [1]. Unlike the widespread MCMC algorithm, *Metropolis-Hastings* (MH), RJMCMC allows one to search in solution spaces of different dimensions

*Corresponding author

Email addresses: oktaykarakus@iyte.edu.tr (O. Karakuş), ercan.kuruoglu@isti.cnr.it (E.E. Kuruoğlu), mustafaaltinkaya@iyte.edu.tr (M.A. Altınkaya)

which has been the main motivation for its use up to date. Classical applications of RJMCMC are model selection in regression and mixture processes [2, 3, 4, 5, 6, 7]. Unlike the classical applications in the literature, the original formulation of RJMCMC in [1] permits a wider interpretations than just exploring the models with different dimensions. As an example of the applicability of RJMCMC beyond model dimension selection: it was utilized to learn polynomial autoregressive (PAR) [8], polynomial moving average (PMA) [9] and polynomial autoregressive moving average (PARMA) [10] processes and identification of Volterra system models [11] by exploring linear and nonlinear model spaces in preliminary works by the authors.

Apart from the classical MCMC methods, during the years, various methods have been developed to solve Bayesian problems. In [12], independence sampler (IS) which is a special case for MH algorithm has been proposed. IS works successfully if the proposal distribution can be defined as a good approximation to the target distribution. In [13] a Gibbs sampling based method has been proposed by Carlin and Chib. This method suggests generating pseudoprior at each realization of the Gibbs sampling and may be computationally inefficient. An alternative to Carlin and Chib's method is using an accept/reject procedure instead of sampling from a full conditional distribution in Gibbs sampling. This method can be named as Metroplised Carlin and Chib (MCC) as in [14]. A modification has been applied to Carlin and Chib's method for variable selection applications and named as Gibbs variable selection in [14]. Methods such as Dellaportas' [14] and Carlin and Chib's [13] have been generally seen as rival methods to RJMCMC. However the application areas of these methods, to the best of our knowledge, are generally limited to regression problems, mixture processes, etc. RJMCMC offers wider meaning and has wider applications. On the other hand, RJMCMC being as an extended version of MH algorithm, is much more general and flexible than these methods as Gibbs sampling has been a special case of MH algorithm.

In [15], Simon J. Godsill provided an important work on generality of RJMCMC and similarities between the Carlin and Chib's method. In that study, a composite product space was created for reversible jump mechanism. The general perspective is to make the model dimension invisible in the operations, and at each iteration, problem turns into a fixed dimension case which can be solved via MCMC methods. The strengths and weaknesses of reversible jump mechanism was provided and authors stated that applying this procedure may be somehow problematic especially in non-nested problems. Applications are on variable and model order selection and show superiority of the RJMCMC.

Apart from all other studies discussed above, this paper contributes to the literature with a generalization on RJMCMC beyond trans-dimensional sampling, which we call *trans-space RJMCMC*. The proposed method follows the generality of the formulation of Green [1] and emphasizes its potential to be a general estimation method by performing the reversible jump mechanism between spaces of different model classes rather than just being a trans-dimensional approach and a model order selection method.

Performing transition between non-nested or different classes of models needs much more attention on generating proposal distributions. In order to increase the convergence speed and avoiding local traps in the algorithm, we propose common feature based proposals, specifically norm based transitions between different classes of models. The proposed usage and common parameter based proposal approach easily exhibit the generality and the potential of the original formulation of RJMCMC. In order to demonstrate this potential in this paper, we focus our attention on a more special but generic problem of choosing between different probability distribution families. The problem is a frequently encountered problem in signal processing and statistics, and their application fields such as in image processing and telecommunications. In various real-life modeling 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, eliminating the need for modeling with each candidate model class separately and comparing. This provides computational gains especially when the number of parameters and candidate model classes are 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 in 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.

General practice is to model noise/data with a Gaussian process especially in communications, network modelling, digital images, due to its analytical ease. In the case of non-Gaussian impulsive noise/data, various model families exist, for example, Middleton Class A, Bernoulli-Gaussian, α-Stable, Generalized Gaussian (GG), Student's t, etc. It has been reported in the literature that noise exhibits non-Gaussian and impulsive characteristics in application areas such as wireless communications [16, 17], *power line communications* (PLC) [18, 19], *digital subscriber lines* (xDSL) [20, 21], image processing [22, 23] and seismology [24].

In this paper, we propose a Bayesian statistical modeling study of impulsive noise/data by estimating the probability distribution among three conventional impulsive distributions families: symmetric α-Stable (SαS), GG and Student's *t*. Other than identifying the distribution family, the proposed method estimates shape and scale parameters of the distribution. These distributions are the most popular statistical models in applications covering diverse areas such as wireless channel modeling, financial time series analysis, seismology, radar imaging.

We study the algorithm extensively on synthetic data providing statistical significance tests. In addition, as case studies, we look into two statistical modeling problems of actual interest impulsive noise on PLC channels and *2-D discrete wavelet transform* (2-D DWT) coefficients. Particularly, PLC impulsive noise measurements in [25, 26] have been utilized in the simulations. Apart from this, statistical modeling for 2-D DWT coefficients have been performed on different kinds of images such as Lena, *synthetic aperture radar* (SAR) [27], *magnetic resonance imaging* (MRI) [28] and mammogram [29].

Rest of the paper is organized as follows: general definitions for trans-dimensional RJMCMC and the proposed method are discussed in Section 2. Section 3 reviews three distribution families and describes the impulsive data modeling scheme of the proposed method. Experimental studies for synthetically generated noise processes and for real applications are explained in Section 4. Section 5 draws conclusions on the results.

2. Reversible jump MCMC

RJMCMC has been first introduced by Peter Green in [1] as an extension of MCMC to a model selection method. 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. In the follow up, to the best of our knowledge almost all publications utilized RJMCMC for model dimension selection. Popular use of RJMCMC is in linear parametric models such as *autoregressive* (AR) [2], *autoregressive integrated moving average* (ARIMA) [3] and *fractional* ARIMA (ARFIMA) [4] and mixture models such as Gaussian mixtures [5], Poisson mixtures [6] and α-stable mixtures [7].

Apart from the popular applications above, RJMCMC has been used in other various applications such as detection of clusters in disease maps [30], graphical models based variable selection and automatic curve fitting [31], log-linear model selection [32], non-parametric drift estimation [33], delimiting species using multilocus sequence data [34], random effect models [35], generation of lane-accurate road network maps from vehicle trajectory data [36].

In this study, our motivation is to draw attention to the generality of the classical RJMCMC beyond trans-dimensionality. The classical RJMCMC algorithm of [1] and the proposed usage, *trans-space* RJMCMC are discussed in the sequel.

The standard MH algorithm [37] accepts a transition from Markov chain state $x \in \mathcal{X}$ to $y \in \mathcal{X}$ with a probability of:

$$A(x \to y) = \min\left\{1, \frac{\pi(y)q(x,y)}{\pi(x)q(y,x)}\right\} \tag{1}$$

where $\pi(\cdot)$ represents the target distribution and $q(y,x)$ refers to the proposal distribution from state x to y.

RJMCMC, in the sense of trans-dimensional MCMC, generalizes MH algorithm by defining multiple parameter subspaces ζ_k of different dimensionality [1]. This is only achieved by defining different types of moves between subspaces providing that the detailed balance is attained. For this condition to hold, a reverse move from state y to x should be defined and dimension matching should be satisfied between parameter subspaces.

Assume that we propose a move m with probability p_m from a Markov chain state κ to κ' each of which has parameter vectors $\boldsymbol{\theta} \in \zeta_1$ and $\boldsymbol{\theta}' \in \zeta_2$, respectively with different dimensions. The move m is reversible and its reverse move m^{R} is proposed with a probability $p_{m^{\mathrm{R}}}$. The general detailed balance condition can be stated as:

$$\pi(\kappa)q(\kappa',\kappa)A(\kappa \to \kappa') = \pi(\kappa')q(\kappa,\kappa')A(\kappa' \to \kappa), \tag{2}$$

where proposal distribution $q(\cdot)$ is directional and includes the probabilities of both the move itself and the proposed parameters. Then, the general expression for the acceptance ratio in (1) turns into [1]:

$$A(\kappa \to \kappa') = \min\left\{1, \frac{\pi(\kappa')p_{m^{\mathrm{R}}}\chi_2(\mathbf{u}')}{\pi(\kappa)p_m\chi_1(\mathbf{u})}\left|\frac{\partial(\boldsymbol{\theta}',\mathbf{u}')}{\partial(\boldsymbol{\theta},\mathbf{u})}\right|\right\}, \tag{3}$$

where $\chi_1(\cdot)$ and $\chi_2(\cdot)$ are the distributions for the auxiliary variable vectors $\mathbf{u}$ and $\mathbf{u}'$, respectively which are required to provide dimension matching for the moves m and m^{R}. The term $\left|\frac{\partial(\boldsymbol{\theta}'\mathbf{u}')}{\partial(\boldsymbol{\theta},\mathbf{u})}\right|$ is the magnitude of the Jacobian.

In each RJMCMC run, the standard Metropolis-Hastings algorithm is applied in moves within the same dimensional models, which is called as *life* move. Sampling is performed in a single parameter space and there is no dimension change in life move. For trans-dimensional transitions between models, moves such as *birth*, *death*, *split* and *merge* are performed which require the creation or the deletion of new variables corresponding to the increased or decreased dimension. Green handles the dimension changing moves as variable transformations and defines a dummy variable to match dimensions which provides a square Jacobian matrix that can be used to update the acceptance ratio easily.

2.1. Trans-space RJMCMC

In spite of RJMCMC's use in trans-dimensional cases, the original formulation in [1] 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 defining the spaces of model parameters.

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) may belong to Borel sets which differ not only in their dimensions but also in the generic models they belong to. Thus, the RJMCMC 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 (e.g. 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.

As in the case of all reversible jump applications, providing such proposals may be somehow hard, however, using a common feature provides users various application areas and an opportunity to utilize RJMCMC on model estimation studies of different classes of models. The proposed method is applicable to the nested cases the model space of which consists of related models. However, the importance of this approach significantly appears for the un-nested cases where the feature-based approach offers flexibility for RJMCMC moves between different classes models. Two examples one can think of firstly, are:

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

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

To this end, we define a combined parameter space $\varphi = \bigcup_k \varphi_k$ for $k > 1$. 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/or $\mathbf{u}_2$ 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 of Green is still applicable.

Consequently, although the candidate spaces are of different classes, since the Borel sets are defined as to be related, the assumption of Green 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). Thus, the acceptance ratio can be written as:

$$A(x \to x') = \min\left\{1, \frac{\pi(x')p_{M^{\mathrm{R}}}Q_2(\mathbf{u}_2)}{\pi(x)p_M Q_1(\mathbf{u}_1)}\left|\frac{\partial h_{12}(\boldsymbol{\theta}_1, \mathbf{u}_1)}{\partial(\boldsymbol{\theta}_1, \mathbf{u}_1)}\right|\right\}. \tag{4}$$

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

Here we recall that in our previous works [8, 9, 10, 11], we have performed model estimation studies with RJMCMC for Volterra based nonlinear models PAR, PMA and PARMA as well as an identification study of Volterra system models. In these studies, RJMCMC has been utilized to explore the model spaces of linear and nonlinear models in polynomial sense instead of performing a model order selection study in a single linear model space. Hence, we add a few concluding remarks.

Remark1. We are going to name this general utilization on RJMCMC as *trans-space*. 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, trans-dimensional cases are subsets of trans-space transitions.

Remark2. Trans-space RJMCMC requires to define new types of moves due to the need for more detailed operations than, e.g. just being birth, death, split and merge 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 norm based mapping between the parameter spaces. *Switch* move (firstly proposed for Volterra system identification study [11]) will be proposed as a between-space move, which performs a switching between the candidate spaces of the generic model classes.

Remark3. As a special case of trans-space sampling, the proposed method can be used to explore the spaces of different distribution families. Therefore, this special case will be named as *trans-distributional*.

3. Trans-distributional RJMCMC for Impulsive Distributions

In this 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 define a reversible jump mechanism which estimates the distribution family among three impulsive distribution families, namely, SαS, GG and Student's t.

These three families cover many different noise modeling studies as stated in the above sections. All of them include Gaussian distribution as a special member, and many real life noise measurements can be modelled with these distribution families. For example, SαS family has various demonstrated application areas such as PLC [38], SAR imaging [23], near optimal receiver design [39], modelling of counterlet transform subbands [40], seismic amplitude data modelling [24], as noise model for molecular communication [41], reconstruction of non-negative signals [42] (Please see [43] and references therein for detailed applications).

GG distributions have found applications in wavelet based texture retrieval [44], image modelling in terms of Markov random fields [45], multicomponent texture discrimination in color images [46], wheezing sound detection [47], modelling sea-clutter data [48].

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 [49, 50], full-waveform inversion of seismic data [51], independent vector analysis for speech separation [52], medical image segmentation [53], growth curve modelling [54].

One might argue that training separate MCMC samplers for each of the seemingly irrelevant distribution families and comparing their modelling performances afterwards would be computationally more advantageous. However, in cases when the number of candidate models is not known or dramatically large, implementing a single Markov chain via RJMCMC could be simpler. In addition, when the number of models are small, one can not conclude that parallel MCMC approach would be a better choice than RJMCMC and this requires an analysis. By efficiently choosing the proposal distributions, the advantage of incorporating reversible jump mechanism can be extended to searching several distribution families which will be described in the sequel.

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 [55, 56]. Particularly, when modelling count data, reversible jump mechanism has been applied to choose between Poisson and negative binomial distributions in [55]. This study deals with the question whether the count data is over-dispersed relative to Poisson distribution. In [56] 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 above have utilized RJMCMC in distribution estimation; however, 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), hence, the methods in these studies can be handled with a single family search (i.e. intra-class sampling in this paper which will be discussed below sections). The proposed usage for RJMCMC, namely *trans-distributional* RJMCMC, is much more general than the examples above and aims to fit a distribution to a given process $\mathbf{x}$ 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.

3.1. Impulsive Distribution Families

3.1.1. 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, its characteristic function, $\varphi(x)$, can be expressed explicitly as:

$$\varphi(x) = \exp(j\delta x - \gamma|x|^{\alpha}) \tag{5}$$

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*. The $\gamma > 0$ provides a measure of the dispersion which is the *scale parameter* expressing the spread of the distribution around δ.

3.1.2. Generalized Gaussian Distribution Family

The univariate GG pdf can be defined as:

$$f(x) = \frac{\alpha}{2\gamma\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x-\delta|}{\gamma}\right)^{\alpha}\right) \tag{6}$$

where $\Gamma(\cdot)$ refers to the gamma function, $\alpha > 0$ is the shape parameter, $-\infty < \delta < \infty$ represents the location parameter and the $\gamma > 0$ 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.

3.1.3. 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:

$$f(x) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma(\alpha/2)\gamma\sqrt{\pi\alpha}}\left(1+\frac{1}{\alpha}\left(\frac{x-\delta}{\gamma}\right)^2\right)^{-((\alpha+1)/2)}. \tag{7}$$

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.

3.2. Parameter Space

RJMCMC construction for impulsive data modeling begins by firstly 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 defines the estimated distribution family has been 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 study, the observed data from all three families are assumed to be symmetric around the origin for simplicity. Therefore, δ, is set to 0 and its effect will be invisible in the simulations. Consequently, parameter vector $\boldsymbol{\theta}$ is reduced to: $\boldsymbol{\theta} = \{k, \alpha, \gamma\}$.

3.3. Hierarchical Bayesian Model

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

$$f(\boldsymbol{\theta}|\mathbf{x}) \propto f(\mathbf{x}|k, \alpha, \gamma) f(\alpha|k) f(k) f(\gamma). \tag{8}$$

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

3.4. 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:

$$f(\mathbf{x}|k, \alpha, \gamma) = \begin{cases} \prod_{i=1}^{n} \mathrm{S}\alpha\mathrm{S}(\gamma), & k = 1 \\ \prod_{i=1}^{n} \mathrm{GG}_{\alpha}(\gamma), & k = 2 \\ \prod_{i=1}^{n} t_{\alpha}(\gamma), & k = 3 \end{cases} \tag{9}$$

3.5. Priors

Priors have been selected as the following:

$$f(\gamma) = \mathcal{IG}(a, b), \tag{10}$$

$$f(k) = \mathbb{I}_{\{1/3,1/3,1/3\}} \quad \text{for } k = 1, 2, 3, \tag{11}$$

$$f(\alpha|k) = \begin{cases} \mathcal{U}(0, 2) & k = 1, \\ \mathcal{U}(0, \alpha_{\text{max,GG}}) & k = 2, \\ \mathcal{U}(0, \alpha_{\text{max},t}) & k = 3, \end{cases} \tag{12}$$

where a and b represent the hyperparameters for scale parameter and they are generally selected as 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 have been defined as $\alpha_{\text{max,GG}}$ and $\alpha_{\text{max},t}$, respectively.

Choosing an inverse gamma prior for scale parameter is a general practice especially for Gaussian problems. Due to the lack of information 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 has been chosen for simplicity. Furthermore, all families are equiprobable *a priori* and shape parameter is uniformly distributed between lower and upper bounds.

3.6. Model Moves

Two RJMCMC model moves have been defined in order to perform trans-distributional transitions discussed in the previous sections. These are: *life* and *switch* moves. Life move performs classical MH algorithm to update γ. 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 $P_{\text{life}}, P_{\text{intra-cl-sw}}$ and $P_{\text{inter-cl-sw}}$, respectively. Different types of moves can, of course, be created to solve this problem. Since the main purpose of this study is to draw attention to the generality of RJMCMC algorithm and to provide its applications on the real data measurements, we only focus on the between-space move switch and its different usages intra and inter class transitions.

In Figure 1 the flow diagram of the proposed method is depicted where the parameter N refers to the maximum number of iterations. The details about the steps of the selected moves are discussed in the sequel.

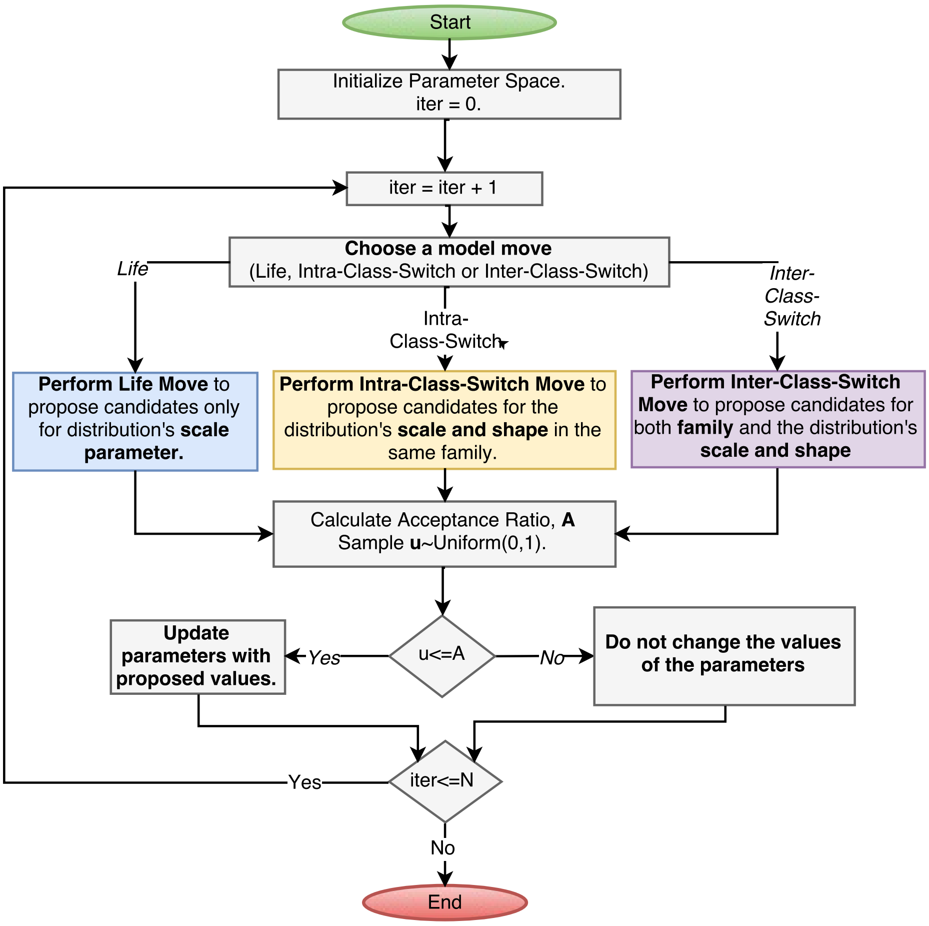


Figure 1: Flow Diagram for the Proposed method.

3.6.1. *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' has been chosen as:

$$q(\gamma'|\gamma) = \mathcal{TN}(\gamma, \xi_{scale}) \quad \text{for interval } (0, \gamma + 1] \tag{13}$$

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:

$$A_{\text{life}} = \min\left\{1, \frac{f(\mathbf{x}|k', \alpha', \gamma')}{f(\mathbf{x}|k, \alpha, \gamma)} \frac{f(\gamma')}{f(\gamma)} \frac{q(\gamma|\gamma')}{q(\gamma'|\gamma)}\right\} \tag{14}$$

3.6.2. FLOM Based Proposals for γ Transitions

As mentioned earlier in this paper, 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 Student's 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. By taking into consideration of the facts that absolute FLOM expressions are defined for all impulsive families, and their success in parameters estimation studies of the SαS distributions, using an absolute FLOM based approach helps 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 study, absolute FLOM-based approach will be used for the proposals of the γ parameter. In particular, to perform sampling between 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.

Absolute FLOMs are defined only for p values lower than alpha 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. [57] suggests $p = \alpha/4$ and [58] suggests $p = 0.2$. However, in [59] 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 [59] suggests the choice $p = \alpha/10$. We use 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 will be the same for both the most recent and candidate distribution spaces. In particular,

$$E_k(|\mathbf{x}|^p) = E_{k'}(|\mathbf{x}|^p) \tag{15}$$

where absolute FLOMs for all three candidate families can be defined as:

$$E_k(|\mathbf{x}|^p) = \begin{cases} C_\alpha(p,\alpha)\gamma^{p/\alpha} & k = 1, \\ C_{\mathrm{GG}}(p,\alpha)\gamma^p & k = 2, \\ C_t(p,\alpha)\gamma^p & k = 3, \end{cases} \tag{16}$$

where

$$C_\alpha(p,\alpha) = \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{-p}{\alpha}\right)}{\alpha\sqrt{\pi}\Gamma\left(\frac{-p}{2}\right)} 2^{p+1}, \tag{17}$$

$$C_{\mathrm{GG}}(p,\alpha) = \frac{\Gamma\left(\frac{p+1}{\alpha}\right)}{\Gamma(1/\alpha)}, \tag{18}$$

$$C_t(p,\alpha) = \frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{\alpha-p}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\alpha}{2}\right)} \alpha^{p/2}. \tag{19}$$

The candidate proposal, γ', has been calculated via reversible functions which are derived by using the relations in (15)-(19) for each transition. These functions have been derived for both of the switch moves and are shown in Tables 1 and 2.

3.6.3. 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 is dependent on the newly proposed α' parameter and firstly one step is performed on shape parameter α to propose α'. The resulting shape parameter values are 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. An example figure of the proposal distribution $q(\alpha'|\alpha)$ is shown in Figure 2(a).

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, this might affect the convergence speed of the algorithm.

Candidate scale parameter γ' has been calculated via reversible functions, $g(\cdot)$, which are derived for intra-class-switch move by using the method in Section 3.6.2. Functions for each family are shown in Table 1.

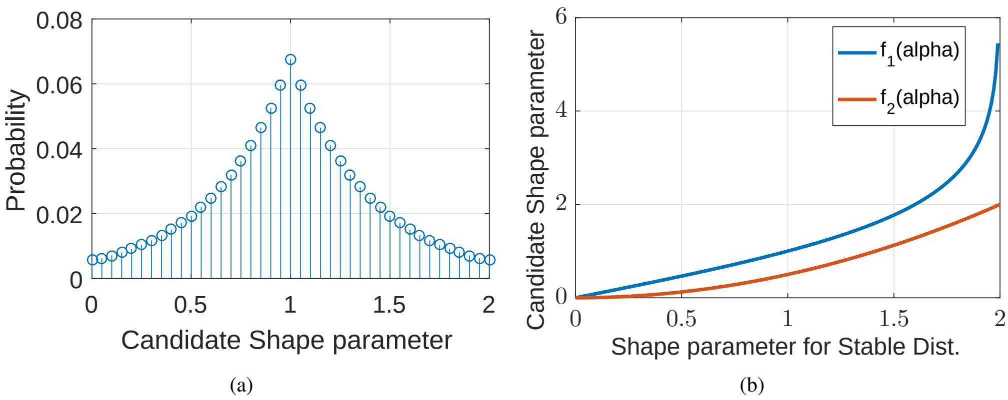


Figure 2: (a) - Proposal distribution, $q(\alpha'|\alpha)$ for intra-class-switch move ($\gamma = 1, \Gamma = 0.4$). (b) - Mapping functions on shape parameter for inter-class-switch move

Consequently, proposals for intra-class-switch move are;

$$q(\alpha'|\alpha) = \mathcal{DL}(\alpha, \Gamma), \tag{20}$$

$$\gamma' = g(\alpha, \alpha', p, \gamma). \tag{21}$$

Table 1: Intra-Class-Switch Details $[(k,\alpha,\gamma)\to(k',\alpha',\gamma')]$

Family	Degree, p	$\gamma' = g(\alpha,\alpha',p,\gamma)$	Jacobian, $\lvert J\rvert$
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	$\alpha'/10$	$\left(\dfrac{C_{\mathrm{GG}}(p,\alpha)}{C_{\mathrm{GG}}(p,\alpha')}\right)^{1/p}\gamma$	$\left(\dfrac{C_{\mathrm{GG}}(p,\alpha)}{C_{\mathrm{GG}}(p,\alpha')}\right)^{1/p}$
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 2: Inter-Class-Switch Details $[(k,\alpha,\gamma)\to(k',\alpha',\gamma')]$

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

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

$$A_{\text{intra-cl-sw}} = \min\left\{1, \frac{f(\mathbf{x}|k',\alpha',\gamma')}{f(\mathbf{x}|k,\alpha,\gamma)}\frac{f(\gamma')}{f(\gamma)}|J|\right\}, \tag{22}$$

where $|J|$ is the magnitude of the Jacobian (See Table 1).

3.6.4. Inter-Class-Switch Move

Different from intra-class-switch move, distribution family has also been 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'.

$$\alpha' = \psi(\alpha,k,k') \tag{23}$$

$$\gamma' = w(\alpha,\alpha',p,\gamma) \tag{24}$$

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 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 2.

In order 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. 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 have been depicted in Figure 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 have been shown in Table 2.

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

$$A_{\text{inter-cl-sw}} = \min\left\{1, \frac{f(\mathbf{x}|k', \alpha', \gamma')}{f(\mathbf{x}|k, \alpha, \gamma)} \frac{f(\gamma')}{f(\gamma)} \frac{f(\alpha|k)}{f(\alpha'|k')} |J| \right\} \tag{25}$$

where $|J| = \dfrac{\partial \gamma'}{\partial \gamma} \dfrac{\partial \alpha'}{\partial \alpha}$.

4. Experimental Study

We study experimentally three cases: synthetically generated noise, impulsive noise on PLC channels and 2-D DWT coefficients. 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 $\alpha_{\max,\mathrm{S}\alpha\mathrm{S}}$, $\alpha_{\max,\mathrm{GG}}$

and $\alpha_{\max,t}$ 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 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 (for details of KL divergence please see [60]). 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. (For details of KS test, please see [61])

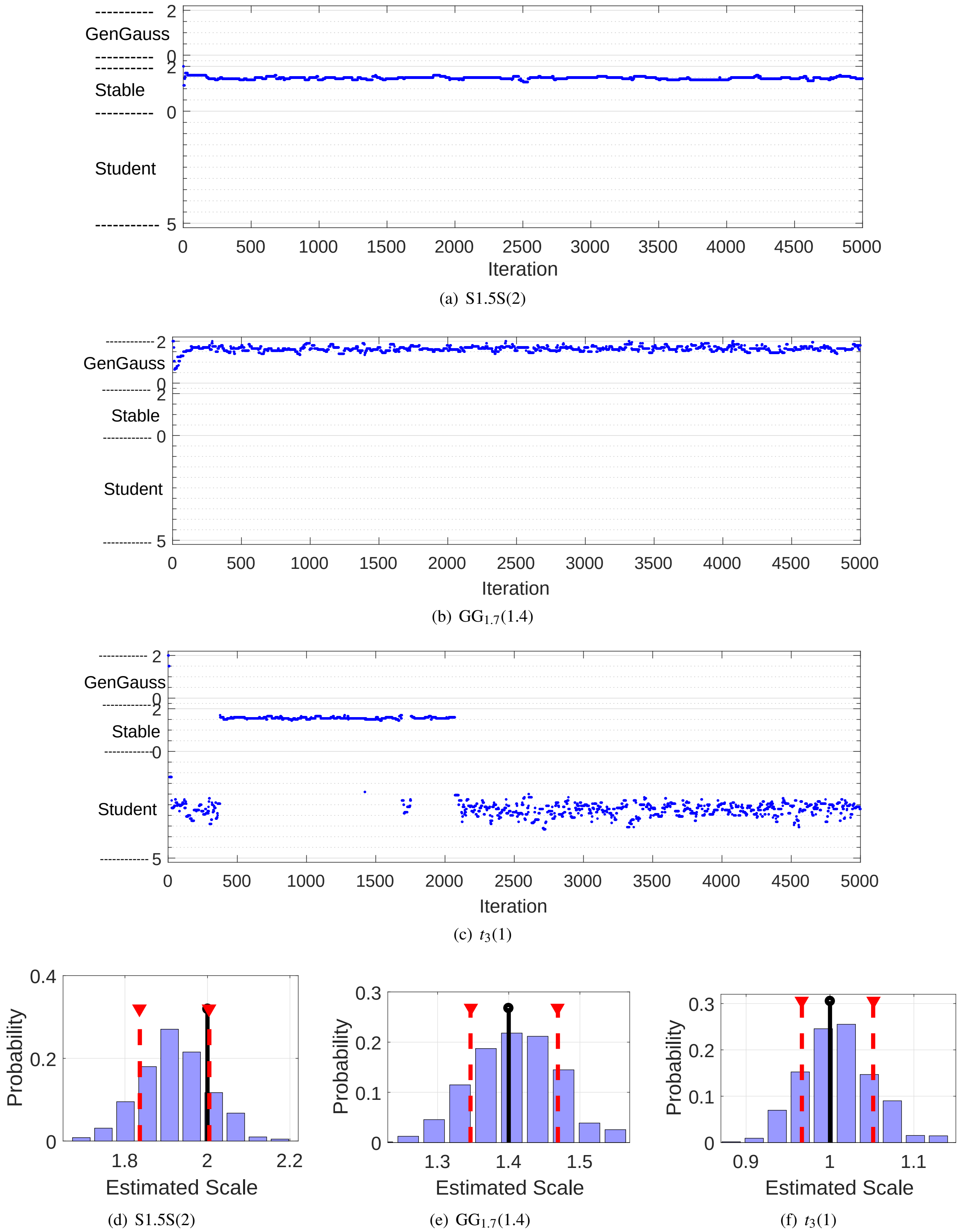
Table 3: Modeling results for synthetically generated processes.

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
$\text{GG}_{0.5}(0.5)$	GG	0.4990	0.5199	0.0229	0.0152	1.0000
$\text{GG}_{1.7}(1.4)$	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

4.1. Case Study 1: Synthetically Generated Noise Modeling

In order to test the proposed method on modeling 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), $\text{GG}_{0.5}(0.5)$, $\text{GG}_{1.7}(1.4)$, $t_3(1)$ and $t_{0.6}(3)$.

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



(a) S1.5S(2)

(b) $GG_{1.7}(1.4)$

(c) $t_3(1)$

(d) S1.5S(2)

(e) $GG_{1.7}(1.4)$

(f) $t_3(1)$

Figure 3: Synthetically generated noise modeling - 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.

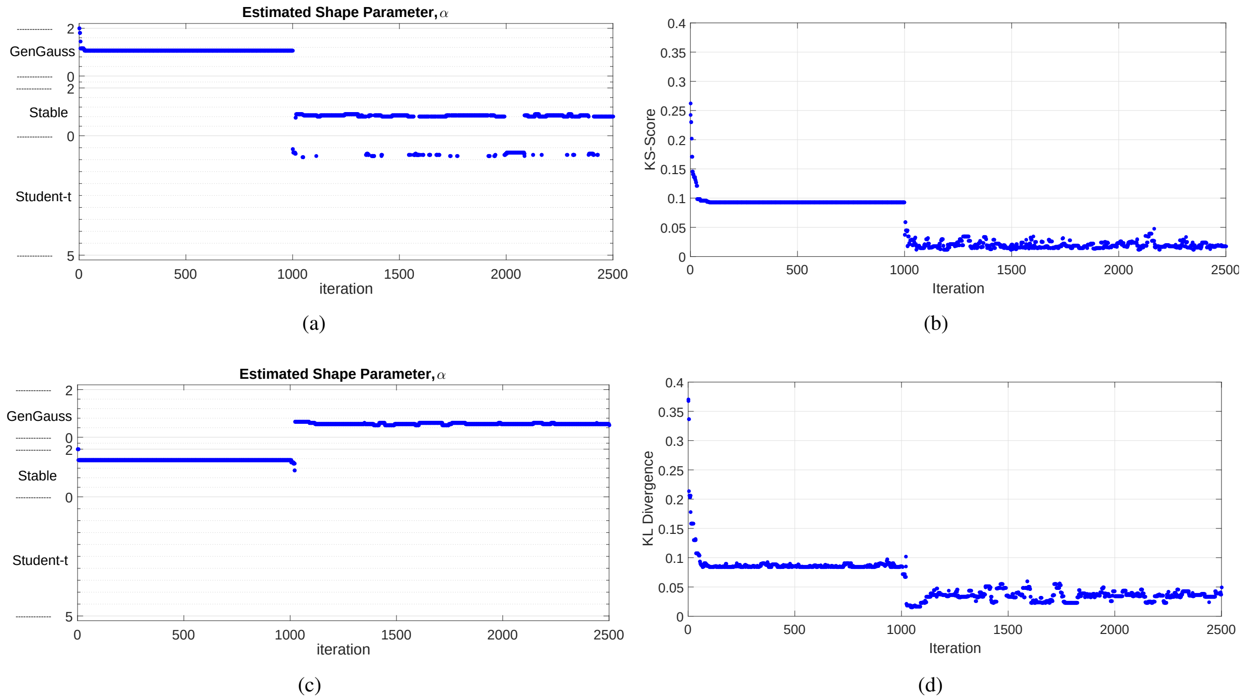


Figure 4: Wrong distribution initialized simulation for Case 1. (a) and (c) refer to the instantaneous shape parameter estimation plots for 2500 iterations. (b) and (d) refer to the instantaneous KS (or KL) statistics plots for 2500 iterations. The correct distributions are Cauchy and Generalized Gaussian for the first and second rows, respectively.

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 3. In Figure 3, instantaneous estimate of shape parameter α and estimated posterior distribution of scale parameter γ are shown 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 subfigures (a)-(c) in order to show the transient characteristics of the algorithm. These plots show that the proposed usage of RJMCMC with FLOM based proposal distributions converges to the correct shape parameters. In subfigures (d)-(f), vertical dashed-lines with ∇ markers refer to $\pm\sigma$ *confidence interval* (CI). Examining these subfigures shows that correct scale parameters lie within the $\pm\sigma$ CI of the posteriors.

As another simulation step, we have created a scenario where the algorithm has been forced to remain at a wrong distribution family for the first 1000 iterations. After that, all the limitations are released and algorithm tries to find the correct distribution for a given data set. This simulation has been named as wrong model initialized simulation and results are shown in Figure 4 for two different synthetically generated data sets. Examining the results in Figure 4-(a) and (c), we can easily see that after the wrong model initialization finishes at iteration 1000, the proposed method tries to find the correct distribution family as soon as possible and achieves this transition within the first 50 iterations (between 1000 and 1050). Even if it has been initialized at a completely wrong model, thanks to the norm based proposals, algorithm can find its way towards the correct model very fast. In Figure 4-(b) and (d) statistical error measures are shown in order to visualize that the algorithm remains in a wrong model at first 1000 iterations. As soon as the transition to the correct family has been performed, error measures exhibit a rapid decrease and remain around these values until the end of the simulation.

Estimated pdfs and CDFs for three example distributions are depicted in Figure 5. In addition to the statistical significance values in Table 3, fitting performance of the algorithm has been presented visually. As can be seen in Figure 5, estimated pdfs are very similar to the data histogram and 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 line of Table 3 is meaningful for an example Cauchy distribution, since the Cauchy distribution is a special member in both SαS and Student's t families.

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

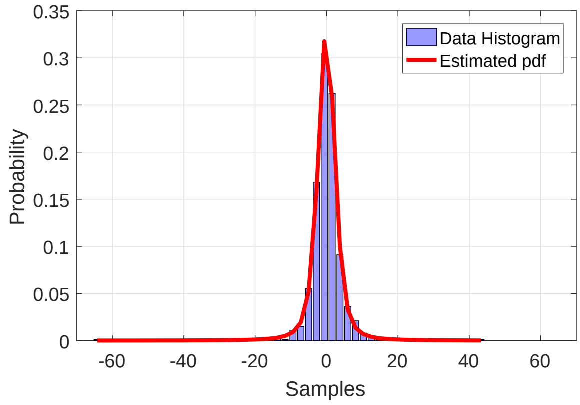
PLC is an emerging technology which utilizes power-lines to carry telecommunication data. Telecommunication speeds up to 200 Mb/s with a good quality of service can be achieved on PLC systems. Apart from this, PLC offers a physical medium for indoor multimedia data traffic without additional cables [38].

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 [25]. Among these, impulsive noise is the most common cause of decoding (or communications) error in PLC systems due to its high amplitudes up to 40 dBs [62].

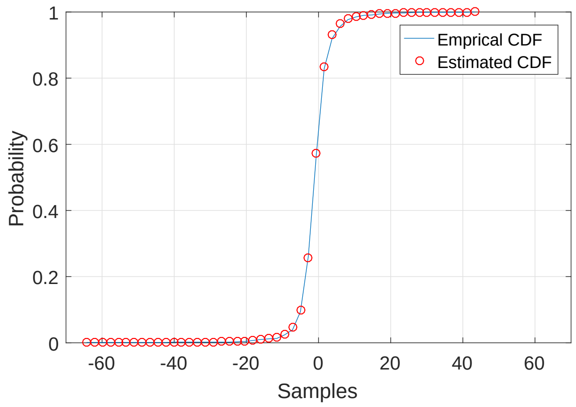
In this case study, we are going to use 3 different PLC noise measurements. First measurement (named as *PLC-1*) has been performed during a project with number PTDC/EEA-TEL/67979/2006. Details for the measurement scheme and other measurements please see [26]. Data utilized in this paper (*PLC-1*) is an amplified impulsive noise measurement from a PLC system with a sampling rate of 200Msamples/sec. Measurements last for 5ms 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 6-(a) a time plot of the utilized downsampled data is depicted (For detailed description of the data please see[2]).

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 [25]). Periodic synchronous measurements last for 4μs and contain 226 noise samples. Periodic

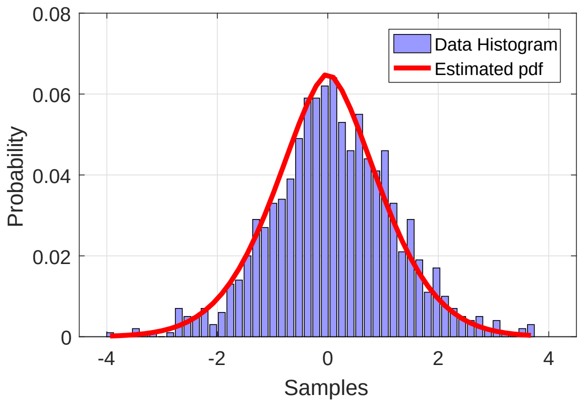
[2]http://sips.inesc-id.pt/~pacl/PLCNoise/index.html



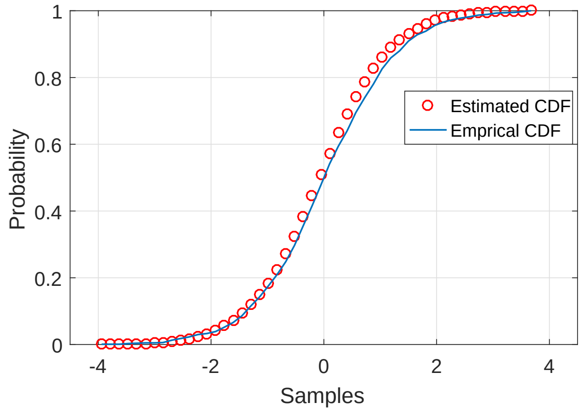
(a) S1.5S(2)



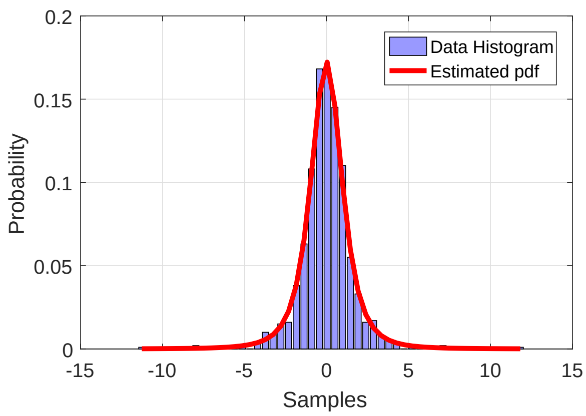
(b) S1.5S(2)



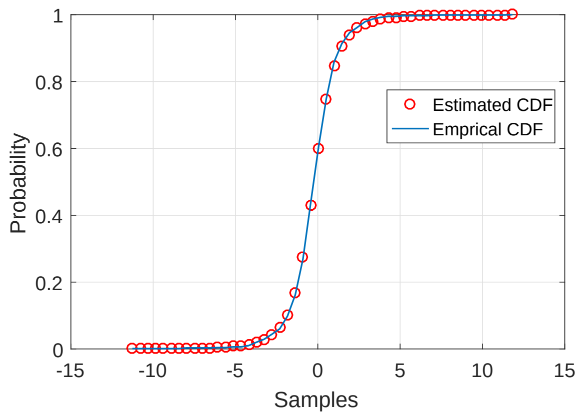
(c) $\mathrm{GG}_{1.7}(1.4)$



(d) $\mathrm{GG}_{1.7}(1.4)$



(e) $t_3(1)$



(f) $t_3(1)$

Figure 5: Synthetically generated noise modeling results. (a)-(c): Estimated pdfs, (d)-(f): Estimated CDFs.

asynchronous measurements contain 1901 noise samples and last for 35μs. In Figures 6 (b) and (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 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 4, we can state that all three considered PLC noise processes follow SαS distribution characteristics. In the literature, there are studies [38, 63] 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 to these studies. According to the results of KL and KS statistics shown in Table 4 on estimated pdfs and CDFs and Figures between 6(d) and 6(i), RJMCMC fits to real data with a remarkable performance. KS p-values are all approximately 1 (> 0.9999) and this provides strong evidence that the estimated and the correct distributions are of the same kind.

4.3. Case Study 3: Statistical Modeling for Discrete Wavelet Transform (DWT) Coefficients

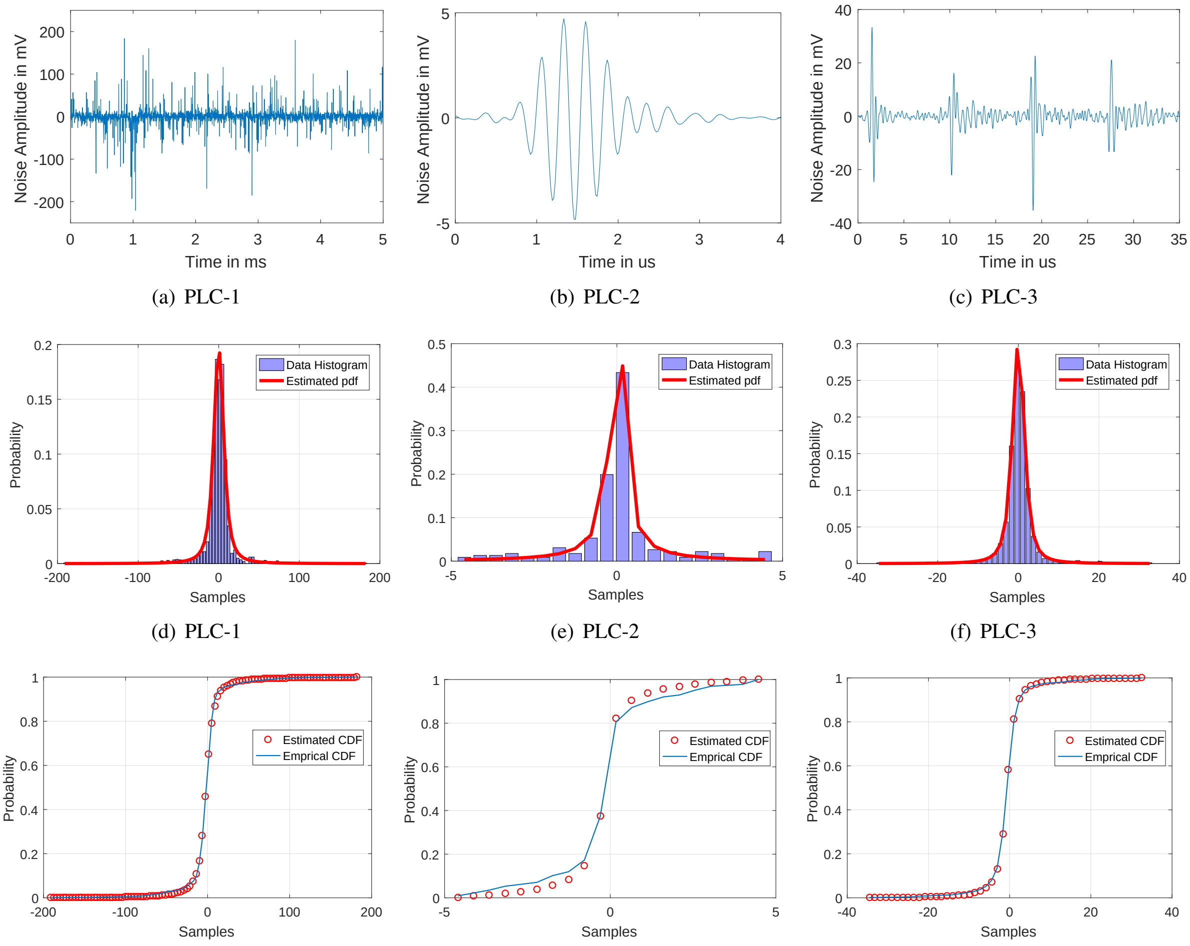
DWT which provides a multiscale representation of an image is a very important tool for recovering local and non-stationary features in an image. The resulting representation is closely related with the processing of the human visual system. DWT obtains this multiscale representation by performing a 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 more heavier tails and sharper peaks than Gaussian distribution [22, 23].

In this 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) [27], *magnetic resonance imaging* (MRI) [28] and mammogram [29] which are shown in the first columns of Figures 7 and 8.

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 as averages of 40 runs.

Estimated distributions for wavelet coefficients of images in Table 5 show different 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

[3]http://www.plc.uma.es/channels.htm



(a) PLC-1 (b) PLC-2 (c) PLC-3

(d) PLC-1 (e) PLC-2 (f) PLC-3

(g) PLC-1 (h) PLC-2 (i) PLC-3

Figure 6: PLC impulsive noise modeling results. (a)-(c): Time plots, (d)-(f): Estimated pdfs, (g)-(i): Estimated CDFs.

Table 4: Modeling 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

Table 5: Modeling results for 2D-DWT coefficients.

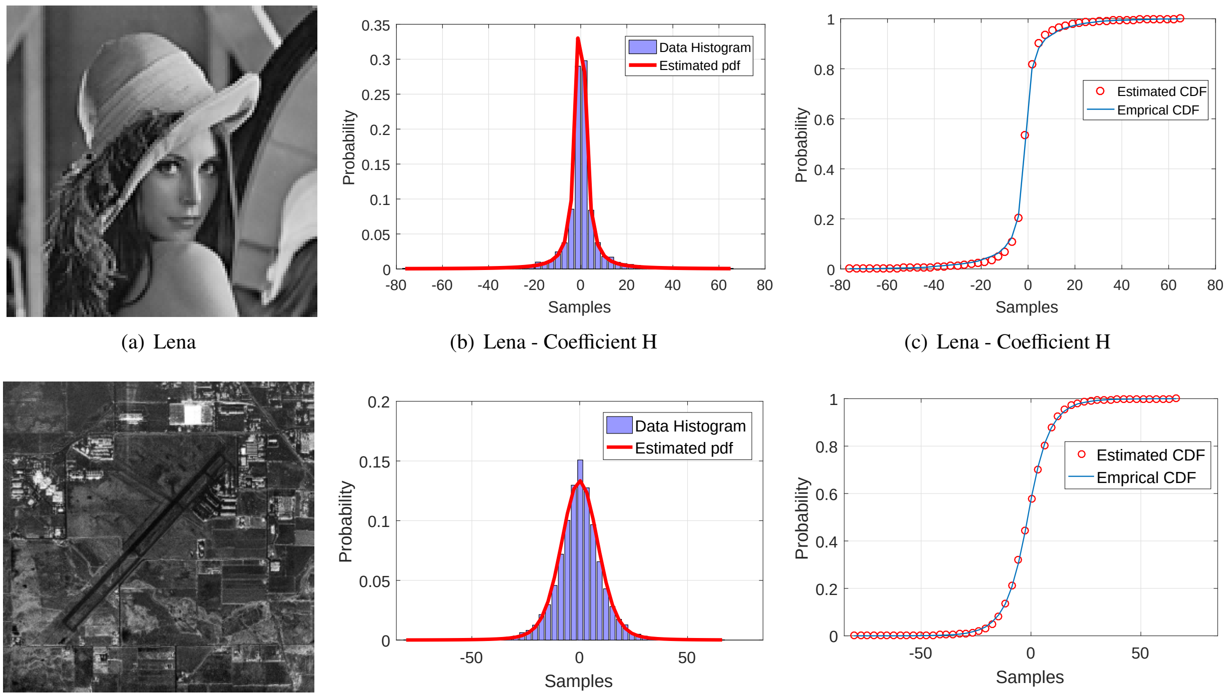
Image	**Est. Family**	**Est. Shape ($\hat{\alpha}$)**	**Est. Scale ($\hat{\gamma}$)**	**KL Div.**	**KS** Score	**KS** p-value
Lena (V)	GG	0.5002	1.7415	0.0271	0.0465	> 0.9999
Lena (H)	t	1.0958	2.2422	0.0094	0.0349	> 0.9999
Lena (D)	t	1.1628	1.7735	0.0145	0.0271	1.0000
SAR(V)	SαS	1.5381	7.7395	0.0025	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

have been modelled successfully according to the KL and KS test scores and p-values. The estimated pdfs and CDFs in Figures 7 and 8 show remarkably good fitting and provide support to the results which are obtained numerically in Table 5.

4.4. Model Switching Analysis for Real Data Sets

As discussed in detail in the previous sections, the proposed usage of RJMCMC in impulsive modelling applications, has 3 different moves. Intra and inter class switch moves perform switching between different distributions and families as well. In order to analyze the model switching capabilities of the proposed model transition approach which is based on a common feature, specifically the FLOMs, instantaneous shape parameter plots are shown Figure 9 for one example data set from each real dat set cases in this study which are PLC and 2D-DWT.

FLOM based proposals demonstrate successful, efficient and fast model transitions leading to the correct (the most suitable of the best matching family) distributions. Except the cases for common distributions in two families such as Cauchy, after reaching the most suitable distribution family, algorithm is more likely to accept sampling in the same family (intra-class switch move) rather than perform sampling between families (inter-class switch move). The most important reason for this is that the norm based transitions highly penalize the transitions from the correct distribution



(a) Lena (b) Lena - Coefficient H (c) Lena - Coefficient H

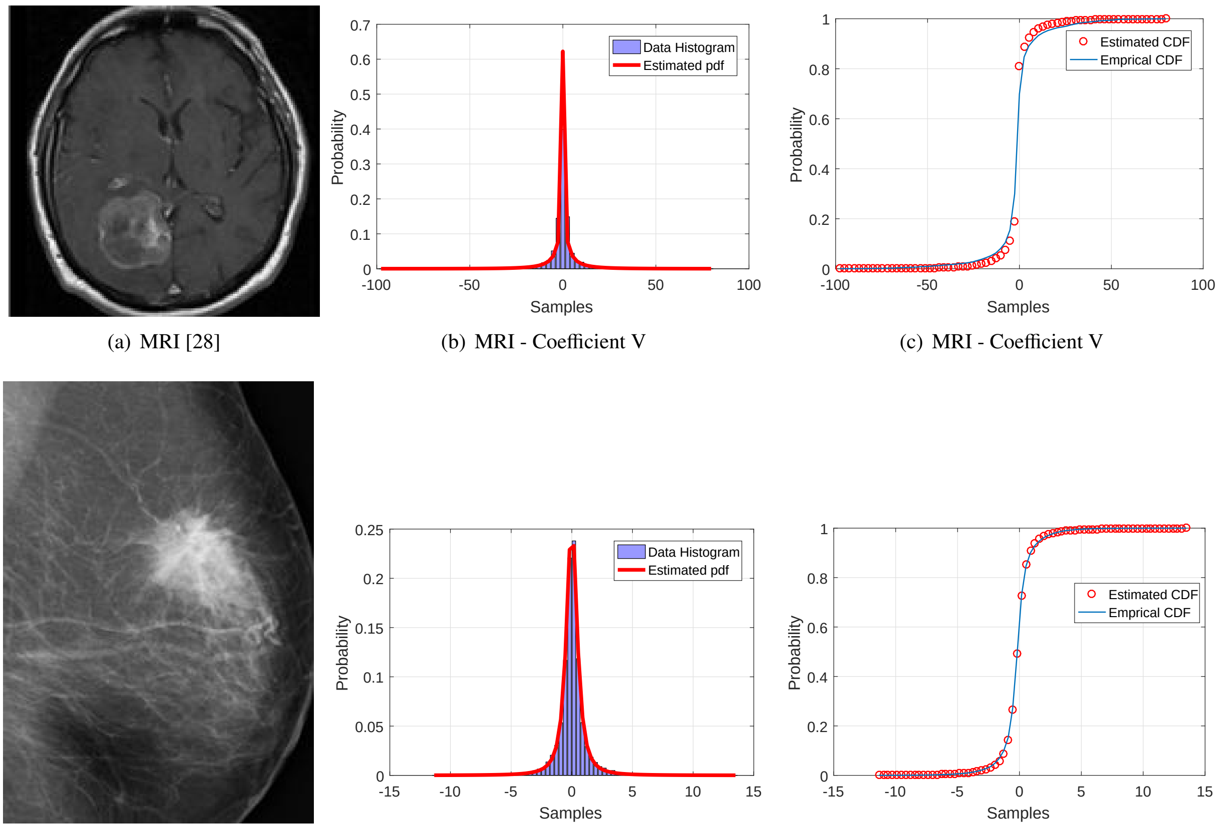
(d) SAR [27] (e) SAR - Coefficient D (f) SAR - Coefficient D

Figure 7: 2D-DWT coefficients modeling results for Lena and SAR images. (a),(d): Images, (b)-(e): Estimated pdfs, (c)-(f): Estimated CDFs.

family to another in the acceptance ratio terms. Although these kinds of transitions were somehow performed in some of the simulation cases, algorithm came back to the correct family after a low number of iterations and performed updates in the correct family. This results can be easily seen in Figures 3-(a) to (c), 4-(a) and (c) and 9-(a) and (c).

5. Conclusion

In this study, we have provided a new usage named as trans-space RJMCMC and draw attention to the generality of RJMCMC algorithm beyond the framework of trans-dimensional sampling. By defining a new combined parameter space of current and target parameter subspaces of possibly different classes or structures, we have shown that the original formulation of RJMCMC offers more general applications than just estimating the model order. This provides users to do model selection between different classes or structures. In particular, exploring solution spaces of linear and nonlinear models or of various distribution families is possible using RJMCMC. 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 RJMCMC considered in this paper have been the common distributions in the impulsive noise families. They made it possible to use the mapping functions benefiting from the FLOMs of the observed data. These functions in turn have enabled



(a) MRI [28] (b) MRI - Coefficient V (c) MRI - Coefficient V

(d) Mammogram [29] (e) Mammogram - Coefficient D (f) Mammogram - Coefficient D

Figure 8: 2D-DWT coefficients modeling results for MRI and Mammogram. (a),(d): Images, (b)-(e): Estimated pdfs, (c)-(f): Estimated CDFs.

to transfer the information learned while searching in one family to the subsequent search after an inter-class-switch move.

Candidate distribution space covers various impulsive densities from three popular families, namely SαS, GG and Student's t. In both synthetically generated noise processes and real PLC noise measurements and wavelet transforms of images, the proposed usage of RJMCMC shows remarkable performance in modeling. 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 generally close to 1.0000 (at least 0.85) for all the example data sets. Moreover, the algorithm indicated SαS distributions for 2D-DWT coefficients of SAR images and noise on PLC channels which is in accordance with the other studies in the literature and confirms the success of the algorithm.

The proposed approach for proposal distributions, FLOM-based proposals, also make it possible to perform transitions between distributions in different families which have similar statistical characteristics easily, even if they have very different values for scale and shape parameters. In other words, matching the FLOMs to calculate the parame-

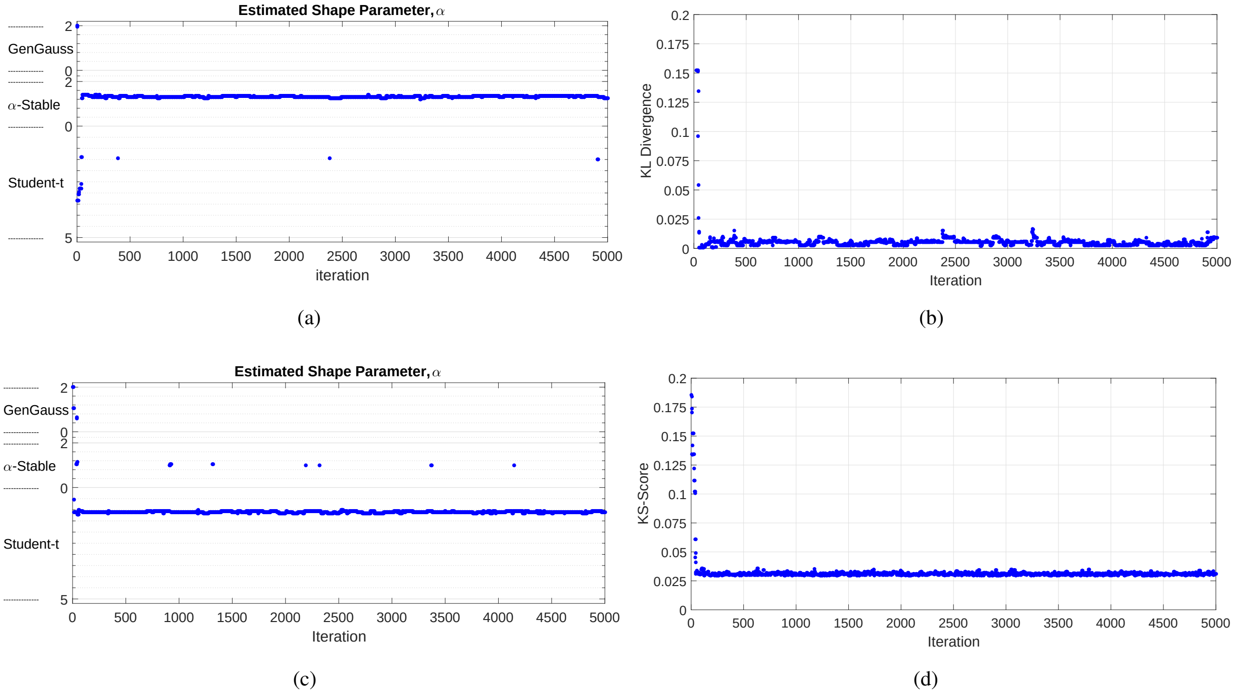


Figure 9: (a) and (c) refer to the instantaneous shape parameter estimation plots. (b) and (d) refer to the instantaneous KS (or KL) statistics plots. Results are for PLC-3 and Lena-H in the first and second rows, respectively.

ters, offers to switch distributions the parameters of which are strictly different. For further studies, this approach has possibility to open research directions to perform simulation studies about the mimicking capabilities of a distribution to another.

We would like to underline that the ideas presented in this paper are not limited only to sampling across distribution families but can be extended to any class of models.

Acknowledgement

Oktay Karakuş is funded as a visiting scholar at ISTI-CNR, Pisa, Italy by The Scientific and Technological Research Council of Turkey (TUBITAK) under grant program 2214/A.

[1] P. J. Green, Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika 82 (4) (1995) 711–732. doi:10.1093/biomet/82.4.711.

[2] P. T. Troughton, S. J. Godsill, A reversible jump sampler for autoregressive time series, in: Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, Vol. 4, IEEE, 1998, pp. 2257–2260.

[3] R. S. Ehlers, S. P. Brooks, Bayesian analysis of order uncertainty in ARIMA models, Federal University of Parana, Tech. Rep.

[4] E. Eğri, S. Günay, et al., Bayesian model selection in ARFIMA models, Expert Systems with Applications 37 (12) (2010) 8359–8364.

[5] S. Richardson, P. J. Green, 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) (1997) 731–792.

[6] V. Viallefont, S. Richardson, P. J. Green, Bayesian analysis of Poisson mixtures, Journal of nonparametric statistics 14 (1-2) (2002) 181–202.

[7] D. Salas-Gonzalez, E. E. Kuruoglu, D. P. Ruiz, Finite mixture of α-stable distributions, Digital Signal Processing 19 (2) (2009) 250–264.

[8] O. Karakuş, E. E. Kuruoğlu, M. A. Altınkaya, Estimation of the nonlinearity degree for polynomial autoregressive processes with RJMCMC, in: 23rd European Signal Processing Conference (EUSIPCO), IEEE, 2015, pp. 953–957.

[9] O. Karakuş, E. E. Kuruoğlu, M. A. Altınkaya, Bayesian estimation of polynomial moving average models with unknown degree of nonlinearity, in: 24th European Signal Processing Conference (EUSIPCO), IEEE, 2016, pp. 1543–1547.

[10] O. Karakuş, E. E. Kuruoğlu, M. A. Altınkaya, Nonlinear Model Selection for PARMA Processes Using RJMCMC, in: 25th European Signal Processing Conference (EUSIPCO), IEEE, 2017, pp. 2110–2114.

[11] O. Karakuş, E. E. Kuruoğlu, M. A. Altınkaya, Bayesian Volterra System Identification Using Reversible Jump MCMC Algorithm, Signal Processing 141 (2017) 125–136.

[12] L. Tierney, Markov chains for exploring posterior distributions, Annals of Statistics 22 (1994) 1701–1762.

[13] B. P. Carlin, S. Chib, Bayesian model choice via Markov chain Monte Carlo methods, Journal of the Royal Statistical Society. Series B (Methodological) (1995) 473–484.

[14] P. Dellaportas, J. J. Forster, I. Ntzoufras, On Bayesian model and variable selection using MCMC, Statistics and Computing 12 (1) (2002) 27–36.

[15] S. J. Godsill, On the relationship between Markov chain Monte Carlo methods for model uncertainty, Journal of computational and graphical statistics 10 (2) (2001) 230–248.

[16] S. A. Bhatti, Q. Shan, I. A. Glover, R. Atkinson, I. E. Portugues, P. J. Moore, R. Rutherford, Impulsive noise modelling and prediction of its impact on the performance of WLAN receiver, in: Signal Processing Conference, 2009 17th European, IEEE, 2009, pp. 1680–1684.

[17] K. L. Blackard, T. S. Rappaport, C. W. Bostian, Measurements and models of radio frequency impulsive noise for indoor wireless communications, IEEE Journal on selected areas in communications 11 (7) (1993) 991–1001.

[18] J. Lin, M. Nassar, B. L. Evans, Impulsive noise mitigation in powerline communications using sparse Bayesian learning, IEEE Journal on Selected Areas in Communications 31 (7) (2013) 1172–1183.

[19] E. Alsusa, K. M. Rabie, Dynamic peak-based threshold estimation method for mitigating impulsive noise in power-line communication systems, IEEE Transactions on Power Delivery 28 (4) (2013) 2201–2208.

[20] T. Y. Al-Naffouri, A. A. Quadeer, G. Caire, Impulsive noise estimation and cancellation in DSL using orthogonal clustering, in: Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on, IEEE, 2011, pp. 2841–2845.

[21] R. Fantacci, A. Tani, D. Tarchi, Impulse noise mitigation techniques for xDSL systems in a real environment, IEEE Transactions on Consumer Electronics 56 (4) (2010) 2106–2114.

[22] E. P. Simoncelli, Statistical models for images: Compression, restoration and synthesis, in: Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on, Vol. 1, IEEE Computer Society, 1997, pp. 673–678.

[23] A. Achim, P. Tsakalides, A. Bezerianos, SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling, IEEE Transactions on Geoscience and Remote Sensing 41 (8) (2003) 1773–1784.

[24] B. Yue, Z. Peng, A validation study of α-stable distribution characteristic for seismic data, Signal Processing 106 (2015) 1–9.

[25] J. A. Cortes, L. Diez, F. J. Canete, J. J. Sanchez-Martinez, Analysis of the indoor broadband power-line noise scenario, IEEE Transactions on electromagnetic compatibility 52 (4) (2010) 849–858.

[26] P. A. Lopes, J. M. Pinto, J. B. Gerald, Dealing with unknown impedance and impulsive noise in the power-line communications channel, IEEE Transactions on power delivery 28 (1) (2013) 58–66.

[27] Artemis, SAR solutions, image samples, http://artemisinc.net/media.php (2017).

[28] MRI image of brain with gadolinium contrast showing enhancing mass in the right, http://mri-scan-img.info/mri-image-of-brain-with-gadolinium-contrast-showing-enhancing-mass-in-the-right (2017).

[29] C. Martinez Lara, M. Martin Perez, I. Martin Garcia, R. Blanco Hernández, B. Sánchez Sánchez, J. Sevillano Sánchez, Radiological findings invasive lobular carcinoma, in: European Congress of Radiology (ECR), 2012, pp. C–1062. doi:10.1594/ecr2012/C-1062.

[30] L. Knorr-Held, G. Raßer, Bayesian detection of clusters and discontinuities in disease maps, Biometrics 56 (1) (2000) 13–21.

[31] D. J. Lunn, N. Best, J. C. Whittaker, Generic reversible jump MCMC using graphical models, Statistics and Computing 19 (4) (2009) 395–408.

[32] P. Dellaportas, J. J. Forster, Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models, Biometrika 86 (3) (1999) 615–633.

[33] F. Van Der Meulen, M. Schauer, H. Van Zanten, Reversible jump MCMC for nonparametric drift estimation for diffusion processes, Computational Statistics & Data Analysis 71 (2014) 615–632.

[34] B. Rannala, Z. Yang, Improved reversible jump algorithms for Bayesian species delimitation, Genetics 194 (1) (2013) 245–253.

[35] C. S. Oedekoven, R. King, S. T. Buckland, M. L. Mackenzie, K. Evans, L. Burger, Using hierarchical centering to facilitate a reversible jump MCMC algorithm for random effects models, Computational Statistics & Data Analysis 98 (2016) 79–90.

[36] O. Roeth, D. Zaum, C. Brenner, 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.

[37] W. Hastings, Monte carlo samping methods using markov chains and their applications, Biometrika 57 (1970) 97–109. doi:10.1093/biomet/57.1.97.

[38] G. Laguna-Sanchez, M. Lopez-Guerrero, On the use of alpha-stable distributions in noise modeling for PLC, IEEE Transactions on Power Delivery 30 (4) (2015) 1863–1870.

[39] E. E. Kuruoglu, W. J. Fitzgerald, P. J. Rayner, Near optimal detection of signals in impulsive noise modeled with a symmetric α-stable distribution, IEEE Communications Letters 2 (10) (1998) 282–284.

[40] H. Sadreazami, M. O. Ahmad, M. S. Swamy, A study of multiplicative watermark detection in the contourlet domain using alpha-stable distributions, IEEE Transactions on Image Processing 23 (10) (2014) 4348–4360.

[41] N. Farsad, W. Guo, C.-B. Chae, A. Eckford, Stable distributions as noise models for molecular communication, in: Global Communications Conference (GLOBECOM), 2015 IEEE, IEEE, 2015, pp. 1–6.

[42] G. Tzagkarakis, P. Tsakalides, Greedy sparse reconstruction of non-negative signals using symmetric alpha-stable distributions, in: Signal Processing Conference, 2010 18th European, IEEE, 2010, pp. 417–421.

[43] J. Nolan, Bibliography on stable distributions, processes and related topics, Tech. rep., Technical report (2010).

[44] M. N. Do, M. Vetterli, Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance, IEEE transactions on image processing 11 (2) (2002) 146–158.

[45] C. Bouman, K. Sauer, A generalized Gaussian image model for edge-preserving MAP estimation, IEEE Transactions on Image Processing 2 (3) (1993) 296–310.

[46] G. Verdoolaege, P. Scheunders, Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination, International Journal of Computer Vision 95 (3) (2011) 265–286.

[47] S. Le Cam, A. Belghith, C. Collet, F. Salzenstein, Wheezing sounds detection using multivariate generalized Gaussian distributions, in: Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, IEEE, 2009, pp. 541–544.

[48] M. Novey, T. Adali, A. Roy, A complex generalized Gaussian distribution—Characterization, generation, and estimation, IEEE Transactions

on Signal Processing 58 (3) (2010) 1427–1433.

[49] A. J. Patton, Modelling asymmetric exchange rate dependence, International economic review 47 (2) (2006) 527–556.

[50] R. F. Engle, T. Bollerslev, Modelling the persistence of conditional variances, Econometric reviews 5 (1) (1986) 1–50.

[51] A. Aravkin, T. Van Leeuwen, F. Herrmann, Robust full-waveform inversion using the Student's t-distribution, in: SEG Technical Program Expanded Abstracts 2011, Society of Exploration Geophysicists, 2011, pp. 2669–2673.

[52] Y. Liang, G. Chen, S. Naqvi, J. A. Chambers, Independent vector analysis with multivariate student's t-distribution source prior for speech separation, Electronics Letters 49 (16) (2013) 1035–1036.

[53] T. M. Nguyen, Q. J. Wu, Robust student's-t mixture model with spatial constraints and its application in medical image segmentation, IEEE Transactions on Medical Imaging 31 (1) (2012) 103–116.

[54] Z. Zhang, K. Lai, Z. Lu, X. Tong, Bayesian inference and application of robust growth curve models using student's t distribution, Structural Equation Modeling: A Multidisciplinary Journal 20 (1) (2013) 47–78.

[55] D. I. Hastie, P. J. Green, Model choice using reversible jump Markov chain Monte Carlo, Statistica Neerlandica 66 (3) (2012) 309–338.

[56] R. J. Barker, W. A. Link, Bayesian multimodel inference by RJMCMC: A Gibbs sampling approach, The American Statistician 67 (3) (2013) 150–156.

[57] G. A. Tsihrintzis, C. L. Nikias, Fast estimation of the parameters of alpha-stable impulsive interference, IEEE Transactions on Signal Processing 44 (6) (1996) 1492–1503.

[58] X. Ma, C. L. Nikias, Parameter estimation and blind channel identification in impulsive signal environments, IEEE transactions on signal processing 43 (12) (1995) 2884–2897.

[59] E. E. Kuruoglu, Density parameter estimation of skewed α-stable distributions, IEEE transactions on signal processing 49 (10) (2001) 2192–2201.

[60] D. J. MacKay, Information theory, inference and learning algorithms, Cambridge university press, 2003.

[61] F. J. Massey Jr, The Kolmogorov-Smirnov test for goodness of fit, Journal of the American statistical Association 46 (253) (1951) 68–78.

[62] N. Andreadou, F.-N. Pavlidou, Modeling the noise on the OFDM power-line communications system, IEEE Transactions on Power Delivery 25 (1) (2010) 150–157.

[63] T. H. Tran, D. D. Do, T. H. Huynh, PLC impulsive noise in industrial zone: measurement and characterization, International Journal of Computer and Electrical Engineering 5 (1) (2013) 48.