# SEQUENTIAL BAYESIAN MODELING OF NON-STATIONARY NON-GAUSSIAN PROCESSES

by

Orhan Deniz Gençağa

B.S., Electronics and Communications Engineering, Yıldız Technical University, 1997

M.S., Electrical and Electronic Engineering, Boğaziçi University, 2000

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Electrical and Electronics Engineering
Boğaziçi University
2007

*To my family*

# ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Ayşın Baytan Ertüzün and co-supervisor Assoc. Prof. Ercan Engin Kuruoğlu for their supervision, guidance and help throughout this thesis. The words are not enough to thank for their continuous efforts and encouragement. I would like to thank both coordinators, Assoc. Prof. Ercan E. Kuruoğlu and Prof. Bülent Sankur, of our joint research project between The Scientific and Technological Research Council of Turkey (TÜBİTAK) and Consiglio Nazionale delle Ricerche (CNR) and I would like to thank the coordinator, Prof. Ayşın Ertüzün, of the project supported by the Boğaziçi University Scientific Research Fund. My first research at CNR would not have been possible without the financial support of the TÜBİTAK-CNR joint project. I would also like to express my gratitude to TÜBİTAK and North Atlantic Treaty Organization (NATO), for awarding me the NATO-TÜBİTAK A2 fellowship which enabled me to continue my research for another five months at CNR, Italy. I gratefully acknowledge the hospitality and support of Assoc. Prof. Ercan E. Kuruoğlu, Dr. Anna Tonazzini, Dr. Luigi Bedini, Dr. Emanuele Salerno and Dr. Diego Herranz during my research at CNR. Moreover, I would like to thank Mauro Costagli and Diego Salas-Gonzalez for their invaluable discussions and friendship. My participation in international conferences would not have been possible without the financial support of the Boğaziçi University Foundation. I would like to express my gratitude to the jury members of my thesis committee, Prof. Ayşın Baytan Ertüzün, Assoc. Prof. Ercan Engin Kuruoğlu, Prof. Lale Akarun, Assist. Prof. F. Kerem Harmancı, Prof. Bülent Sankur, for their suggestions. I would also like to thank Assist. Prof. Kevin H. Knuth for motivating me. This dissertation is a consequence of the continuous support of my family throughout my education life. So, I would like to express my gratitude to my family and especially to my parents to whom I am indebted for all I have. Finally, I would like to thank all my friends at Boğaziçi University for their friendship and discussions.

This thesis has been supported by TÜBİTAK-CNR project numbers: 104E101, 102E027; NATO-TÜBİTAK A2 Fellowship and Boğaziçi University Scientific Research Fund project number: BAP-04A201.

# ABSTRACT

## SEQUENTIAL BAYESIAN MODELING OF NON-STATIONARY NON-GAUSSIAN PROCESSES

This thesis brings a unifying approach for modeling non-stationary non-Gaussian signals which are widely encountered in many multidisciplinary research fields. In the literature, different approaches have been used to model non-stationary signals. However, they could not fulfill the increasing needs where non-Gaussian processes are involved until the development of Sequential Monte Carlo techniques (particle filters). In general particle filtering, the problem is expressed in terms of nonlinear and/or non-Gaussian state-space equations and we need information about the functional form of the state variations. In this thesis, we bring a general solution for cases where these variations are unknown and the process distributions cannot be expressed by a closed form probability density function. We propose a novel modeling scheme which is as unified as possible to cover these problems. First, a novel technique is proposed to model Time-Varying Autoregressive Alpha Stable processes where unknown, time-varying autoregressive coefficients and distribution parameters can be estimated. Successful performances have been supported by posterior Cramer Rao Lower Bound values. Next, we extend our methodology to model cross-correlated signals where vector autoregressive processes with non-Gaussian driving signals can also be modeled. Later, this extension is used as a building block to provide a more unifying solution where both mixing matrix and latent processes are modeled from their mixtures. This can be interpreted as a solution for non-stationary Dependent Component Analysis. Successful simulation results verify that our methodology is very flexible and provides a unifying solution for the modeling of non-stationary processes in all cases described above.

# ÖZET

## DURAĞAN-OLMAYAN GAUSS-DIŞI DAĞILIMLI SÜREÇLERİN BAYESÇİ YAKLAŞIMLARLA MODELLENMESİ

Bu tez, birçok disiplinde sıkça karşılaşılan durağan-olmayan Gauss-dışı dağılımlı süreçlerin modellenmesi için genel bir yaklaşım sunmaktadır. Literatürde durağan-olmayan süreçlerin modellenebilmesi için birçok farklı yöntem kullanılmıştır. Ancak, Ardışıl Monte Carlo (Parçacık Süzgeçleri) metodlarının geliştirilmesine kadar bahsi geçen yöntemler Gauss-dışı süreçlerin yer aldığı artan modelleme talebine karşı yetersiz kalmışlardır. Genel parçacık süzgeçlerinde, problem doğrusal olmayan ve/veya Gauss-dışı durum-uzay denklemleri ile ifade edilir ve bizim durum geçişleri hakkında işlevsel bilgiye ihtiyacımız vardır. Bu tezde, bu geçişlerin bilinmediği ve süreç dağılımlarının kapalı forma sahip olasılık yoğunluk işlevleri tarafından ifade edilemediği problemler için genel bir çözüm getirilmektedir. Burada, bu problemlerin hepsini kapsayabilecek ve mümkün olduğu kadar genel, yeni bir modelleme yöntemi öne sürülmektedir. İlk olarak, hem bilinmeyen hem de zamanla değişen öz-bağlanım katsayıları ve dağılım parametrelerine sahip Zamanla-Değişen Öz-bağlanımlı Alfa Kararlı süreçlerin modellenmesi için yeni bir teknik geliştirilmiştir. Olumlu başarım sonuçları, sonsal Cramer Rao Alt Sınırları ile desteklenmiştir. Daha sonra, Gauss-dışı süreçlerin sürdüğü vektörel öz-bağlanımlı süreçlerin de modellenebildiği öz-ilintili işaretlerin modellenebilmesi için metodumuz genelleştirilmiştir. Ardından, bu genelleştirme, karışımlarından hem karışım matrisi hem de saklı süreçlerin modellenebildiği daha genel bir çözüm için yapı-taşı olarak kullanılmıştır. Bu, durağan-olmayan Bağımlı Bileşen Analizi için bir çözüm olarak düşünülebilir. Başarılı simülasyon sonuçları metodolojimizin çok esnek olduğunu ve yukarıda anlatılan tüm problemlerdeki durağan-olmayan süreçlerin modellenebilmesi için genel bir çözüm sağladığını doğrulamıştır.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF SYMBOLS / ABBREVIATIONS

| | |
|---|---|
| $\mathbf{A}_t$ | Mixing matrix at time *t* |
| $\mathcal{B}$ | Matrix hyperparameter of the Inverted Wishart distribution probability distribution of the covariance matrix $\mathbf{\Sigma_n}$ pertaining to the observation noise |
| $\mathcal{C}$ | Augmented matrix consisting of the background mean vector and the mixing matrix |
| $\mathbf{C_0}$ | Hyperparameter of $\mathcal{C}$ |
| $d_1$ | Dimension of the observation vector $\mathbf{y}_t$ |
| $d_2$ | Dimension of the source vector $\mathbf{s}_t$ |
| **D** | Submatrix of Information Matrix used in Cramer Rao Lower Bound |
| $e_d$ | *d* x *1* dimensional vector of ones |
| E | Total number of realizations in an ensemble |
| $E[.]$ | Expectation operator |
| $f_t$ | General state-transition function of a state space representation |
| **F** | Linear state-transition function of a state space representation |
| **G** | Augmented VAR coefficient matrix |
| $h_t$ | General observation function of a state space representation |
| $\hbar$ | Differential entropy |
| **H** | Linear observation function of a state space representation |
| $\mathcal{H}$ | Hyperparameter of $\mathcal{C}$ |
| *i* | Particle index |
| $\mathbf{I_d}$ | *d* x *d* dimensional identity matrix |
| $\mathcal{I}$ | Monte Carlo integration |
| $\mathcal{IG}(\eta,\kappa)$ | Inverted Gamma distribution with hyperparameters $\eta$ and $\kappa$ |
| $\mathcal{IW}(\mathcal{V},d,\overline{\eta})$ | Inverted Wishart distribution with hyperparameters $\mathcal{V},d,\overline{\eta}$ |
| $\mathbb{I}$ | Indicator function |

| | |
|---|---|
| $j$ | $\sqrt{-1}$ |
| $\mathcal{J}$ | Negentropy |
| **J** | Information Matrix used in Cramer Rao Lower Bound |
| $k$ | Index of an Autoregressive coefficient |
| K | Order of an Autoregressive process |
| $K_t$ | Kalman Gain |
| $\mathcal{K}$ | Kullback Leibler Divergence |
| $l$ | Particle index |
| L | Total number of Gibbs sampling iterations after the burnin period |
| $m$ | Iteration index of the outer Gibbs sampling in Two-Stage Gibbs Sampling method |
| **m** | Mean vector |
| M | Total number of iterations in the outer Gibbs sampling in Two-Stage Gibbs Sampling method |
| $\mathcal{MN}$ | Matrix Normal distribution |
| $\mathbf{n}_t$ | Observation noise vector at time $t$ |
| $\tilde{\mathbf{n}}_t$ | Augmented observation noise vector at time $t$ |
| $\hat{n}_t$ | Estimate of innovations process at time $t$ |
| N | Total number of particles |
| **N** | Noise matrix consisting of $\tau$ number of ($d_1$ x $1$) dimensional noise vectors |
| $\mathcal{N}$ | Gaussian distribution |
| $p$ | Probability density function |
| **P** | Covariance matrix of a Gaussian distribution |
| Pr($x$) | Probability of $x$ |
| $\mathcal{P}$ | Permutation matrix |
| $q$ | Importance function (proposal distribution) |
| **Q** | Covariance matrix of a Gaussian distributed observation noise $\mathbf{n}_t$ |
| $r$ | Iteration index of the inner Gibbs sampling in Two-Stage Gibbs Sampling method |

| | |
|---|---|
| R | Total number of iterations in the inner Gibbs sampling in Two-Stage Gibbs Sampling method |
| $\mathbf{s}_t$ | Source vector at time *t* |
| $\mathbf{S}_t$ | Predicted mean vector of a Gaussian observation process in Kalman filter |
| $\mathcal{S}$ | ($d_1$ x $\tau$) dimensional Matrix Normal distributed source matrix |
| $S_\alpha(\gamma,\beta,\mu)$ | Stable probability distribution |
| $T(x';x)$ | Transition kernel of a Markov Chain |
| $\mathcal{U}$ | Uniform distribution |
| $\mathbf{V}_t$ | Noise vector of Artificial Random Walk model for state-transitions |
| $\mathcal{V}$ | Matrix hyperparameter of the Inverted Wishart distribution probability distribution of the covariance matrix $\mathbf{\Sigma_s}$ pertaining to the source vector |
| $w_t^{(i)}$ | Importance weight of the $i^{th}$ particle at time *t* |
| $\mathbf{W}$ | Demixing matrix |
| $\mathbf{x}_t$ | State vector in particle filtering |
| $x_{k,t}$ | $k^{th}$ entry of state vector $\mathbf{x}_t$ |
| $\mathbf{Y}$ | ($d_1$ x $\tau$) dimensional observation matrix consisting of τ number of (*d* x *1*) dimensional observation vectors |
| $\mathcal{Z}$ | Augmented matrix consisting of $e_d$ and $\mathcal{S}$ |
| | |
| $\alpha$ | Characteristic Exponent parameter of a stable distribution |
| $\beta$ | Skewness parameter of a stable distribution |
| $\gamma$ | Dispersion parameter of a stable distribution |
| $\delta$ | Delta-dirac function |
| $\mathbf{\Delta}$ | Diagonal matrix |
| $\varepsilon$ | Acceptance ratio in Metropolis Hastings algorithm |
| $\zeta$ | Dummy variable in the characteristic function of a stable distribution |
| $\eta$ | Hyperparameter of an Inverted Gamma distribution |
| $\boldsymbol{\theta}$ | Distribution parameter vector in the Two-Stage Gibbs Sampler |
| $\kappa$ | Hyperparameter of an Inverted Gamma distribution |

| | |
|---|---|
| $\lambda$ | Positive stable distributed auxiliary random variable sample |
| $\Lambda$ | Positive stable distributed auxiliary random variable |
| $\mu$ | Location parameter of a stable distribution |
| $\xi$ | Forgetting factor |
| $\rho$ | Correlation coefficient |
| $\sigma_k^2$ | Elements of covariance matrix $\mathbf{\Sigma}_{\mathbf{V}_t}$ |
| $\mathbf{\Sigma}_{\mathbf{n}_t}$ | Covariance matrix of noise vector $\mathbf{n}_t$ |
| $\mathbf{\Sigma}_{\mathbf{s}_t}$ | Covariance matrix of source vector $\mathbf{s}_t$ |
| $\mathbf{\Sigma}_{\mathbf{V}_t}$ | Covariance matrix of $\mathbf{V}_t$ |
| $\phi_{k,t}$ | $k^{th}$ order autoregressive coefficient of a univariate Autoregressive process |
| $\mathbf{\Phi}_{k,t}$ | $k^{th}$ order autoregressive coefficient matrix of a Vector Autoregressive process |

| | |
|---|---|
| APF | Auxiliary Particle Filter |
| AR | Autoregressive |
| ARMA | Autoregressive Moving Average |
| CLT | Central Limit Theorem |
| CRLB | Cramer Rao Lower Bound |
| DBN | Dynamic Bayesian Network |
| DCA | Dependent Component Analysis |
| DSMC | Direct Sequential Monte Carlo method |
| fMRI | Functional Magnetic Resonance Imagery |
| EKF | Extended Kalman Filter |
| EM | Expectation Maximization |
| HMC | Hybrid Monte Carlo method |
| HSMC | Hybrid Sequential Monte Carlo method |
| i.i.d. | Independent Identically Distributed |
| ICA | Independent Component Analysis |
| IFA | Independent Factor Analysis |
| IFT | Inverse Fourier Transform |

| | |
|---|---|
| IRLS | Iteratively Reweighted Least Squares |
| IS | Importance Sampling |
| KLD | Kullback Leibler Divergence |
| kurt($y$) | Kurtosis of $y$ |
| LS | Least Squares |
| MA | Moving Average |
| MAP | Maximum a Posteriori |
| MCARS | Mixture Modeling of Cross-Correlated AR Sources |
| MCMC | Markov Chain Monte Carlo |
| MCS | Mixture Modeling of Cross-Correlated Sources |
| MH | Metropolis-Hastings |
| ML | Maximum Likelihood |
| MMSE | Minimum Mean Squared Error |
| NMSE | Normalized Mean Squared Error |
| PCRLB | Posterior Cramer Rao Lower Bound |
| RLS | Recursive Least Squares |
| RS | Rejection Sampling |
| SIS | Sequential Importance Sampling |
| SMC | Sequential Monte Carlo |
| SOBI | Second Order Blind System Identification |
| SSMC | Single Sequential Monte Carlo method |
| $S\alpha S$ | Symmetric Alpha Stable |
| TVAR | Time-varying autoregressive |
| UKF | Unscented Kalman Filter |
| VAR | Vector Autoregressive |
| $\alpha$ *stable* | Alpha Stable |

# 1. INTRODUCTION

Modeling of non-stationary signals constitutes a very important part of modern statistical signal processing. These signals are widely encountered in many multidisciplinary research fields such as telecommunications, astrophysics, biomedicine, seismology, geophysics and finance. In telecommunications, recent developments in mobile communications have led to an increasing need for non-stationary signal processing. It is well known that, due to the mobility of the users, time-varying channel characteristics are involved in the processing of communication signals which cannot be modeled to be stationary in their most general form (Rappaport, 2001). A similar scenario is also valid in the cocktail party problem which is a typical example of a blind source separation application (Haykin, 2000). In this case, different moving source signals (speaking people in the cocktail party problem) are separated from their mixtures. Here, the mobility of sources causes them to be non-stationary. Another application area is the separation of astrophysical sources in astrophysical maps (Kuruoğlu *et al*., 2003). In this problem, some latent sources are observed in mixtures which are also corrupted by an additive noise. It is well known that this noise component has a spatially varying nature which causes the need for the implementation of non-stationary signal processing techniques. Moreover, speech and audio signal enhancement is also an example where non-stationarity arises due to the time varying autoregressive (TVAR) nature of speech signals (Andrieu and Godsill, 2000; Vermaak *et al.,* 2002). In addition to these; system identification, echo cancellation and blind deconvolution are just a few of many more application areas where non-stationary signal processing is needed (Haykin, 1996). Despite the richness of potential application areas, existing techniques for non-stationary signal processing cannot be considered to fulfill the needs in the literature.

Most of these non-stationary problems have been addressed using Kalman filtering and its variants until the beginning of last decade (Kalman, 1960; Van Trees, 1968). It is well known that Kalman filter gives the optimal recursive Bayesian solution provided that the non-stationary problem can be represented by a linear, state-space form where the process and the observation noises are additive and Gaussian distributed (Van Trees, 1968;

Shanmugan and Breihpohl, 1988; Kailath, 2000). Recursive Least Squares (RLS) is a variant of Kalman filtering and it is also optimal only for Gaussian processes (Haykin, 1996).

If the system functions are nonlinear, Extended Kalman Filter (EKF) (Anderson and Moore, 1979) and recently Unscented Kalman Filter (UKF) (Julier and Uhlmann, 1996; Wan and van der Merwe, 2000) have been widely used in literature to overcome the nonlinearity problem. However, the required probability density function (pdf) of the state vector is approximated by Gaussian distributions in both of these methods.

On the other hand, solving Gaussian nonlinear problems is not sufficient in physical world, since non-Gaussian signals are involved in most cases, such as the application areas mentioned above: Most of the mobile communication signals possess non-Gaussian distributions due to the fading phenomenon (Rappaport, 2001). In astrophysical source separation, some of the source signals are distributed as Mixture of Gaussians, such as the galactic dust and synchrotron emissions (Kuruoğlu *et al*., 2003). In literature, useful techniques have been developed to handle non-Gaussian signals. Some of these include Maximum Likelihood (ML) (Van Trees, 1968) and Bayesian approaches (Gelman *et al*., 1995; Tanner, 1996; Gilks *et al*., 1998; Robert and Casella, 1999; MacKay, 2003).

As outlined above, different methodologies have been proposed to handle stationary, non-stationary; Gaussian, non-Gaussian; linear, nonlinear or different combinations of these characteristics according to the problem. However, they could not be covered under one umbrella until the development of Sequential Monte Carlo techniques which are also known as the particle filters. In particle filtering, the problem is expressed in terms of state-space equations where the linearity and Gaussianity requirements of the Kalman filtering are generalized. However, we need knowledge about the functional form of the state variations. In this thesis, first, we bring solution for more general cases where these variations are unknown and the process distributions cannot be expressed by any closed form pdf's. Here, we propose a novel modeling scheme which is as unified as possible to cover all these problems. We study our unifying particle filtering methodology to model non-stationary alpha stable ($\alpha$-stable) processes. It is well known that the pdf's of these processes cannot be expressed in closed form, except for some limited number of cases.

Moreover, this distribution family presents a direct generalization from Gaussian to non-Gaussian distributions, since they have common properties, such as the stability property and the Central Limit Theorem. In order to model time structures of such non-Gaussian processes, we propose using linear autoregressions which is commonly preferred due to its wide application areas (Hamilton, 1994). Moreover, by imposing time-variations on autoregressive (AR) coefficients and distribution parameters of the $\alpha$-stable driving process, we generalize our model so that non-stationary $\alpha$-stable processes can also be modeled. To the best of our knowledge, this is the first general method which can model TVAR $\alpha$-stable processes where the distribution parameters are also time-varying in addition to the AR coefficients. According to the computer simulations, our method provides very successful results for the modeling of non-stationary non-Gaussian processes.

Secondly, we extend our technique so that cross-correlations between different signals can also be modeled. It is known that speech signals in a cocktail party problem are not statistically independent and they show cross-correlations among them (Rowe, 2003). In another reference, relationship of different chemical processes has been studied by vector autoregressive (VAR) representations (Hsu, 1997). VAR representation is known as the multivariate generalization of univariate AR proceses, where the current value of a process depends both on the past values of itself and other processes (Hamilton, 1994; Lütkepohl, 1993). However, these references merely consider stationary cross-correlations between the processes. On the other hand, in physical world, correlation information is indeed non-stationary. For example, in cocktail party problem, such non-stationary correlations arise due to the mobility of the speakers. A similar case is also encountered in mobile communications. Recently, mobile communication channels have been modeled by non-stationary VAR structures (Jachan and Matz, 2005).

After presenting an extension of our methodology which is capable of modeling cross-correlated processes, we propose another solution for cases where modeling of mixtures of cross-correlated processes is needed. This is even a more challenging problem since the estimation of cross-correlated processes needs to be done by using their mixtures, i.e. both the mixing mechanism and the latent cross-correlated processes are to be modeled. Astrophysical source separation is a typical example for such a problem where cross-correlated galactic dust and synchrotron emissions are among the sources to be separated

by using their mixtures (Kuruoğlu *et al*., 2003; Bedini *et al*., 2005). Separation of cross-correlated speech signals in a cocktail party problem constitutes another application area (Rowe, 2003). In this thesis, we show that our method can also be used successfully in the modeling of non-stationary cross-correlated processes and their mixtures as well. Here, we propose a novel solution for the separation of non-stationary mixtures of cross-correlated processes which can be interpreted as Dependent Component Analysis (DCA) problem. Even though statistical dependencies between different processes are modeled in terms of cross-correlations for the moment, our methodology opens a new and realistic research direction in handling the dependent source separation problem which has been widely approximated so far by Independent Component Analysis (ICA) techniques (Hyvarinen *et al*., 2001). Therefore, our methodology serves as an important alternative to ICA approaches by avoiding the assumption of statistical independency.

In the following section, we present a comprehensive literature survey on the previous studies performed in the literature. This is followed by the main contributions and the scope of thesis.

### 1.1. Literature Survey

In nature, non-Gaussian distributed signals are encountered widely. However, most of the time, these signals are approximated by Gaussian distributions due to the modeling and processing simplifications of such an approximation (Van Trees, 1968; Haykin, 1996). A direct generalization of Gaussian distributions to non-Gaussian distributions has been developed by the introduction of $\alpha$-stable modeling of heavy-tailed signals (Zolotarev, 1989; Samorodnitsky and Taqqu, 1994; Nikias and Shao, 1995). This family of distributions has been extensively utilized in the modeling of impulsive signals which are widely witnessed in many areas such as radar and sonar communications, financial time-series modeling, telecommunications and teletraffic modeling (Nolan and Swami, 1999). These impulsive signals possess outliers which cause the most extreme effects, when the physical nature of the data is considered, such as cases in geophysical, climatologic and oceanographic data. A remarkable example for this discussion can be the modeling of oceanic wave heights, where these are generally modeled by heavy-tailed distributions (Pierce, 1997). During signal modeling, neglecting the outlier data could cause some

catastrophic consequences, such as the case that was witnessed in Holland in 1953, where an increase of 3.5m in the sea level led to the death of approximately 2000 people.

Another important issue in the modeling of heavy-tailed data is expressing its temporal dependency. Similar to Gaussian cases, linear and nonlinear parametric temporal modeling of impulsive data can be expressed in terms of AR (Kuruoğlu *et al*., 1997), moving average (MA), autoregressive moving average (ARMA) (Davis *et al*., 1992; Samorodnitsky and Taqqu, 1994; Nikias and Shao, 1995) or nonlinear AR models (Kuruoğlu, 2002). Among these, AR structure is widely used in teletraffic data modeling in computer communications (Resnick, 1997). Moreover, time evolution of these data model parameters is also an active research area. According to the observations, there is an ongoing debate about time dependency of teletraffic data, where it is stated that AR coefficients may have a time-varying nature (Bates and McLaughlin, 1997). In literature, several techniques have been developed to model *time-invariant* AR *α-stable* processes which are outlined as follows:

A generalized version of the Yule-Walker equations have been derived to estimate the AR model coefficients for symmetric alpha stable ( $S\alpha S$ ) processes (Kanter and Steiger, 1974). However, the performance of this method drops significantly when small numbers of data are available.

Later, this shortcoming of the above technique has been overcome by the iteratively reweighted least squares (IRLS) approach proposed in (Kuruoğlu *et al*., 1997). In this method, AR model coefficients of a $S\alpha S$ process have been estimated successfully even for shorter data lengths by using an $l_p$ - norm minimization problem.

By the utilization of the Bayesian methodology, Godsill and Kuruoğlu proposed a Markov Chain Monte Carlo (MCMC) method to estimate the AR model coefficients of a $S\alpha S$ process (Godsill and Kuruoğlu, 1999). In addition to its effectiveness, the flexibility of this approach makes it more superior to its predecessors.

AR coefficients have been modeled to be time invariant in all methods discussed above. However, a very *limited* number of works have been done to model TVAR *α-stable*

processes which can be encountered as the most general case. To the best of our knowledge, the only approach for such a modeling has been proposed in (Thavaneswaran and Peiris, 1999). In this technique, a penalized minimum dispersion method was proposed in case of a presumably *known* shape parameter $\alpha$ of the *α-stable* process which was taken to be larger than one $(\alpha > 1)$. However, the distribution parameters have been assumed to be known in this approach and therefore they have not been estimated.

In this thesis, we will focus on particle filtering to provide a unifying modeling scheme where both the *unknown* TVAR coefficients and the distribution parameters of an *α-stable* process can be estimated. With the increase in computational capabilities, particle filters have been proposed to approximate the optimal recursive Bayesian solution in the case of non-stationary non-Gaussian problems (Gordon, 1993; Doucet *et al*. 2000; Doucet *et al*. 2001;). In general particle filtering framework, the problem is expressed in terms of nonlinear and /or non-Gaussian state-space representation. Generally, functional form of the state-transition equation needs to be known (Doucet *et al*. 2001).

Due to its eligibility to process non-stationary non-Gaussian signals, particle filtering has been proposed for the enhancement of a TVAR Gaussian signal which was embedded in a $S\alpha S$ noise process (Lombardi and Godsill, 2004). Despite its satisfactory performance, this algorithm is not very flexible since the modeling of a *Gaussian* signal is performed constituting only a subset of a more general class of non-Gaussian distributions. This is the main shortcoming of this algorithm beside the fact that it can only be used provided that we have full information regarding the time variations of the AR coefficients, i.e. state-transition equation.

When there is no a priori information about the time variations of the unknown parameters, a random walk with discounting of old measurements has been proposed to model the unknown state-transition equation (Djuric *et al*., 2001; Djuric *et al*., 2002). Here, modeling of TVAR non-Gaussian signals have been performed within the particle filtering scheme. Although successful performances has been obtained, only non-Gaussian distributions having closed form pdf's, such as the Laplacian or Mixture of Gaussian distributions have been used with *known* statistical parameters. That is why, despite its

success, this methodology does not provide a flexible scheme for a generalization to a non-Gaussian case whose pdf cannot be expressed in a closed form.

In addition to various approaches used in the modeling of non-stationary and/or non-Gaussian processes, different techniques have also been developed to study cross-correlated signals as well. (Hsu, 1997) used least squares (LS) approach to estimate the parameters of the VAR structure which was used to model relationships among chemical processes. In other application areas, the need for a time-varying VAR modeling has arisen, which cannot be solved by a batch natured technique. To overcome this difficulty, (Möller *et al*., 2001) proposed a modified RLS technique to fit a VAR model with time-dependent parameters to multidimensional signals such as multichannel EEG data. Later, (Sato *et al*., 2006) proposed another method to estimate time-varying VAR processes using wavelet expansions of AR coefficients. This technique has been applied to functional magnetic resonance imaging (fMRI) with success (Sato *et al*., 2006). In (Jachan and Matz, 2005), mobile communication channels have been modeled as VAR processes and a modified Yule-Walker approach has been proposed to model channel characteristics. Despite its innovation in the field, this approach is not practical in the physical world since channel realizations cannot be directly observed. Moreover, in all of these works, distribution of the driving processes has been taken to be Gaussian distributed. Therefore, even if the non-stationarity case has been handled by different approaches, modeling cross-correlated non-Gaussian processes is still an open research area.

A similar situation arises in the field of source separation, where mixtures of latent processes are observed. In literature, many techniques have been developed to separate sources from their mixtures (Haykin, 2000). However, separation of cross-correlated (dependent) sources has been rarely studied in literature contrary to the widely known ICA approaches (Hyvarinen *et al*., 2001). Many techniques have been proposed in the ICA literature such as the “Natural Gradient Method” (Amari, 1998), “Relative Gradient Technique” (Cardoso and Laheld, 1996) and the “FastICA” algorithm of (Hyvarinen and Oja, 2000). Despite their successful performances, all of these basic ICA techniques consider an ideal mixture where no corrupting noise is taken into account and the number of sources is taken to be equal to that of sensors. Noise is included to the problem by (Hyvarinen, 1998; Hyvarinen, 1999) using contrast functions based on Gaussian moments.

However, by using these techniques, it has been noted that separation performance degraded when the noise level was increased (Maino, *et al*., 2002).

A more flexible methodology has been proposed by (Moulines *et al*., 1997; Attias, 1999) where sources have been modeled by mixtures of Gaussians. This technique was named as Independent Factor Analysis (IFA) by Attias and provided successful results for stationary data by the utilization of Expectation Maximization (EM) based optimization algorithm. However, a major shortcoming of these techniques is that they have been designed to process stationary data. A very successful solution for non-stationary signal modeling has been provided by (Kuruoğlu *et al*., 2003) by extending IFA with simulated annealing strategy. In this research, sources have been successfully modeled by mixtures of Gaussians.

On the other hand, in order to recommend a more flexible methodology, model constraints on sources have to be removed. In most of the aforementioned ICA approaches either a fixed (or hypothetical) or an application specific model such as mixtures of Gaussians has been employed. However, if source signals can be modeled by temporal structures unlike the independent, identically distributed (i.i.d) sources in the above methods, then even Gaussian signals can be separated from their mixtures by using this temporal information (Cardoso, 2001). Two types of temporal structures have been utilized in literature. (Belouchrani *et al*., 1997) proposed a second order blind identification (SOBI) algorithm to separate stationary signals with temporal dependence, i.e. non-white processes, provided that no two sources have proportional spectra. The second case considers separation of non-stationary (possibly white) signals (Cardoso and Pham, 2001).

Another ICA approach has been proposed by (Parga and Nadal, 2000) where time varying mixing matrices are considered provided that they vary slowly and smoothly. However, in this research, sources are constrained to be stationary and ergodic which constitutes a drawback in our way to form a flexible and unifying methodology.

In addition to the above blind techniques, studies of (Djafari, 1999) and (Knuth, 1999) have led to a new direction in handling the source separation problem. In these approaches, utilization of *a priori* information about the problem has been encouraged. As a result of

the additional flexibility brought by the use of *a priori* information, particle filtering techniques have been proposed to model non-stationary mixtures of *statistically independent* non-Gaussian sources (Everson and Roberts, 2000; Andrieu and Godsill, 2000; Ahmed *et al*., 2000). Although these latter methods provide us with the most unifying modeling schemes discussed until now, they are not capable of modeling the non-stationary mixtures of *cross-correlated* (dependent) sources.

When compared to the ICA techniques, a very limited number of works have been done to separate dependent sources from their mixtures. Barros proposed an algorithm to separate dependent signals by minimizing the mutual spectral overlap of the mixture (Barros, 2000). In this approach, the averaged coherence function is suggested to be minimized. However, minimizing this cost function does not guarantee a full separation of every dependent (non-orthogonal) signal. This is the main drawback of this approach and it has been depicted in (Barros, 2000). This method is an important development for a realistic source separation since it drops the assumption of statistical independency between the sources to separate them. However, its batch type of processing nature avoids it to be used in non-stationary dependent source separation problems.

Separation of dependent processes has also been studied in (Nuzillard and Nuzillard, 1999) within the context of separating non-orthogonal signals. Here, the frequency domain representation of the mixture is divided into different zones. Then, a region holding the orthogonality condition is taken. In order to extract the sources, the inverse Fourier Transform (IFT) of the spectral components is taken, in which orthogonality holds. After IFT, time correlated data are obtained in the time domain. Then, Second Order Blind Identification (SOBI) algorithm (Belouchrani *et al*, 1997) is applied to time-correlated (but spatially uncorrelated) data to find the mixing matrix. However, this is not a flexible approach since it is only applicable provided that a spectral region possessing the orthogonality condition is available. Moreover, its batch nature avoids the utilization of this method to be used in non-stationary problems.

Bach and Jordan have proposed to separate dependent sources by a tree based algorithm where a transform of the mixture is tried to be found so that the transformed data fits to a tree-structured graphical model (Bach and Jordan, 2002). The optimal transform is

found by minimizing a contrast function and dependency of the sources are modeled by a graphical model. Despite its successful performance, this technique cannot be used for non-stationary DCA, either; due to its batch nature.

Although many novel techniques have been developed in literature in order to model the mixtures of cross-correlated (dependent) processes, it is observed that most of these methods have been developed to bring solution to a specific area of interest. Therefore, they are application specific and cannot be used in different scenarios. To overcome this difficulty, we propose a particle filtering scheme which is capable of modeling non-stationary non-Gaussian processes. Due to its generality, this model can also be used in sub-cases, such as a stationary Gaussian problem. In this thesis, our method provides a framework which is general enough to model non-stationary cross-correlated processes and their mixtures successfully. To the best of our knowledge, this is the first approach that is capable of separating (modeling) cross-correlated processes from their non-stationary mixtures. Next, we summarize the main contributions and the motivation of this thesis.

## 1.2. Main Contributions and Motivation

Here, we propose a novel and flexible modeling scheme by using particle filtering. It is well known that non-stationary and/or non-Gaussian problems can be efficiently solved by particle filters. By the choice of appropriate state variables and importance functions, major contributions in modeling of the following processes are presented for the first time in the literature:

- Modeling of TVAR *α-stable* processes.
- Modeling of cross-correlated non-stationary non-Gaussian processes.
- Modeling of non-stationary mixtures of cross-correlated AR processes.

Flexibility of the proposed methodology enables the solution of these three problems by the modification of the state variables and the importance functions. In all cases, it is assumed that we have no a priori information about the time variations of the state variables. This constitutes the most challenging part of the proposed techniques, since the functional form of the state-transition equation needs to be known in a typical particle

filtering framework (Doucet *et al*., 2001). For this purpose, different artificial state-transition equations are proposed for each problem in this thesis. Moreover, *Bootstrap* particle filtering is utilized in all modeling schemes for the sake of simplicity. It is well known that importance function is chosen to be *a priori state-transition pdf* in Bootstrap particle filtering (Doucet *et al*., 2001). In other words, proposing an artificial state-transition equation corresponds to the choice of the importance function.

Modeling of TVAR non-Gaussian processes has been performed by (Djuric *et al*., 2001; Djuric *et al*., 2002) using particle filtering. Despite its successful performance, the statistics of the non-Gaussian driving processes need to be known in these references. In this thesis, in order to provide a flexible methodology, we present a novel solution to model TVAR *α-stable* processes which constitutes a generalization of Gaussian processes. Here, *unknown statistics* (distribution parameters) of the *α-stable* processes are also estimated beside the TVAR coefficients. In order to present a solution for the most general case, we approach the problem by assuming two simplifications:

First, distribution parameters of the *Symmetric-α-stable* ( $S\alpha S$ ) process are assumed to be known and only TVAR coefficients are estimated. In this case, AR coefficients are modeled as state variables and an artificial random walk is utilized to model the unknown state-transition equation as in (Djuric *et al*., 2001; Djuric *et al*., 2002). However, generally, pdf of *α-stable* processes cannot be expressed analytically (except for Gaussian, Cauchy and Pearson distributions) unlike the cases in (Djuric *et al*., 2001; Djuric *et al*., 2002). Therefore, IFT of the characteristic function of the corresponding *α-stable* process is taken numerically to estimate the importance weights of each particle pertaining to the state variables. Performance of this method is also verified by the presentation of Posterior Cramer Rao Lower Bounds (PCRLB) for various AR values.

Secondly, distribution parameters of the $S\alpha S$ process are assumed to be unknown and these are estimated in addition to the unknown TVAR coefficients. Here, we propose a Hybrid Sequential Monte Carlo (HSMC) method constituting a two-stage Gibbs sampling algorithm. This method is composed of a particle filter and a Hybrid Monte Carlo (HMC) method, which are used iteratively to estimate the unknown TVAR coefficients and the unknown, *constant* distribution parameters, respectively. HMC is also a modified Gibbs

sampling algorithm. In the particle filtering stage, unknown AR coefficients constitute the state vector. Similar to the previous case, unknown state-transition is modeled by an artificial random walk model, given the estimated values of the distribution parameters. Then, HMC is utilized to estimate the unknown distribution parameters using the outputs (TVAR coefficients) of the particle filter. These two coupled Gibbs sampling stages are used iteratively until a convergence is obtained.

Finally, we propose a novel technique to model the most general case that can be encountered in the modeling of TVAR *α-stable* processes. In this situation, both AR coefficients and distribution parameters of an *α-stable* process change in time. Here, unknown and time-varying AR coefficients and distribution parameters are modeled by a single sequential Monte Carlo (SSMC) (particle filter) framework. Moreover, by this approach, skewed *α-stable* processes can also be modeled in addition to the $S\alpha S$ processes. For this purpose, a hybrid importance function is proposed by means of the artificial state-transition model: A random walk with time-varying covariances are utilized to model the transition of the states pertaining to the TVAR coefficients, whereas random walk with constant covariances are used for the state transitions of the distribution parameters. Importance weights are evaluated by the numerical calculation of the IFT of the characteristic function of the stable processes.

To the best of our knowledge, these are the first techniques that can be used to model TVAR *α-stable* processes in the literature. In addition to this successful non-stationary non-Gaussian process modeling, we also consider cases where relationships between different non-stationary non-Gaussian processes need to be modeled. For this purpose, we modify our particle filtering algorithm so that time-varying VAR processes can be modeled. This is an extension of the previous method to multivariate cases. By taking the elements of the TVAR coefficient *matrix* as state variables, an artificial random walk is used to model the unknown state-transition equations. Again, Bootstrap particle filtering is utilized and the importance weights of each particle corresponding to the state variables are calculated by the likelihood function. To the best of our knowledge, by the proposition of this approach, time-dependent relationships between different *non-Gaussian* processes are modeled for the first time via time-varying VAR modeling.

Finally, with our flexible modeling scheme, we bring a solution to modeling of the non-stationary *mixtures* of cross-correlated AR processes. This modeling scheme can also be interpreted as a DCA problem, since latent processes are modeled (separated) by observing their instantaneous mixtures. In order to approach to the problem from a simplified point of view, non-stationary mixtures of spatially correlated (cross-correlated) but temporally independent Gaussian processes (sources) are considered first. Moreover, we assume that the statistics of the Gaussian sources are known a priori. First, mixing matrix is estimated by particle filtering. Then, having estimated the time-varying mixing matrix, a MCMC algorithm is used to extract the sources. Here, the result of the particle filtering is used as *a priori* information in the MCMC scheme.

Then, a more general problem is solved where the assumption of informative *a priori* information of the sources is relaxed and non-stationary mixtures of cross-correlated Gaussian AR sources are modeled where mixtures of VAR models are observed. In addition to the state variables used for the mixing matrix and sources, AR coefficients and the dependency information are also included to the state vector. An artificial random walk model for the transitions of the AR coefficients, mixing matrix elements and noise variances is assigned with constant drift parameters. Here, a hybrid importance function is proposed where a mixture of the predicted states and the MAP estimate of the sources are utilized. By using this importance function, we bring solution to the separation of stationary and cross-correlated Gaussian AR sources from their non-stationary (time-varying) mixtures. With this flexible modeling scheme, non-stationary (time-varying) and cross-correlated Gaussian AR sources are also separated from their time-invariant mixtures. Next, the general outline of this thesis is presented.

## 1.3. Scope of the Thesis

This thesis is organized as follows: Chapter 1 gives an introduction including the literature survey in the field, main contributions brought by this thesis and an overview of the methods proposed in this work. In Chapter 2, brief background information on the theory of *α-stable* distributions is presented constituting a generalization of Gaussian processes. Then, fundamentals of VAR processes are given which will be useful in the modeling of cross-correlated AR processes. Finally, Chapter 2 is concluded with a short

review of the classical ICA problem where mixtures of statistically independent processes are modeled. Chapter 3 focuses on the Bayesian signal processing techniques such as the Monte Carlo integration, sampling methods, Markov Chain Monte Carlo approaches and particle filtering. In Chapter 4 we present our first contribution where we propose a novel method to model TVAR *α-stable* processes. After describing this non-stationary non-Gaussian process modeling strategy, we discuss an extension of this flexible approach to model cross-correlated VAR processes in Chapter 5. Later, we conclude this chapter with our third contribution on the flexible modeling of non-stationary non-Gaussian processes where non-stationary mixtures of cross-correlated processes are modeled. Both Chapter 4 and Chapter 5 are supported by simulations and then conclusions are drawn in Chapter 6. These are followed by the suggestions for future work in the field and finally some additional material is presented in the Appendix.

# 2. BACKGROUND INFORMATION

## 2.1. Introduction

This chapter starts with a brief presentation of *α-stable* distributions. In order to propose a new flexible methodology for the modeling of non-stationary non-Gaussian processes, we suggest using *α-stable* distributions to model non-Gaussian processes. The reason of this choice is that *α-stable* distributions present a generalization of Gaussian processes by satisfying the generalized central limit theorem and stability properties (Samorodnitsky and Taqqu, 1994). Moreover, their pdf's are analytically inexpressible unlike other non-Gaussian distributions such as generalized Gaussian or mixture of Gaussian distributions. Thus, modeling *α-stable* processes are considered here to provide a general modeling scheme.

The second part of this chapter provides a background material to model relationships between different time-structured processes. Due to its wide application areas (Hamilton, 1994), these time-structures are modeled by linear autoregressions here. In order to analyze cross correlated processes, multivariate extension of autoregressive modeling is introduced in this section by the utilization of VAR models.

This chapter is concluded by a brief description on mixture modeling of multivariate processes. Here, modeling of mixtures of statistically independent processes (sources) (ICA) is introduced.

## 2.2. Alpha Stable Processes

It is well known by the Central Limit Theorem (CLT) that, if we add a large number of random variables of different distributions, the summation variable tends to be Gaussian distributed as the number of terms goes to infinity. Moreover, it is necessary that each added random variable is of finite variance. Otherwise, CLT becomes insufficient and Generalized Central Limit Theorem should be used (Feller, 1966). In this case, the limiting

distribution is an *α-stable* distribution. Another common point between Gaussian and *α-stable* distributions is the stability property. This property states that linear combination of *α-stable* random variables having the same exponent is also an *α-stable* random variable but with different dispersion. This property allows us to utilize some assets of linear system theory developed for Gaussian distributions, since a linear system produces an *α-stable* output for an *α-stable* input.

*α-stable* distributions are defined in terms of their characteristic functions, since their pdf's cannot be obtained analytically, except for some limited cases (α=2, β=0 Gaussian; α=1, β=0 Cauchy; α=0.5, β=-1 Pearson) (Nikias and Shao, 1995). The characteristic function of *α-stable* distributions is given as follows:

$$\varphi(\zeta) = \exp\left\{ j\mu\zeta - \gamma|\zeta|^{\alpha}\left[1 + j\beta\, sign(\zeta)\omega(\zeta,\alpha)\right]\right\} \tag{2.1}$$

where $j = \sqrt{-1}$ and the parameters are defined within the following intervals: $-\infty < \mu < \infty,\ \gamma > 0,\ 0 < \alpha \le 2,\ -1 \le \beta \le 1$.

$$\omega(\zeta,\alpha) = \begin{cases} \tan\dfrac{\alpha\pi}{2} & ,\alpha \neq 1 \\ \dfrac{2}{\pi}\log|\zeta| & ,\alpha = 1 \end{cases} \qquad \text{and} \qquad sign(\zeta) = \begin{cases} 1, & \zeta > 0 \\ 0, & \zeta = 0 \\ -1, & \zeta < 0 \end{cases} \tag{2.2}$$

As shown above, an *α-stable* distribution is defined by four parameters and will be represented by $S_{\alpha}(\gamma,\beta,\mu)$, from now on[1]. Among these, $\alpha$ and $\beta$ are known as the characteristic exponent and symmetry parameters and they determine the thickness of the tails and the skewness of the distribution, respectively. $\beta = 0$ corresponds to *SαS* processes. As $\alpha$ gets smaller, distributions become more impulsive. $\mu$ and $\gamma$ are known as the measures of location and the dispersion around it, respectively. $\mu = 0$ and $\gamma$=1 is

---

[1]With this notation, a Gaussian distribution with $\mathcal{N}(\mu, 2\gamma^2)$ corresponds to $S_2(\gamma, 0, \mu)$.

known as a standard distribution.

In the literature, theory of *α-stable* distributions is widely studied and properties of these distributions are published in works such as (Zolotarev, 1989; Samorodnitsky and Taqqu, 1994; Nikias and Shao, 1995). However, one property of these distributions plays an utmost importance in this thesis for the understanding of the proposed methods which is known as the "Product Decomposition of Stable Densities" (Samorodnitsky and Taqqu, 1994; Godsill and Kuruoğlu, 1999). This property can be explained as follows:

**Theorem 2.1. Product Decomposition of Stable Densities**

- Let $z$ and $\lambda$ be independent random variables with the following distributions:

$$z \sim S_{\alpha'}(\gamma,0,0) \qquad \text{and} \qquad \lambda \sim S_{\alpha/\alpha'}\left\{\left(\cos\left(\frac{\pi\alpha}{2\alpha'}\right)^{\alpha'/\alpha}\right),1,0\right\} \tag{2.3}$$

where $0<\alpha'<2$ and $0<\alpha<\alpha'$. Here, $\lambda$ denotes an $\alpha/\alpha'$-stable random variable totally skewed to the right.

Then $\hat{n} = z\lambda^{1/\alpha'}$ is stable with distribution $\hat{n} \sim S_{\alpha}(\gamma,0,0)$.

This theorem allows us to represent a $S\alpha S$ random variable as a product of two *α-stable* random variables where one of them is a positive stable random variable ($\lambda$). For a proof of this theorem the reader is referred to (Samorodnitsky and Taqqu, 1994).

**Corollary to Theorem 2.1: (Scale Mixtures of Gaussians)**

If $\alpha'=2$ is chosen, it means that $z$ is a Gaussian random variable, i.e. $z \sim S_2(\gamma,0,0) = \mathcal{N}\left(0,2\gamma^2\right)$. If a positive stable random variable, $\lambda$, is drawn as shown above, i.e. $\lambda \sim S_{\alpha/2}\left\{\left(\cos\left(\frac{\pi\alpha}{4}\right)^{2/\alpha}\right),1,0\right\}$, then their product is distributed as a $S\alpha S$ random variable:

$$\hat{n} = z\lambda^{1/2} \sim S_{\alpha}(\gamma,0,0) \tag{2.4}$$

This property means, provided that a random variable $\lambda_t$ is sampled as in (2.3), at time *t*, $\hat{n}_t$ is distributed as *conditionally Gaussian* with distribution $\hat{n}_t \sim N\left(0,2\lambda_t\gamma^2\right)$. . Although $\hat{n}_t$ is conditionally Gaussian with different variances at each time instant (or data sample), a collection of $\tau$ samples of these $\hat{n}_t \sim N\left(0,2\lambda_t\gamma^2\right)$, provide an overall $S\alpha S$ distribution of $\hat{n} \sim S_\alpha\left(\gamma,0,0\right)$ with dispersion parameter $\gamma$. This is known as a heteroscedastic model (Godsill and Kuruoğlu, 1999) and enables one to express the distribution of a *SαS* random variable as conditionally Gaussian with distribution $\hat{n}_t \sim N\left(0,2\lambda_t\gamma^2\right)$ at a specific time instant *t*.

In the next section, VAR modeling will be introduced to analyze relationships between different processes. This model will be very useful to study cross correlated AR processes in the sequel. By means of VAR modeling, theoretical background will be presented which will be utilized in extending our non-stationary non-Gaussian modeling scheme to multivariate cases.

### 2.3. Vector Autoregressive Models

It is well known that AR modeling is widely used to examine temporal correlation information of a time series (Hamilton, 1994; Lütkepohl, 1993). In its most general form, a scalar (univariate) AR time series of order *K* is shown as follows:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_K y_{t-K} + n_t \tag{2.5}$$

where $\boldsymbol{\varphi} = \left[\phi_1,\phi_2,\ldots,\phi_K\right]^T$, $n_t$ denote the AR coefficient vector and the driving process, respectively. Here, $(.)^T$ represents the transposition operator. In literature, generally, driving process is modeled by a Gaussian distribution with the following statistics:

$$E\left[n_t\right] = 0, \qquad E\left[n_{t_1} n_{t_2}\right] = \begin{cases} \sigma_n^2 & t_1 = t_2 \\ 0 & otherwise \end{cases} \tag{2.6}$$

However, a scalar AR modeling cannot be used to interpret relationships among different processes. Therefore, cross-correlated processes form a vector which is denoted by $\mathbf{y}_t = \left[ y_{1,t}, y_{2,t}, \ldots, y_{d_1,t} \right]^T$. Before presenting its multivariate form, a scalar AR(1) model is given below for a recapitulation:

$$y_t = \phi_1 y_{t-1} + n_t \tag{2.7}$$

For illustrative purposes, multivariate generalization is first shown by a bivariate VAR process of order 1:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix} \tag{2.8}$$

where $d_1 = 2$ and $K = 1$ is taken. It can be observed that (2.8) models the relationships between two processes, $y_1$ and $y_2$, through the *AR matrix* and positive definite covariance matrix of the driving process which can be non-diagonal:

$$\mathbf{\Phi}_1 = \left( \boldsymbol{\varphi}_1^T, \boldsymbol{\varphi}_2^T \right)^T = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \quad \text{and} \quad E\left[ \mathbf{n}_{t_1} \mathbf{n}_{t_2}^T \right] = \begin{cases} \mathbf{\Sigma}_\mathbf{n}, & t_1 = t_2 \\ \mathbf{0}, & otherwise \end{cases} \tag{2.9}$$

After presenting the above simplified model, a full vector generalization of (2.5) and (2.6) can be represented by the following VAR($K$) model:

$$\mathbf{y}_t = \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_K \mathbf{y}_{t-K} + \mathbf{n}_t \tag{2.10}$$

where $\mathbf{y}_t = \left[ y_{1,t}, y_{2,t}, \ldots, y_{d_1,t} \right]^T$, $\mathbf{n}_t = \left[ n_{1,t}, n_{2,t}, \ldots, n_{d_1,t} \right]^T$ and $\mathbf{\Phi}_\mathrm{j}$ denote the VAR($K$) process, driving noise and ($d_1$ x $d_1$) matrix of AR coefficients for $j = 1, 2, \ldots, K$, respectively. With this notation, (2.6) is represented by the following covariance matrix:

$$E\left[\mathbf{n}_t\right]=\mathbf{0} \text{ and } E\left[\mathbf{n}_{t_1}\mathbf{n}_{t_2}^T\right]=\begin{cases}\boldsymbol{\Sigma}_{\mathbf{n}}, & t_1=t_2\\ \mathbf{0}, & otherwise\end{cases} \tag{2.11}$$

It is well known that a vector process $\mathbf{y}_t$ is wide-sense-stationary provided that its first and second moments are independent of $t$ (Hamilton, 1994). Moreover, it can be easily shown that any VAR($K$) process can be expressed in terms of a VAR(1) by making the following definitions:

$$\tilde{\mathbf{y}}_t=\begin{bmatrix}\mathbf{y}_t-E\left[\mathbf{y}_t\right]\\ \vdots\\ \mathbf{y}_{t-K+1}-E\left[\mathbf{y}_t\right]\end{bmatrix} \tag{2.12}$$

$$\mathbf{G}=\begin{bmatrix}\boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \boldsymbol{\Phi}_3 & \cdots & \boldsymbol{\Phi}_{K-1} & \boldsymbol{\Phi}_K\\ \mathbf{I}_{d_1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{d_1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{d_1} & \mathbf{0}\end{bmatrix} \text{ and } \tilde{\mathbf{n}}_t=\begin{bmatrix}\mathbf{n}_t\\ 0\\ \vdots\\ 0\end{bmatrix} \tag{2.13}$$

By using these definitions, a VAR($K$) process can be rewritten by the following VAR(1):

$$\tilde{\mathbf{y}}_t=\mathbf{G}\tilde{\mathbf{y}}_{t-1}+\tilde{\mathbf{n}}_t \tag{2.14}$$

where

$$E\left[\tilde{\mathbf{n}}_{t_1}\tilde{\mathbf{n}}_{t_2}^T\right]=\begin{cases}\boldsymbol{\Omega}, & t_1=t_2\\ \mathbf{0}, & otherwise\end{cases} \text{ and } \boldsymbol{\Omega}=\begin{bmatrix}\boldsymbol{\Sigma}_{\mathbf{n}} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\\ \vdots & \vdots & \cdots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}\end{bmatrix} \tag{2.15}$$

In many applications, it is usually desired to express the correlation matrix of $\tilde{\mathbf{y}}_t$ in terms of the correlation matrix of the driving process. This relationship is easily obtained to be as follows:

$$\boldsymbol{\Sigma} = \mathbf{G}\boldsymbol{\Sigma}\mathbf{G}^T + \boldsymbol{\Omega} \tag{2.16}$$

where the correlation matrix of $\tilde{\mathbf{y}}_t$ is expressed by $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{\Sigma} = E\left[\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_t^T\right]$. A closed form solution to (2.16) is found by the following relationship (Hamilton, 1994):

$$vec(\boldsymbol{\Sigma}) = \left[\mathbf{I}_{d^2} - \mathcal{A}\right]^{-1} vec(\boldsymbol{\Omega}) \tag{2.17}$$

where $d = d_1 K$, $\mathcal{A} \equiv (\mathbf{G} \otimes \mathbf{G})$ are used. Here, $\otimes$ denotes the Kronecker product and $vec(\boldsymbol{\Sigma}) = \left[\sigma_1^2, \sigma_3^2, \sigma_2^2, \sigma_4^2\right]^T$ for a 2x2 matrix of $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_2^2 \\ \sigma_3^2 & \sigma_4^2 \end{bmatrix}$. Using these information, the $j^{th}$ autocovariance of $\tilde{\mathbf{y}}_t$ can be obtained by the following equation (Hamilton, 1994):

$$E\left[\tilde{\mathbf{y}}_t \tilde{\mathbf{y}}_{t-j}^T\right] = \boldsymbol{\Sigma}_j = \mathbf{G}^j \boldsymbol{\Sigma} \text{ for } j = 1, 2, \ldots \tag{2.18}$$

However, ensuring the stationarity of a VAR($K$) process is of utmost importance in studying these signals. The following proposition provides the condition so that a VAR($K$) is wide-sense-stationary:

**Proposition 2.1: Wide-Sense-Stationary VAR processes**

- A VAR($K$) process is wide-sense-stationary as long as, $|e| < 1$, for all values of $e$ satisfying the following equation:

$$\left|\mathbf{I}_{d_1} e^K - \mathbf{\Phi}_1 e^{K-1} - \mathbf{\Phi}_2 e^{K-2} - \cdots - \mathbf{\Phi}_K\right| = 0 \tag{2.19}$$

where $e$ denotes the eigenvalues of the matrix **G** in (2.13).
The reader is referred to (Hamilton, 1994) for a proof of this proposition.

Although VAR modeling offers us an invaluable tool to study the relationships between various cross correlated AR processes, there are also many situations in nature that these processes cannot be observed directly. Next section provides a background material to model such scenarios where mixtures of processes are observed. These problems have been widely studied within the concepts of ICA, in literature. As the name implies, in these techniques mixtures of statistically independent processes have been examined. However, our main objective is to extend our flexible modeling scheme so that the mixtures of cross-correlated processes can also be handled.

## 2.4. Independent Component Analysis

ICA has been an extremely valuable tool used for the modeling of the mixtures of statistically independent signals (Comon, 1994; Cardoso, 1998; Hyvarinen *et al*., 2000; Hyvarinen *et al*., 2001). A typical mixing problem can be formulated as follows, where **A,** **y**(t)**, s**(t) are the mixing matrix, observation vector and source vector, respectively.

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{A}\mathbf{s}(t) \\ \mathbf{s}(t) &= [s_1(t), ....., s_{d_2}(t)]^T \\ \mathbf{y}(t) &= [y_1(t), ....., y_{d_1}(t)]^T \end{aligned} \tag{2.20}$$

where $d_1$ and $d_2$ denote the dimensions of the sensors and the sources, respectively.

ICA can be classified into two groups according to the nature of the mixing matrix **A**. It can be either time-varying or constant. If **A** is constant, then the mixture is called as an instantaneous mixture. ICA problem, which is considered in (2.20), can be visualized by the following figure:

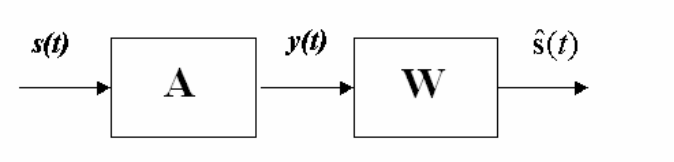


Figure 2.1. ICA scheme

In Figure 2.1., matrix **W** is known as the separating matrix and is adapted so that the output, $\hat{s}(t)$, has independent components. Ideally, the separating matrix **W** should be equal to the inverse of the mixing matrix **A**, at the end of the adaptations. Here, since there is no other reference signal than the mixture itself, the adaptation process is unsupervised (Haykin, 2000). It should also be noted that the utilized contrast measure is making the output components as independent as possible. So, at the output, sources can be reconstructed within a scale and a permuted way, as shown below:

$$\hat{s}(t) = Wy(t) = WAs(t) = \Delta \mathcal{P} s(t) \tag{2.21}$$

In (2.21), $\Delta$ is a nonsingular diagonal matrix and $\mathcal{P}$ is the permutation matrix and the components of $\hat{s}(t)$ are independent, as desired. So, after this point, the objective is to define a performance index, which is a contrast function, to measure how independent the reconstructed components are. Mostly, Kullback-Leibler divergence (KLD) is used as a contrast function (Comon, 1994). The reason for this can be explained as follows: The statistical independence of the components of a multivariate vector is obtained if the multiplication of the marginal pdf's of the components is equal to the vector's joint pdf as given below:

$$p_{\hat{S}}(\hat{s}_1, \hat{s}_2, ....., \hat{s}_{d_2}) = p(\hat{s}_1)p(\hat{s}_2).....p(\hat{s}_{d_2}) \tag{2.22}$$

where, $p_{\hat{S}}(.)$ denotes the joint pdf and $p(.)$'s are the marginal pdf's. So, it should be measured how much the joint pdf is similar to (2.22). The similarity between two pdf's can be measured by the KLD (Haykin, 2000), which is given below:

$$\mathcal{K}[p_1(\hat{\mathbf{s}}) \mid p_2(\hat{\mathbf{s}})] = \int_{-\infty}^{\infty} p_{\hat{S}}(\hat{s}_1, \hat{s}_2, ....., \hat{s}_{d_2}) \log\left( \frac{p_{\hat{S}}(\hat{s}_1, \hat{s}_2, ....., \hat{s}_{d_2})}{p(\hat{s}_1)p(\hat{s}_2).....p(\hat{s}_{d_2})} \right) d\hat{\mathbf{s}} \tag{2.23}$$

where, $p_1(\hat{\mathbf{s}}) = p_{\hat{S}}(\hat{s}_1, \hat{s}_2, ....., \hat{s}_{d_2})$ is the joint pdf and $p_2(\hat{\mathbf{s}}) = p(\hat{s}_1)p(\hat{s}_2).....p(\hat{s}_{d_2})$ is the multiplication of the marginal pdf's.

That is, if this KLD is estimated, one can understand how far vector $\hat{\mathbf{s}}$ is from independency. So, the objective is to make this contrast function to be equal to zero. This is also equal to the minimization of the mutual information (Cover and Thomas, 1991; Haykin, 2000). The mutual information between two random variables $\hat{s}_1$ and $\hat{s}_2$ can be given as follows:

$$I(\hat{s}_1; \hat{s}_2) = \iint p_{\hat{S}}(\hat{s}_1, \hat{s}_2) \log\left( \frac{p_{\hat{S}}(\hat{s}_1 \mid \hat{s}_2)}{p(\hat{s}_1)} \right) d\hat{s}_1 d\hat{s}_2 = \hbar(\hat{s}_1) - \hbar(\hat{s}_1 \mid \hat{s}_2) \tag{2.24}$$

where $\hbar(.)$ is known as the differential entropy, which is defined below:

$$\hbar(\hat{s}) = -E[\log p(\hat{s})] = -\int p(\hat{s}) \log p(\hat{s}) d\hat{s} \tag{2.25}$$

According to the Bayes Theorem (2.24) can be written as follows:

$$I(\hat{s}_1; \hat{s}_2) = \iint p_{\hat{S}}(\hat{s}_1, \hat{s}_2) \log\left( \frac{p_{\hat{S}}(\hat{s}_1, \hat{s}_2)}{p(\hat{s}_1)p(\hat{s}_2)} \right) d\hat{s}_1 d\hat{s}_2 \tag{2.26}$$

which is equal to the KLD given in (2.23) for two variables. In another KLD based ICA method, (Amari, 1998) approached the problem by proposing an optimization technique

where demixing matrix, **W**, is iteratively updated by a gradient descent method known as the "Natural Gradient Algorithm". (Cardoso and Laheld, 1996) approached the problem from another perspective and ended up with the same method recognized as the "Relative Gradient Algorithm". Both methods try to obtain independent components by minimizing the KLD divergence defined between vectors $\mathbf{s}(t)$ and $\hat{\mathbf{s}}(t)$ shown in Figure 2.1. The update equation can be given by the following equation:

$$\mathbf{W}(t) = \mathbf{W}(t-1) - \tilde{\mu}(t)\tilde{\nabla}\mathcal{K}(\mathbf{W}) \tag{2.27}$$

where $\tilde{\mu}(t)$ denotes the step size. Here, $\tilde{\nabla}\mathcal{K}$ denotes the natural gradient and it is given as follows:

$$\tilde{\nabla}\mathcal{K}(\mathbf{W}) \propto \frac{\partial \mathcal{K}(\mathbf{W})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W} \tag{2.28}$$

In another approach, (Jutten and Herault, 1991) used a basic feedback circuit where the elements of the input vector, $\mathbf{y}(t)$, are decorrelated in a nonlinear manner. Later, (Bell and Sejnowski, 1995) proposed to maximize the output entropy of a neural network with nonlinear outputs. Here, demixing mechanism is modeled by a neural network.

Beside these approaches, higher order statistical information about data has been utilized widely (Hyvarinen and Oja, 2000). It is seen from (2.20) and Figure 2.1. that each component of the observation vector **y**(t), that is $y_i$(t), is obtained by a linear combination of the components of the independent sources $s_1(t)$ through $s_{d_2}(t)$. According to the CLT, linear combination of various random variables tends to a Gaussian distribution. Thus, the distribution of mixture **y**(t) should be closer to the Gaussian distribution than the sources. Therefore, in order to obtain the independent components from the mixture, the non-Gaussianity can be maximized. One such contrast function is the kurtosis. Kurtosis is known as the fourth-order cumulant and is given as follows:

$$kurt(y) = E[y^4] - 3\left(E[y^2]\right)^2 \tag{2.29}$$

where the first term on the right-hand side of (2.29) is the fourth moment. Since the fourth moment of a Gaussian variable is given by $3\left(E[y^2]\right)^2$, it can easily be seen that the kurtosis is zero for Gaussian variables and thus a good measure for non-Gaussianity. But, this measure is not preferred much, since the kurtosis is not a robust measure of non-Gaussianity. Instead of this, more robust negentropy can be used. Negentropy can be defined as the difference of the differential entropy of the data and the differential entropy of a Gaussian variable with the same variance and shown as follows by $\mathcal{J}(\mathbf{y})$:

$$\mathcal{J}(\mathbf{y}) = h(\mathbf{y}_{GAUSS}) - h(\mathbf{y}) \tag{2.30}$$

Since a Gaussian random variable has the largest differential entropy among the random variables with the same variances but different distributions (Haykin, 2000), negentropy is always non-negative and equal to zero for the Gaussian data. In this method it should be noted that the negentropy is maximized. In literature, maximizing negentropy is widely used, such as in FastICA algorithm of (Hyvarinen and Oja, 2000).

# 3. BAYESIAN SIGNAL PROCESSING

## 3.1. Introduction

In this chapter we present a brief introduction to the Bayesian approaches in signal processing. We start our discussion with the presentation of the Bayes' theorem and essential definitions such as *a priori*, *a posteriori* distributions and likelihood function. Then, we will focus on obtaining point estimates from *a posteriori* distributions which will be used in the evaluation of the estimation methodologies.

Secondly, Monte Carlo integration will be introduced where computational methods are used for the calculation of point estimates via sampling techniques. This brief introduction will be followed by various sampling methods such as the rejection sampling (RS), importance sampling (IS), Markov Chain Monte Carlo (MCMC) algorithms and sequential importance sampling (SIS) methodology, which is widely known as the particle filtering.

## 3.2. The Bayesian Philosophy

Bayesian theory allows us to formulate our prior knowledge about the data which are ignored by the classical frequentist approaches. Mainly, by means of Bayesian approach, uncertainties about the parameters of interest are taken into account. That is, parameters are considered to be random variables unlike the frequentist approaches where they are assumed to be deterministic. Basically, our knowledge on the parameters of interest is expressed by their *a posteriori* pdf which is calculated by the Bayes' formula as shown below:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})}{p(\mathbf{y})} \tag{3.1}$$

where $p(\mathbf{x}|\mathbf{y})$, $p(\mathbf{y}|\mathbf{x})$, $p(\mathbf{x})$ are known as *a posteriori* pdf, likelihood function and *a priori* pdf, respectively. $p(\mathbf{y})$ is known as the evidence and it can be calculated by the following integration:

$$p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x} \tag{3.2}$$

Evidence is usually discarded during the interpretation of *a posteriori* pdf, since it is not dependent on the parameter of interest **x**. Therefore, (3.1) is usually written as follows in Bayesian literature (Box and Tiao, 1973; Gelman *et al*., 1995; Sivia, 1998):

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x}) \tag{3.3}$$

As (3.3) implies, our *a posteriori* information is formed when our *a priori* belief is shaped by the observed data, i.e. the likelihood function. If we have some *a priori* belief about the parameter before the data becomes available, this helps us to obtain inference about the parameter without the need of collecting infinitely many data samples unlike the case encountered in frequentist approaches.

Moreover, once the whole *a posteriori* pdf information about the parameter is obtained, any point estimate regarding the parameter can be calculated easily. For example, minimum mean squared error (MMSE) and maximum a posteriori (MAP) estimates of $x$ can be obtained respectively as follows:

$$\hat{\mathbf{x}}_{MMSE} = \int \mathbf{x} p(\mathbf{x}|\mathbf{y})\, d\mathbf{x} \tag{3.4}$$

$$\hat{\mathbf{x}}_{MAP} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) \tag{3.5}$$

where (3.4) and (3.5) are integration and optimization problems, respectively. Most of the time, *a posteriori* pdf's cannot be expressed analytically. Therefore, it is almost impossible calculating integrals of these pdf's which is highly needed to estimate point estimates, such

as the MMSE estimate shown in (3.4). (3.4) is a special case of a more general class of integrations known as the expectations. In order to estimate an expectation of a random vector **x**, the following integration must be calculated:

$$I(f) = \int f(\mathbf{x}) p(\mathbf{x} \mid \mathbf{y}) d\mathbf{x} \tag{3.6}$$

where (3.6) reduces to (3.4) in case of $f(\mathbf{x}) = \mathbf{x}$.

### 3.3. Monte Carlo Integration

In cases where posterior pdf and its integrations cannot be expressed analytically, numerical sampling procedures are of utmost importance to perform these tasks. If we can sample random variables from the desired (target) posterior pdf, then its MAP estimate given in (3.5) can be found by using numerical optimization techniques. Sampling a random variable from the desired pdf is shown by the following notation:

$$\mathbf{x} \sim p(\mathbf{x}) \tag{3.7}$$

By sampling *N* random variables from the desired pdf, analytically inexpressible pdf can be represented in terms of the drawn samples as shown below:

$$p(\mathbf{x}) \approx \sum_{i=1}^{N} \left[ \frac{p\left(\mathbf{x}^{(i)}\right)}{\sum_{j=1}^{N} p\left(\mathbf{x}^{(j)}\right)} \right] \delta\left(\mathbf{x} - \mathbf{x}^{(i)}\right) \tag{3.8}$$

where $\delta(.)$ is the Dirac delta function. Here, two requirements must be satisfied:

- We must be able to sample random variables from the target pdf
- We must be able to evaluate these random variables in the functional form of the desired pdf.

Once we can satisfy these conditions, we can both express the desired pdf and its integrations, such as the MMSE estimate (3.4). In this case, (3.6) can be numerically represented by the following equation:

$$E_{p(\mathbf{x}|\mathbf{y})}\left[f(\mathbf{x})\right] = \mathcal{I}(f) \approx \sum_{i=1}^{N} f\left(\mathbf{x}^{(i)}\right) \left[ \frac{p\left(\mathbf{x}^{(i)}\middle|\mathbf{y}\right)}{\sum_{j=1}^{N} p\left(\mathbf{x}^{(j)}\middle|\mathbf{y}\right)} \right] \tag{3.9}$$

In Monte Carlo integration, sampling from the posterior distribution constitutes the most important part. It is well known that deterministic integration suffers from computational complexity. Therefore, in Monte Carlo techniques, samples are drawn from the highly probable regions of the desired pdf.

In order to draw samples from the probable regions of the desired pdf, many techniques have been used in the literature. RS and IS are two powerful sampling schemes where i.i.d. samples from the target distribution are drawn (Gelman *et al*., 1995; Tanner, 1996; Robert and Casella, 1999; MacKay, 2003). On the other hand, MCMC approaches are widely utilized computational methods where drawn samples are not i.i.d. Here, samples are drawn so that the sequence of these samples constitutes a Markov Chain. After its transient, samples are drawn from the desired distribution when the Markov Chain converges to steady state (Gelman *et al*., 1995; Tanner, 1996; Gilks *et al*. 1998; Robert and Casella, 1999; MacKay, 2003). Below, these sampling techniques will be elaborated.

## 3.4. Sampling Techniques

This section starts with the presentation of RS and IS methods. Then, MCMC algorithms are introduced. In the literature, these techniques are widely used to draw samples almost from any distribution and they are usually preferred to process data in a batch framework. However, when data has a temporal or spatial structure, the need of sequential data processing arises. For such cases, SIS has been widely utilized in the literature (Doucet *et al*., 2000, Doucet *et al*., 2001). Finally, to study non-stationary non-Gaussian processes, we discuss the theory of particle filtering (SIS) in detail.

### 3.4.1. Rejection Sampling

It is usually tedious to draw samples directly from the target distribution (MacKay, 2003). Therefore, another pdf is chosen to sample from. This pdf is chosen so that drawing samples from this distribution is easier than that of the desired one. In rejection sampling, first a pdf $q(\mathbf{x})$ is proposed in such a way that its product by a constant takes higher values than the original pdf $p(\mathbf{x})$. This condition should be satisfied throughout the whole support region of random variables, named **x**'s. A one dimensional example to this scheme is shown below:

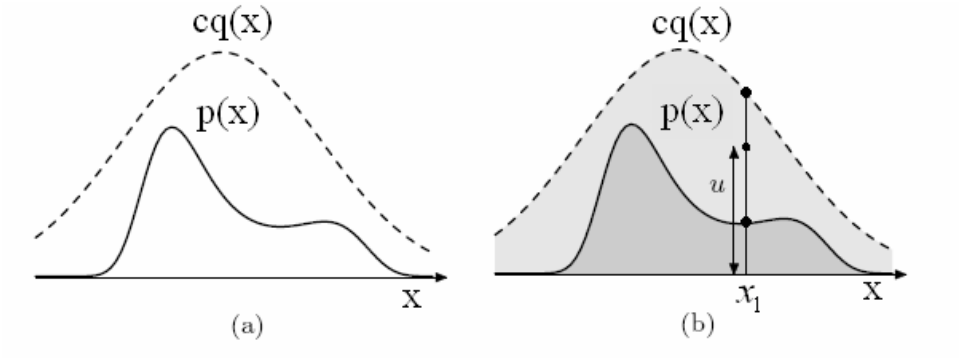


Figure 3.1. Rejection Sampling (MacKay, 2003)

(Note that $cq(x) > p(x)$ for all support region)

After selecting such a proposal density, a random sample, $x_1$, is drawn from this new density $q(x)$. Then the density functions $q(x)$ and $p(x)$ are evaluated at this random variable and the values of the pdf's are denoted by $cq(x_1)$ and $p(x_1)$, respectively. After performing this, a uniformly distributed random variable, *u*, is drawn from the interval $\left[0, cq(x_1)\right]$. Then, according to this random variable, sample $x_1$ is either accepted or rejected. If *u* is between $\left[p(x_1), cq(x_1)\right]$, $x_1$ is rejected, otherwise it is accepted.

### 3.4.2. Importance Sampling

It is seen that RS can be used for both generating samples from a desired distribution, $p(x)$, and estimating its expectations via substitution of these samples in (3.9). However,

IS is not a method for generating samples from $p(x)$; it is a method for estimating its expectations as shown by (3.9). As in RS, in order to draw samples, another distribution is chosen instead of the target pdf. This distribution is known as the importance function. If this pdf is denoted by $q(x)$, expectation of (3.9) can be represented as follows:

$$\mathcal{I}(f) \approx \frac{\sum_{i=1}^{N} f\left(x^{(i)}\right)\left[\frac{p\left(x^{(i)}\right)}{q\left(x^{(i)}\right)} q\left(x^{(i)}\right)\right]}{\sum_{i=1}^{N}\left[\frac{p\left(x^{(i)}\right)}{q\left(x^{(i)}\right)} q\left(x^{(i)}\right)\right]} = \frac{E_{q(x)}\left[f(x)\frac{p(x)}{q(x)}\right]}{E_{q(x)}\left[\frac{p(x)}{q(x)}\right]} \tag{3.10}$$

where conditionality on the observation is dropped for the sake of simplicity, i.e. $p\left(x^{(i)}\right) = p\left(x^{(i)} \middle| y\right)$ and $q\left(x^{(i)}\right) = q\left(x^{(i)} \middle| y\right)$ is used in short.

It should be noted that the expectation in (3.9) is taken with respect to the desired pdf $p(x)$, whereas in (3.10) it is transformed into an expectation with respect to $q(x)$, from which sampling is easier. The right hand side of (3.10) can be shown as follows:

$$\mathcal{I}(f) \approx \frac{E_{q(x)}\left[f(x)\frac{p(x)}{q(x)}\right]}{E_{q(x)}\left[\frac{p(x)}{q(x)}\right]} = \frac{\sum_{i=1}^{N} f\left(x^{(i)}\right) w^{(i)}}{\sum_{j=1}^{N} w^{(j)}} \tag{3.11}$$

where samples are drawn from $q(x)$, i.e. $x \sim q(x)$. In (3.11), importance weights are defined with the following definition:

$$w(x) \triangleq \frac{p(x)}{q(x)} \tag{3.12}$$

### 3.4.3. Markov Chain Monte Carlo methods

In addition to the computational complexity problem of the i.i.d. sampling methods, rejection sampling method has also one more disadvantage: The proposal pdf $q(\mathbf{x})$ should look like the original pdf $p(\mathbf{x})$ throughout the whole support region of $\mathbf{x}$ (MacKay, 2003). Choosing such a distribution is too difficult and becomes even harder when the dimension increases (Metropolis *et al*., 1953; Gelman *et al*., 1995; Gilks *et al*., 1998; Robert and Casella, 1999; MacKay, 2003). Instead of choosing such a whole distribution to fit, we can construct a Markov chain whose samples are asymptotically distributed according to the desired pdf $p(\mathbf{x})$. These are known as the MCMC techniques. Below, some important definitions and properties used in MCMC literature are presented.

Let the probability distribution of the state is represented by $p^{(t)}(\mathbf{x})$ at iteration *t*. The objective is to find a Markov chain such that as $t \to \infty$, $p^{(t)}(\mathbf{x})$ approaches to the desired distribution $p(\mathbf{x})$ (Metropolis *et al*., 1953; Gelman *et al*., 1995; Gilks *et al*., 1998; Robert and Casella, 1999; MacKay, 2003). In order to define a Markov chain, an initial probability distribution $p^{(0)}(\mathbf{x})$ and a transition probability $T(\mathbf{x}';\mathbf{x})$ need to be specified. The probability distribution of the state at the $(t+1)th$ iteration of the Markov chain is given by the following equation:

$$p^{(t+1)}(\mathbf{x}') = \int T(\mathbf{x}';\mathbf{x}) p^{(t)}(\mathbf{x})\, d\mathbf{x} \tag{3.13}$$

where integration is taken over $d_1$ dimensional $\mathbf{x}$ vector, defined as $\mathbf{x} = \left[ x_1, x_2, \ldots, x_{d_1} \right]^T$.

A MCMC method should satisfy the following criteria (MacKay, 2003):

- The desired distribution $p(\mathbf{x})$ is an *invariant* distribution of the chain. A distribution $\pi(\mathbf{x})$ is an invariant distribution of the transition probability $T(\mathbf{x}';\mathbf{x})$ if the following equation is satisfied:

$$\pi(\mathbf{x}') = \int T(\mathbf{x}';\mathbf{x})\pi(\mathbf{x})\,d\mathbf{x} \tag{3.14}$$

(3.14) tells us that the Markov chain has completed its transient and reached to a steady state where samples from $\pi(\mathbf{x})$ are drawn. This transient is known as the burn-in period.

- Markov chain must be ergodic:

$$p^{(t)}(\mathbf{x}) \to \pi(\mathbf{x}) \text{ as } t \to \infty \text{ for any } p^{(0)}(\mathbf{x}) \tag{3.15}$$

Also, in order for a Markov chain to be ergodic, its matrix must be irreducible and the chain has to be aperiodic. These concepts are defined next:

- Irreducibility: The matrix of the chain is said to be irreducible if all subsets of states can be reached from each other.
- Aperiodicity: The chain should not visit the explored parts periodically, i.e. for some initial conditions, $p^{(t)}(\mathbf{x})$ doesn't tend to a limit cycle.

Moreover, in order to draw samples from an invariant distribution, Markov chain should satisfy the *detailed balance* condition which is explained below:

If a Markov chain satisfies detailed balance, this means that the probability of passing from a particular state $\mathbf{x}_i$ to another $\mathbf{x}_j$ is equal to passing from $\mathbf{x}_j$ to $\mathbf{x}_i$:

$$T(\mathbf{x}_j;\mathbf{x}_i)\,p(\mathbf{x}_i) = T(\mathbf{x}_i;\mathbf{x}_j)\,p(\mathbf{x}_j) \text{ for all } \mathbf{x}_i \text{ and } \mathbf{x}_j \tag{3.16}$$

If a Markov chain satisfies detailed balance it is also called as a reversible Markov chain. Detailed balance property implies the invariance of $p(\mathbf{x})$ under the Markov chain $T$, which is our objective (Metropolis *et al.*, 1953; Gelman *et al.*, 1995; Gilks *et al.*, 1998; Robert and Casella, 1999; MacKay, 2003).

Generation of such Markov Chains is a product of the examination of the electron distributions in solid-state physics by (Metropolis *et al.*, 1953). In this work, such Markov

Chains has been utilized for optimization purposes. The optimized function is selected to be proportional to the energy of the states. The idea was exploited to avoid trapping in local extramum points by looking at the values of the criterion function at different temperature values. As a result of this search performed by jumping from one temperature value to another, the global optimum is found which corresponds reaching to the target distribution in our case. As a result of these jumps, a Markov chain is formed from the desired distribution. Therefore, this optimization method has led to the development of a general framework for the construction of Markov chains with the above properties. This method, first proposed by (Metropolis *et al*., 1953), is known as the Metropolis-Hastings method. This technique is elaborated first, which is followed by its variants in the following sections.

### 3.4.3.1. Metropolis-Hastings method

Here, the state space is explored by a Markov chain which simplifies the need of constructing a proposal distribution $q(\mathbf{x})$ which should resemble the target density in the whole space. Here, a Markov chain is used to propose new states based on the current ones. These new states, $\mathbf{x}^{(t+1)}$, are drawn from simpler functions which are located at the current values of the state variables $\mathbf{x}^{(t)}$. Such a proposal distribution, $q(\mathbf{x}';\mathbf{x})$, is illustrated below for a one dimensional case where the target distribution is denoted by $p(\mathbf{x})$:

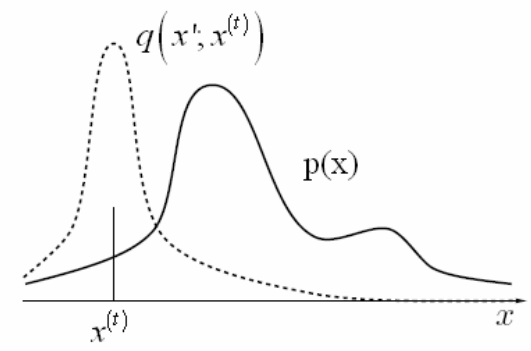


Figure 3.2. Metropolis-Hastings method in one dimension (MacKay, 2003)

After drawing a new state $\mathbf{x}'$ from the proposal distribution, $q(\mathbf{x}';\mathbf{x})$, the following ratio, known as the acceptance ratio, is calculated:

$$\varepsilon = \frac{p(\mathbf{x}')\,q\left(\mathbf{x}^{(t)};\mathbf{x}'\right)}{p\left(\mathbf{x}^{(t)}\right)q\left(\mathbf{x}';\mathbf{x}^{(t)}\right)} \tag{3.17}$$

If $\varepsilon \geq 1$, then $\mathbf{x}'$ is accepted to the Markov chain: $\overline{X} = \left\{\ldots, \mathbf{x}^{(t-1)}, \mathbf{x}^{(t)}, \mathbf{x}'\right\}$, i.e. $\mathbf{x}^{(t+1)} = \mathbf{x}'$. If $\varepsilon < 1$, then $\mathbf{x}'$ is accepted with probability $\varepsilon$. If it is rejected, then the Markov chain proceeds with $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)}$, i.e. $\overline{X} = \left\{\ldots, \mathbf{x}^{(t-1)}, \mathbf{x}^{(t)}, \mathbf{x}^{(t)}\right\}$.

### 3.4.3.2. Metropolis method

The Metropolis algorithm is a special case of the Metropolis-Hastings method where the proposal distribution is selected to be a symmetrical density centered around the current state $\mathbf{x}^{(t)}$. In this case, the acceptance ratio reduces to the following form, since $q\left(\mathbf{x}';\mathbf{x}^{(t)}\right) = q\left(\mathbf{x}^{(t)};\mathbf{x}'\right)$:

$$\varepsilon = \frac{p(\mathbf{x}')}{p\left(\mathbf{x}^{(t)}\right)} \tag{3.18}$$

### 3.4.3.3. Gibbs sampling

If drawing samples from a target pdf, $p(\mathbf{x})$, is difficult, we could draw random samples from another pdf, $q(\mathbf{x})$, as explained in the preceding sections. On the other hand, if we cannot easily draw samples from $p(\mathbf{x})$, but easily draw samples from its conditional probabilities, then we can also use Gibbs sampling method. Gibbs sampling is a version of MH algorithm where the proposal distributions are taken to be the conditional distributions of their joint pdf, $p(\mathbf{x})$, where state vector is composed of $d_1$ components, i.e. $\mathbf{x} = \left[x_1, x_2, \ldots, x_{d_1}\right]^T$. So, Gibbs sampling can be used for at least two-dimensional cases. For this case $(d_1 = 2)$, the method can be explained as follows: At iteration *t*, let's take our random sample vector to be $\mathbf{x}^{(t)} = \left[x_1^{(t)}, x_2^{(t)}\right]^T$. First, $x_1$ is drawn from the conditional

density: $x_1^{(t+1)} \sim p\left(x_1 \middle| x_2^{(t)}\right)$. Then a sample of $x_2$ is drawn from its conditional pdf: $x_2^{(t+1)} \sim p\left(x_2 \middle| x_1^{(t+1)}\right)$. Here, it should be noted that, at every sampling, the new value of the previously drawn component is substituted in the conditional pdf for drawing the next component. This is illustrated in Figure 3.3.

The general scheme for $k$ dimensions can be expressed by the following set of equations:

$$\begin{aligned}
x_1^{(t+1)} &\sim p(x_1 \mid x_2^{(t)}, x_3^{(t)}, \ldots, x_k^{(t)}) \\
x_2^{(t+1)} &\sim p(x_2 \mid x_1^{(t+1)}, x_3^{(t)}, \ldots, x_k^{(t)}) \\
x_3^{(t+1)} &\sim p(x_3 \mid x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_k^{(t)}) \\
&\vdots \qquad\qquad \vdots \\
x_k^{(t+1)} &\sim p(x_k \mid x_1^{(t+1)}, x_2^{(t+1)}, \ldots, x_{k-1}^{(t)})
\end{aligned} \tag{3.19}$$

In the above Gibbs sampling scheme, it should be noted that it can be viewed as a Metropolis algorithm where every proposal is accepted.

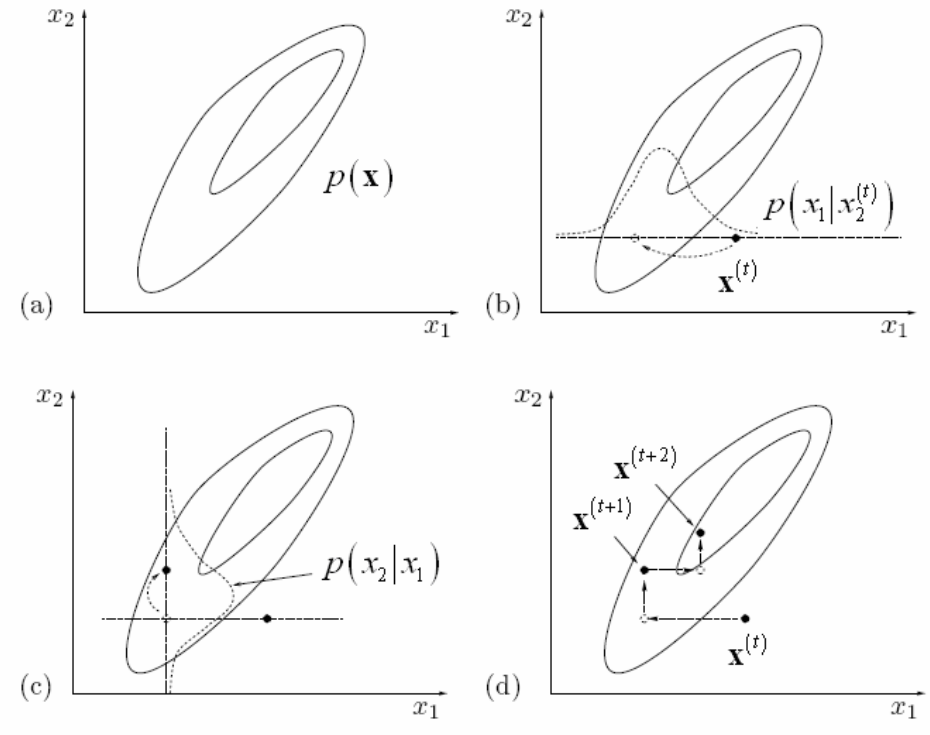


Figure 3.3. Gibbs sampling in two dimensions (MacKay, 2003)

### 3.4.4. Sequential Monte Carlo methods

In many application areas, observation order of the data is of utmost importance (Haykin, 1996; Doucet *et al*., 2001, Ristic *et al*., 2004). Especially in non-stationary cases, pdf of the related parameter changes over time. Therefore, in such cases, pdf and expectations need to be updated sequentially as the new data samples are observed. For this purpose, these dynamic systems are usually expressed in terms of state space equations as shown below:

$$\mathbf{x}_t = f_t\left(\mathbf{x}_{t-1}, \mathbf{v}_t\right) \tag{3.20.a}$$

$$\mathbf{y}_t = h_t\left(\mathbf{x}_t, \mathbf{n}_t\right) \tag{3.20.b}$$

where $\mathbf{x}_t$ and $\mathbf{y}_t$ represent the hidden state and the observation vectors at current time $t$, respectively. Here, the process and observation noises are denoted by $\mathbf{v}_t$ and $\mathbf{n}_t$, respectively. $f_t$ and $h_t$ are known as the process and observation functions and in their most general case, they are nonlinear. The noise processes in (3.20) are modeled to be non-Gaussian. Here, the objective is to estimate sequentially *a posteriori* distribution of the state variables obtained via the observation data gathered up to that time, i.e. $p\left(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}\right)$, where $\mathbf{x}_{0:t} = \left[\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t\right]$ and $\mathbf{y}_{1:t} = \left[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t\right]$.

Until the development of particle filters, the general modeling of (3.20) had been approximated by linear and Gaussian state-space equations which reduce to Kalman filter modeling. Below, the ideal recursive Bayesian solution of (3.20) will be given first. Next, in case of linear and Gaussian forms of (3.20), Kalman filtering equations will be recapitulated. After mentioning on some variants of Kalman filtering, particle filtering method will be explained in detail.

Sequential estimation of the posterior distributions of $\mathbf{x}$ is performed in two stages, namely prediction and update. In the prediction stage, the current value of the hidden variable, $\mathbf{x}_t$, is predicted from the previous observations and shown by the following Chapman-Kolmogorov equation:

$$p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1} \tag{3.21}$$

where we used $p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}\right) = p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}\right)$, since (3.20.a) describes a first order Markov process. At time step *t*, the prior is updated by the measurement $\mathbf{y}_t$ via Bayes rule as follows:

$$\begin{aligned} p(\mathbf{x}_t \mid \mathbf{y}_{1:t}) &= p(\mathbf{x}_t \mid \mathbf{y}_t, \mathbf{y}_{1:t-1}) \\ &= \frac{p(\mathbf{y}_t \mid \mathbf{x}_t, \mathbf{y}_{1:t-1}) p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1})} \\ &= \frac{p(\mathbf{y}_t \mid \mathbf{x}_t) p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1})} \end{aligned} \tag{3.22}$$

where the normalizing constant is given by the following integration:

$$p(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}) = \int p(\mathbf{y}_t \mid \mathbf{x}_t) p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}) d\mathbf{x}_t \tag{3.23}$$

In this general model, the statistics of $\mathbf{n}_t$ and $\mathbf{v}_t$ are assumed to be known. Recurrence relations (3.21) and (3.22) give the optimal Bayesian solution. This recursive solution cannot be determined analytically in general (Ristic *et al*., 2004). However, if the $f(.)$ and $h(.)$ functions are linear and the **v**, **n** noise terms are Gaussian, then model (3.20) can be written as follows:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{v}_t \\ \mathbf{y}_t &= \mathbf{H}_t \mathbf{x}_t + \mathbf{n}_t \end{aligned} \tag{3.24}$$

where $\mathbf{F_t}$ and $\mathbf{H_t}$ are linear. In this case, instead of propagating and updating the whole pdf by (3.21) and (3.22), mean and covariance matrices become sufficient to analyze as a consequence of the CLT. In the following equations, covariances of the prediction and observation noise terms will be denoted by $\mathbf{Q_t}$ and $\mathbf{R_t}$, respectively. For this situation, the exact analytical solution to (3.21) and (3.22) can be given by the classical Kalman Filter equations (Kalman, 1960), which are summarized below for the sake of completeness:

$$
\begin{aligned}
p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}) &= \mathcal{N}(\mathbf{x}_{t-1}; \mathbf{m}_{t-1|t-1}, \mathbf{P}_{t-1|t-1}) \\
p(\mathbf{x}_t \mid \mathbf{y}_{1:t-1}) &= \mathcal{N}(\mathbf{x}_t; \mathbf{m}_{t|t-1}, \mathbf{P}_{t|t-1}) \\
p(\mathbf{x}_t \mid \mathbf{y}_{1:t}) &= \mathcal{N}(\mathbf{x}_t; \mathbf{m}_{t|t}, \mathbf{P}_{t|t})
\end{aligned}
\tag{3.25}
$$

where

$$
\begin{aligned}
\mathbf{m}_{t|t-1} &= \mathbf{F}_t \mathbf{m}_{t-1|t-1} \\
\mathbf{P}_{t|t-1} &= \mathbf{Q}_{t-1} + \mathbf{F}_t \mathbf{P}_{t-1|t-1} \mathbf{F}_t^T \\
\mathbf{m}_{t|t} &= \mathbf{m}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \mathbf{m}_{t|t-1}) \\
\mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \mathbf{P}_{t|t-1}
\end{aligned}
\tag{3.26}
$$

In the equations given above, $\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P})$ is a Gaussian density with parameter **x**, mean **m**, covariance **P** and

$$
\begin{aligned}
\mathbf{S}_t &= \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T + \mathbf{R}_t \\
\mathbf{K}_t &= \mathbf{P}_{t|t-1} \mathbf{H}_t^T \mathbf{S}_t^{-1}
\end{aligned}
\tag{3.27}
$$

are the covariance of the innovation term $\mathbf{y}_t - \mathbf{H}_t \mathbf{m}_{t|t-1}$ and the Kalman gain, respectively.

However, when the $f(.)$ and $h(.)$ functions in (3.20) are not linear, two main approximations have been used in the literature: One of them is known as the EKF where the state distribution is modeled by a Gaussian random variable and then propagated analytically through the first-order linearization of the nonlinear system (Anderson and Moore, 1979). As a result of this linearization, large errors can occur and sometimes the filter diverges. In order to avoid these problems, Unscented Kalman Filter (UKF) has been used in nonlinear estimation by (Wan and van der Merwe, 2000). In UKF, state distribution is again approximated by a Gaussian random variable but it is represented by a carefully chosen minimal number of points. These points are *sampled deterministically* and then passed through the true nonlinear system. It has been observed that these points preserve the true mean and covariance of the state variable up to second order with any nonlinearity, while the EKF could only achieve first-order accuracy.

On the other hand, for the most general case, the state distributions are non-Gaussian distributed. Therefore, analytical solutions are not available unlike the well known Kalman filter and the use of EKF or UKF are not satisfactory due to the non-Gaussian distribution of the states. So, new techniques should be used to approximate the optimal solution equations (3.21) and (3.22). Since these expressions involve intractable integrals, numerical solutions, such as the Monte Carlo methods, should be used, which makes use of *stochastic sampling*. On the other hand, Monte Carlo methods discussed in the preceding sections need to be modified in order to compute the optimal solution as the new data become available, in a similar manner in Kalman filter theory. So, in order to compute the equations (3.21) and (3.22) of Bayesian filtering in a sequential manner, the IS method needs to be modified. The new method is known as the "Sequential Importance Sampling (SIS)". In SIS, the posteriori distribution can be approximated by *N* discrete particles with normalized weights as follows:

$$p\left(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}\right) = \sum_{i=1}^{N} \tilde{w}_t^{(i)} \delta\left(\mathbf{x}_{0:t} - \mathbf{x}_{0:t}^{(i)}\right) \tag{3.28}$$

where $\tilde{w}_t^{(i)}$'s represent the normalized importance weights of particles $\left\{\mathbf{x}_{0:t}^{(i)}, i = 1,...,N\right\}$, which are calculated as follows:

$$\tilde{w}_t^{(i)} = \frac{w\left(\mathbf{x}_{0:t}^{(i)}\right)}{\sum_{j=1}^{N} w\left(\mathbf{x}_{0:t}^{(j)}\right)} \tag{3.29}$$

where $w\left(\mathbf{x}_{0:t}\right)$ denotes the unnormalized importance weight, which is given by the following equation:

$$w\left(\mathbf{x}_{0:t}\right) = \frac{p\left(\mathbf{x}_{0:t}\left|\mathbf{y}_{1:t}\right.\right)}{q\left(\mathbf{x}_{0:t}\left|\mathbf{y}_{1:t}\right.\right)} \text{ implying that } \tilde{w}\left(\mathbf{x}_{0:t}\right) \propto \frac{p\left(\mathbf{x}_{0:t}\left|\mathbf{y}_{1:t}\right.\right)}{q\left(\mathbf{x}_{0:t}\left|\mathbf{y}_{1:t}\right.\right)} \tag{3.30}$$

Here, $q\left(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}\right)$ is known as the importance function and the samples are drawn from this distribution. In (3.30) the importance weight of the joint state vector, $\mathbf{x}_{0:t}$, is shown instead of the instantaneous weight, $\mathbf{x}_t$, of (3.12) (Note that the conditionality on $\mathbf{y}_t$ had been dropped in (3.12) for the sake of simplicity).

In order to perform the IS in a sequential manner, dependence of (3.30) on the joint state vector should be expressed in terms of the sequentially arriving samples. Let's suppose that we have an approximation for the posterior at time $(t-1)$,i.e. $p\left(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}\right)$, and choose our importance function in the following form:

$$q\left(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}\right) \triangleq q\left(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}\right) q\left(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}\right) \tag{3.31}$$

As a result of (3.31), we can draw the following conclusions:

a) Suppose that we have samples drawn at time $(t-1)$: $\mathbf{x}_{0:t-1}^{(i)} \sim q\left(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}\right)$

b) Then we draw our new samples from the following distribution: $\mathbf{x}_t^{(i)} \sim q\left(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}\right)$ meaning that the new samples depend on both the previous ones, $\mathbf{x}_{0:t-1}$, and the new measurement $\mathbf{y}_t$.

c) We augment the state vector given in (a) with those drawn in (b): $\mathbf{x}_{0:t}^{(i)} = \left\{\mathbf{x}_{0:t-1}^{(i)}, \mathbf{x}_t^{(i)}\right\}$

By using such an importance function, the importance weight calculation given in (3.30) can be expressed sequentially by using the following derivations:

$$\begin{aligned} p\left(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}\right) &= p\left(\mathbf{x}_{0:t}|\mathbf{y}_t,\mathbf{y}_{1:t-1}\right) \\ &= \frac{p\left(\mathbf{y}_t|\mathbf{x}_{0:t},\mathbf{y}_{1:t-1}\right) p\left(\mathbf{x}_{0:t}|\mathbf{y}_{1:t-1}\right)}{p\left(\mathbf{y}_t|\mathbf{y}_{1:t-1}\right)} \\ &= \frac{p\left(\mathbf{y}_t|\mathbf{x}_{0:t},\mathbf{y}_{1:t-1}\right) p\left(\mathbf{x}_t|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t-1}\right) p\left(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}\right)}{p\left(\mathbf{y}_t|\mathbf{y}_{1:t-1}\right)} \end{aligned} \tag{3.32}$$

$$= \frac{p\left(\mathbf{y}_t|\mathbf{x}_t\right) p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)}{p\left(\mathbf{y}_t|\mathbf{y}_{1:t-1}\right)} p\left(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}\right) \tag{3.33}$$

$$\propto p\left(\mathbf{y}_t|\mathbf{x}_t\right) p\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right) p\left(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}\right) \tag{3.34}$$

where we made use of the first order Markov property of the state equation to obtain (3.33) from (3.32). By substituting (3.31) and (3.34) in (3.30), the weight update equation is obtained as follows:

$$\begin{aligned} \tilde{w}_t^{(i)} &\propto \frac{p\left(\mathbf{y}_t|\mathbf{x}_t^{(i)}\right) p\left(\mathbf{x}_t^{(i)}\middle|\mathbf{x}_{t-1}^{(i)}\right) p\left(\mathbf{x}_{0:t-1}^{(i)}\middle|\mathbf{y}_{1:t-1}\right)}{q\left(\mathbf{x}_t^{(i)}\middle|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t}\right) q\left(\mathbf{x}_{0:t-1}^{(i)}\middle|\mathbf{y}_{1:t-1}\right)} \\ &= \frac{p\left(\mathbf{y}_t|\mathbf{x}_t^{(i)}\right) p\left(\mathbf{x}_t^{(i)}\middle|\mathbf{x}_{t-1}^{(i)}\right)}{q\left(\mathbf{x}_t^{(i)}\middle|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t}\right)} \tilde{w}_{t-1}^{(i)} \\ &= \frac{p\left(\mathbf{y}_t|\mathbf{x}_t^{(i)}\right) p\left(\mathbf{x}_t^{(i)}\middle|\mathbf{x}_{t-1}^{(i)}\right)}{q\left(\mathbf{x}_t^{(i)}\middle|\mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t}\right)} \tilde{w}_{t-1}^{(i)} \end{aligned} \tag{3.35}$$

As a result of these, posterior filtered density $p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right)$ can be approximated as follows:

$$p\left(\mathbf{x}_t|\mathbf{y}_{1:t}\right) = \sum_{i=1}^{N} \tilde{w}_t^{(i)} \delta\left(\mathbf{x}_t - \mathbf{x}_t^{(i)}\right) \tag{3.36}$$

Unfortunately, expressing the update equation in a recursive way does not suffice to solve the problem. This time, a new problem which is called as the "Degeneracy Problem" arises as a major obstacle. This phenomenon can be explained as follows: Due to the selection of the importance function as in (3.31), the variance of the importance weights can only increase in time (Doucet *et al.*, 2000). After several iterations, almost all normalized importance weights tend to be zero, resulting in unsuccessful estimations. In order to limit the degeneracy effect, the following choice of importance weight is utilized in (Doucet *et al.*, 2000) to minimize the variance of the importance weights $w_t^{(i)}$ conditional upon $\mathbf{x}_{0:t-1}^{(i)}$ and $\mathbf{y}_{1:t}$ :

$$q\left(\mathbf{x}_t \left| \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t} \right.\right) = p\left(\mathbf{x}_t \left| \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t} \right.\right) \tag{3.37}$$

This is known as the optimal importance function (Doucet *et al*., 2000, Doucet *et al*., 2001). But despite these efforts, we cannot solve the problem of degeneracy. That's why; resampling methods are introduced to eliminate the trajectories with small importance weights. In order to control the degeneracy of the algorithm, an optimal criterion can be calculated as follows (Doucet *et al*., 2001):

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N} (\tilde{w}_t^i)^2} \tag{3.38}$$

So, if the value found by (3.38) is greater than a threshold value (This threshold value can be chosen to be $2N/3$), no resampling is performed. Otherwise, random samples are resampled in the following way: For $i = 1, 2, \ldots, N$, a new index value, *j*, for each *i* is found by sampling from the discrete distribution by using the cumulative mass function of the importance weights. This procedure is illustrated below:

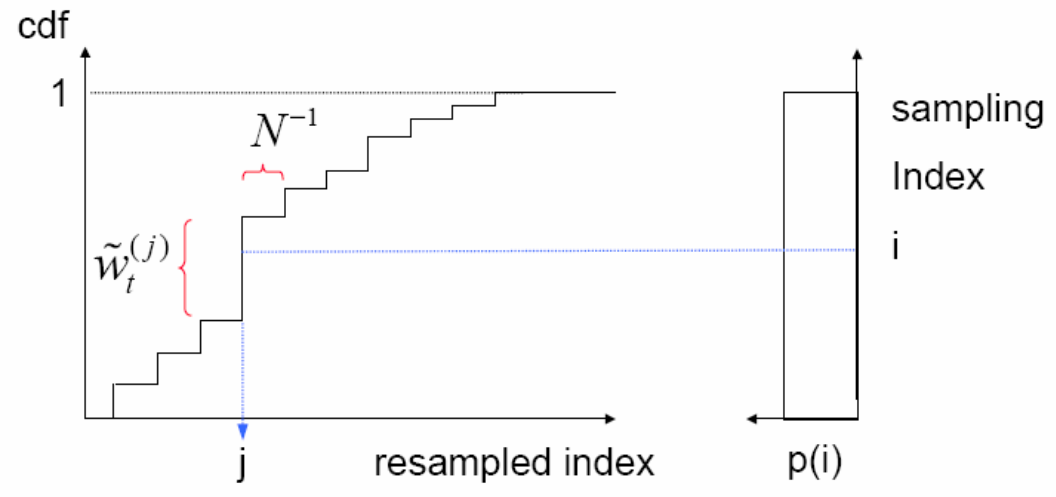


Figure 3.4. Resampling procedure (Van der Merwe *et al*., 2000)

where cdf stands for the cumulative distribution function. As can be understood from the figure given above, after the resampling, particles with high importance weights will be replicated, since they are assigned the same new indices, while the lower ones will probably assigned no index and die out. The replicated new particles will take equal importance weights of $1/N$ in this new situation. This resampling scheme is explained by the following pseudo-code:

Table 3.1. Pseudo-code of systematic resampling algorithm

$(j) = SR(N, i)$

Generate random number $u \sim \mathcal{U}\left(0, \frac{1}{N}\right)$

s = 0

for $i = 1 : N$

    k = 0

    s = s + $\tilde{w}_t^{(i)}$

    while $s > u$

        k = k + 1

        $u = u + 1/N$

    end

$j(i) = k$

end

Many resampling algorithms have been developed in the literature (Liu and Chen, 1998; Carpenter *et al*., 1999; Bolic *et al*., 2003). However, in our algorithms we have not observed diverse results when different resampling algorithms have been tried. That is why; we made use of the systematic resampling scheme given above.

In the following figure, a single iteration of a particle filter is illustrated where importance sampling and resampling are shown for 10 particles.

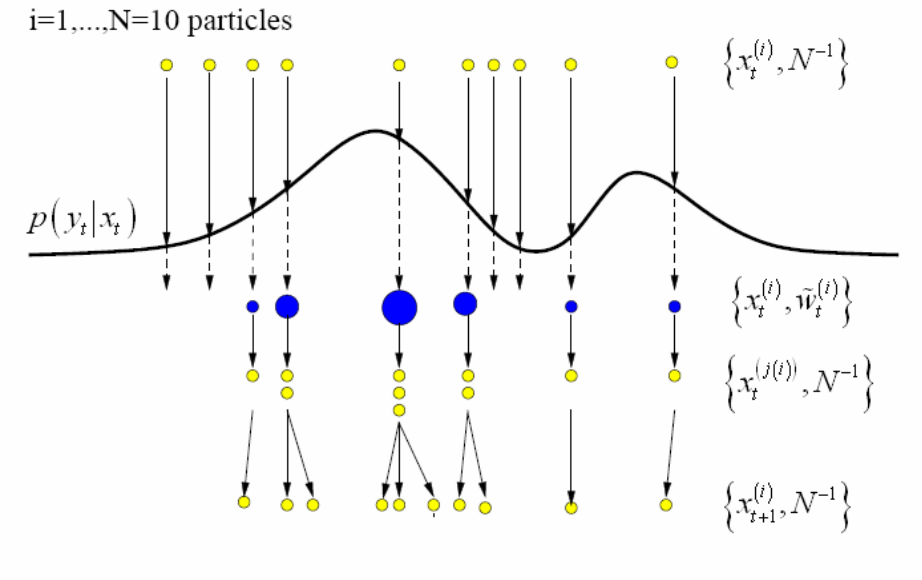


Figure 3.5. A single iteration of a particle filter (Van der Merwe, 2000)

Although the variance of the importance weights can be minimized by the selection of the optimal importance function of (3.37), it is generally tedious to obtain this function in practice, due to its dependency on the observation. Therefore, many approximations to the optimal importance function have been developed in the literature (Doucet *et al*., 2000, Doucet *et al*., 2001). Among these, *a priori state transition pdf* is highly utilized because of its simplicity. This pdf can be easily obtained from the process equation (3.20.a) of the general state-space formulation and it is shown as follows:

$$q\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t}\right) = p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}^{(i)}\right) \tag{3.39}$$

If the optimal importance function is approximated as in (3.39), particle filter is called as the "Bootstrap particle filter" (Gordon *et al*., 1993, Doucet *et al*., 2001) and the importance weight calculation of (3.35) takes the following form:

$$\tilde{w}_t^{(i)} \propto p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) \tilde{w}_{t-1}^{(i)} \tag{3.40}$$

which reads as the evaluation of the particle drawn from $\mathbf{x}_t^{(i)} \sim p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}^{(i)}\right)$ in the likelihood function, i.e. $p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right)$. Here, it should be noted that, if resampling is performed at each iteration, $\tilde{w}_{t-1}^{(i)}$ can be dropped from (3.40) during the importance weight calculation at time $t$, since $\tilde{w}_{t-1}^{(i)} = 1/N$ for each particle. Thus, (3.40) can be expressed as follows in this case:

$$\tilde{w}_t^{(i)} \propto p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) \tag{3.41}$$

which will be utilized in our algorithms.

Although particle filters can be successfully applied to nonlinear, non-stationary and non-Gaussian problems, the functional forms of the system dynamics (functional form of the state transition equation (3.20.a) and the measurement equation (3.20.b)) need to be known. Even if they are known, it is very tedious to sample from the optimal importance function of (3.37) in most cases. Therefore, in order to approximate the optimal importance function, various versions of particle filters have been proposed in the literature in addition to the Bootstrap particle filter given above. Some of these include "auxiliary particle filters" (APF) (Pitt and Shephard, 1999) and "resample-move particle filters" (Gilks and Berzuini, 2001), where the deteriorating effect of the resampling stage is minimized. However, in this thesis we model processes whose state-transition equations (time-variations of the related parameters) are unknown. Additionally, the functional form of the measurement equation is also unknown in mixture modeling cases. Thus, these are highly challenging problems since we have to model the process (in mixture modeling both the process and the measurement) equations. According to our observations, none of the above techniques have been found to be superior to the Bootstrap particle filter, for our problems.

Finally, a typical particle filtering scheme is given as a pseudo-code in the following table, which is followed by the pseudo-code of a Bootstrap particle filter where resampling is performed at each iteration.

Table 3.2. Pseudo-code of a generic particle filter

1. For $i = 1,....., N$ , sample $\mathbf{x}_t^{(i)} \sim q\left(\mathbf{x}_t \middle| \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t}\right)$ and set $\mathbf{x}_{0:t}^{(i)} = \left\{\mathbf{x}_{0:t-1}^{(i)}, \mathbf{x}_t^{(i)}\right\}$

2. For $i = 1,....., N$ , evaluate the importance weights:

$$w_t^{(i)} = \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) p\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}\right)}{q\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t}\right)} w_{t-1}^{(i)}$$

3. For $i = 1,....., N$ , normalize the importance weights:

$$\tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^{N} w_t^{(j)}}$$

4. Evaluate $\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N} (\tilde{w}_t^i)^2}$

5. If $\hat{N}_{eff} \geq N_{threshold}$ , do not resample and keep the particles drawn in step (1),

otherwise

For $i = 1,....., N$ , sample an index $j(i)$ distributed according to the discrete distribution with $N$ elements satisfying $\Pr\left[j(i) = l\right] = \tilde{w}_t^{(l)}$ for $l = 1,....., N$ .

For $i = 1,....., N$ , set $\mathbf{x}_{0:t}^{(i)} = \mathbf{x}_{0:t}^{(j(i))}$ and $w_t^{(i)} = \frac{1}{N}$ .

Table 3.3. Pseudo-code of Bootstrap particle filter with resampling at each iteration

1. For $i = 1,....., N$ , sample $\mathbf{x}_t^{(i)} \sim p\left(\mathbf{x}_t \middle| \mathbf{x}_{0:t-1}^{(i)}\right)$ and set $\mathbf{x}_{0:t}^{(i)} = \left\{\mathbf{x}_{0:t-1}^{(i)}, \mathbf{x}_t^{(i)}\right\}$

2. For $i = 1,....., N$ , evaluate the importance weights:

$$w_t^{(i)} = p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right)$$

3. For $i = 1,....., N$ , normalize the importance weights:

$$\tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^{N} w_t^{(j)}}$$

4. For $i = 1,....., N$ , sample an index $j(i)$ distributed according to the discrete distribution with $N$ elements satisfying $\Pr\left[j(i) = l\right] = \tilde{w}_t^{(l)}$ for $l = 1,....., N$.

For $i = 1,....., N$ , set $\mathbf{x}_{0:t}^{(i)} = \mathbf{x}_{0:t}^{(j(i))}$ and $w_t^{(i)} = \frac{1}{N}$.

As a final remark, degeneracy problem and resampling procedures will be illustrated in a 3-dimensional framework in order to provide a better insight to explain particle filtering. For this purpose, the following nonlinear model will be simulated by using a Bootstrap particle filtering.

**Example 3.1.**

$$\begin{aligned} x_t &= \frac{x_{t-1}}{2} + \frac{25 x_{t-1}}{1 + x_{t-1}^2} + 8\cos\left(1.2t\right) + v_t \\ y_t &= \frac{x_t^2}{20} + n_t \end{aligned} \tag{3.42}$$

where $v_t$ and $n_t$ are zero mean Gaussian random variables with variances $\sigma_v^2 = 10$ and $\sigma_n^2 = 1$, respectively. For this example, a priori state-transition pdf and the likelihood function can be easily obtained as follows:

$$p\left(x_t \middle| x_{t-1}\right) = \mathcal{N}\left(x_t; \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos\left(1.2t\right), 10\right)$$
$$p\left(y_t \middle| x_t\right) = \mathcal{N}\left(y_t; \frac{x_t^2}{20}, 1\right) \tag{3.43}$$

By using the Bootstrap particle filtering, the original state, $x_t$, and its MMSE estimate, $\hat{x}_t$, are shown below:

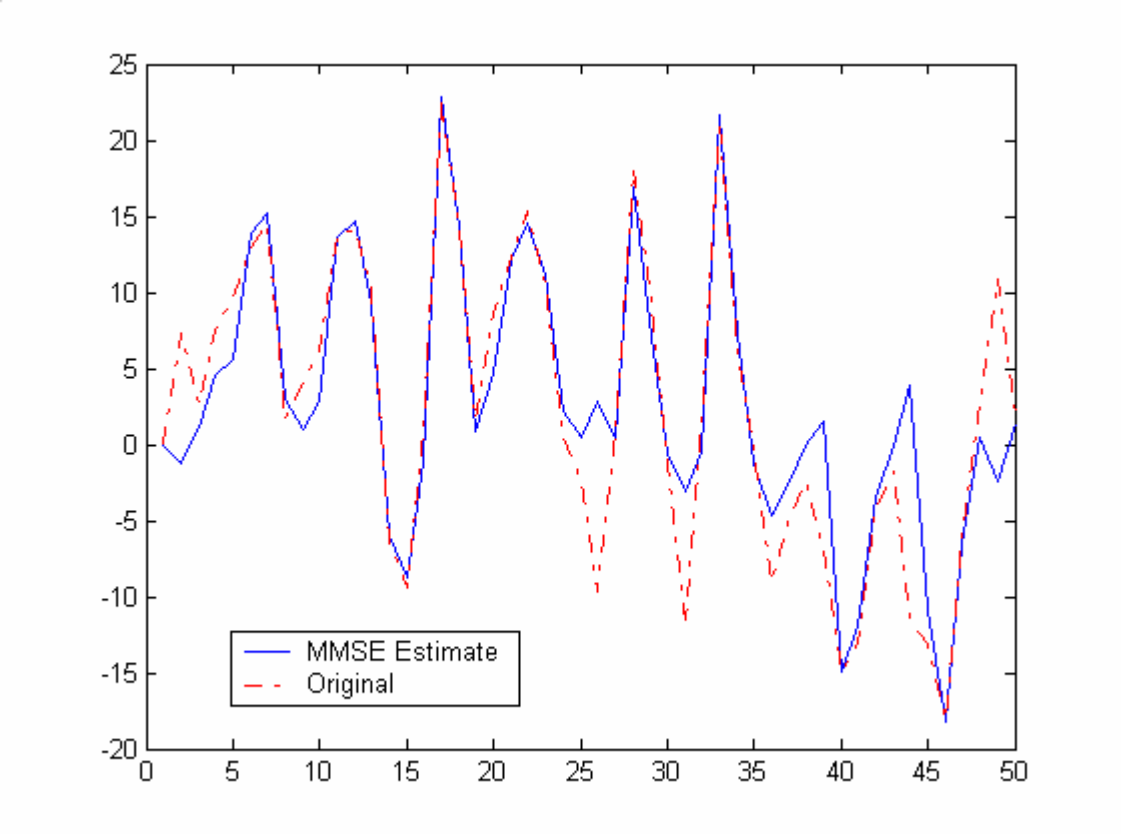


Figure 3.6. Original and estimated states for Example 3.1.

The corresponding 3-dimensional propagation plot of the MMSE estimate is illustrated in the following figure:

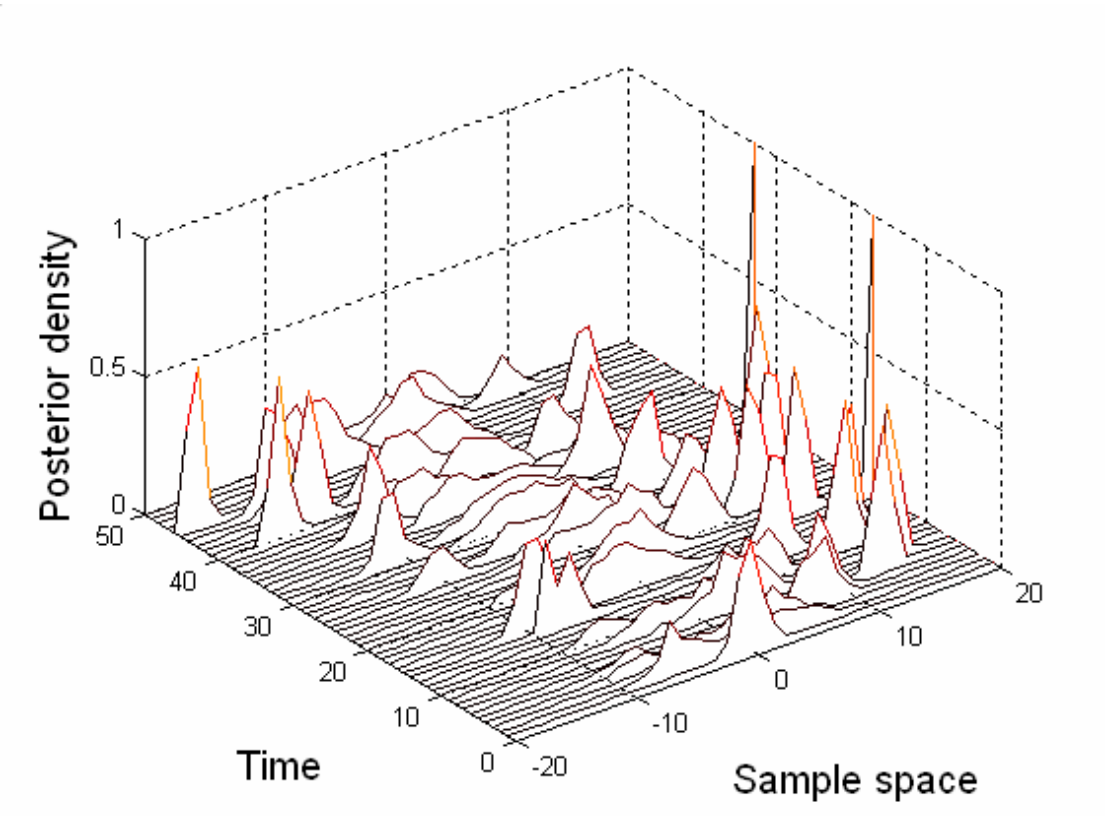


Figure 3.7. Three dimensional posterior density propagation

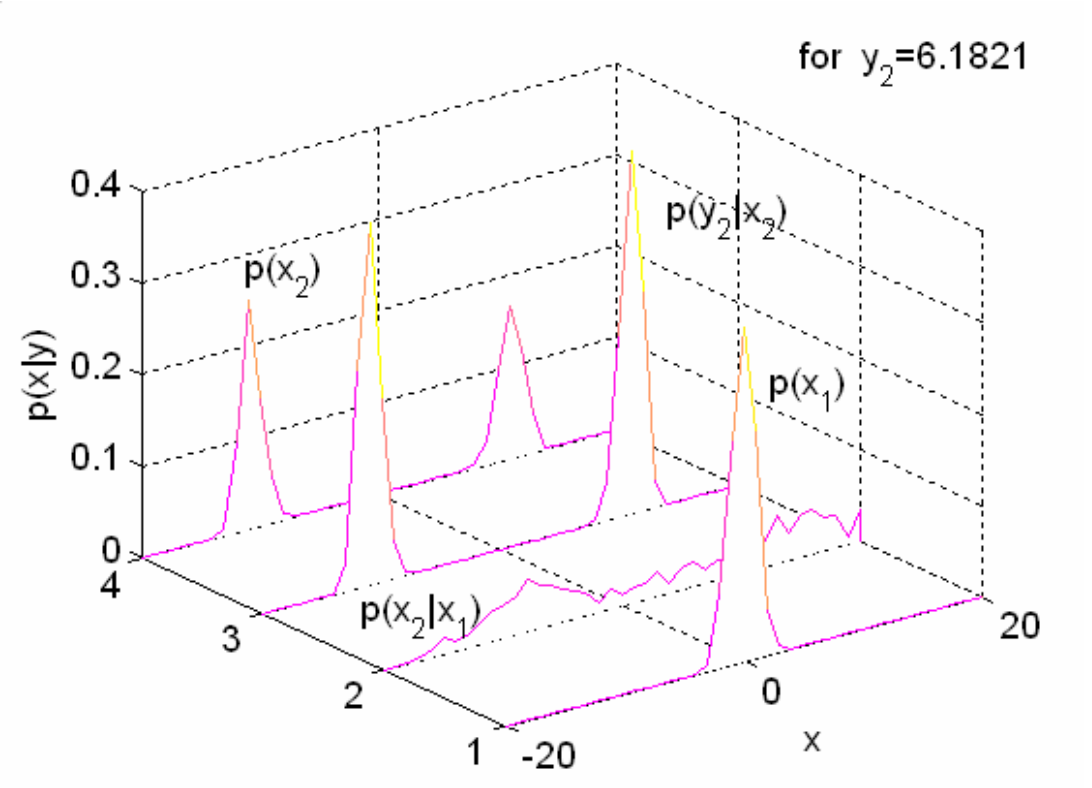


Figure 3.8. One step propagation of the particle filter for an observed $y_2 = 6.1821$

Now, in order to show the degeneracy effect, resampling is not performed in Example 3.1. Again, Bootstrap filtering is used. This is illustrated as follows:

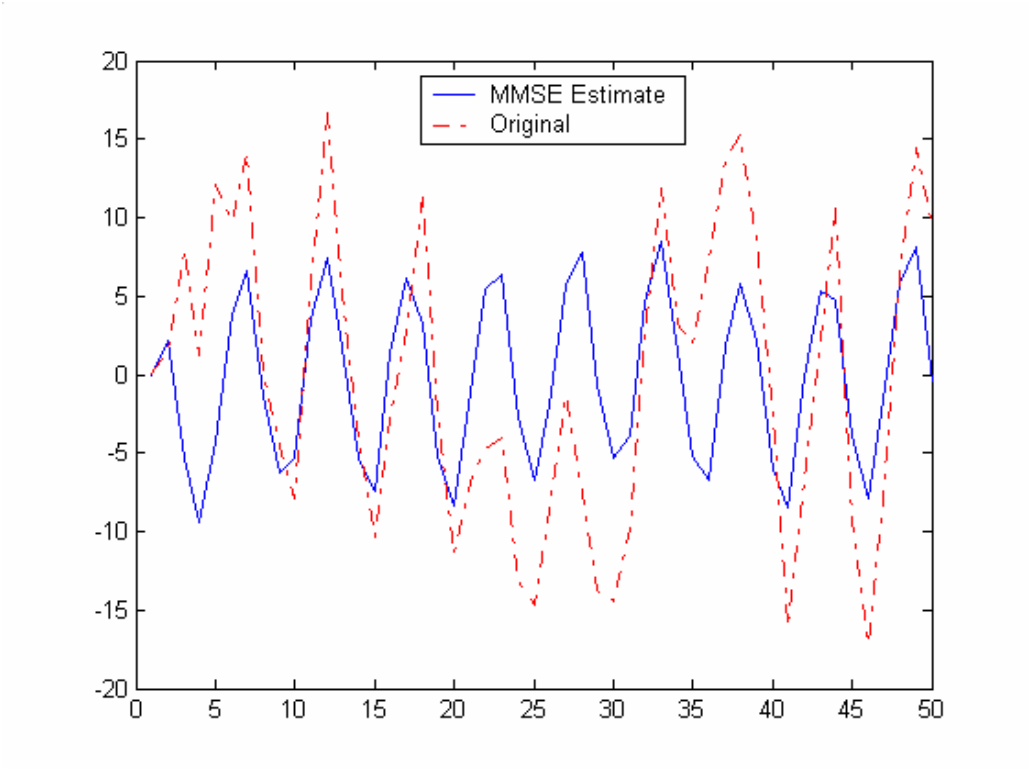


Figure 3.9. Original and estimated states for Example 3.1. without resampling

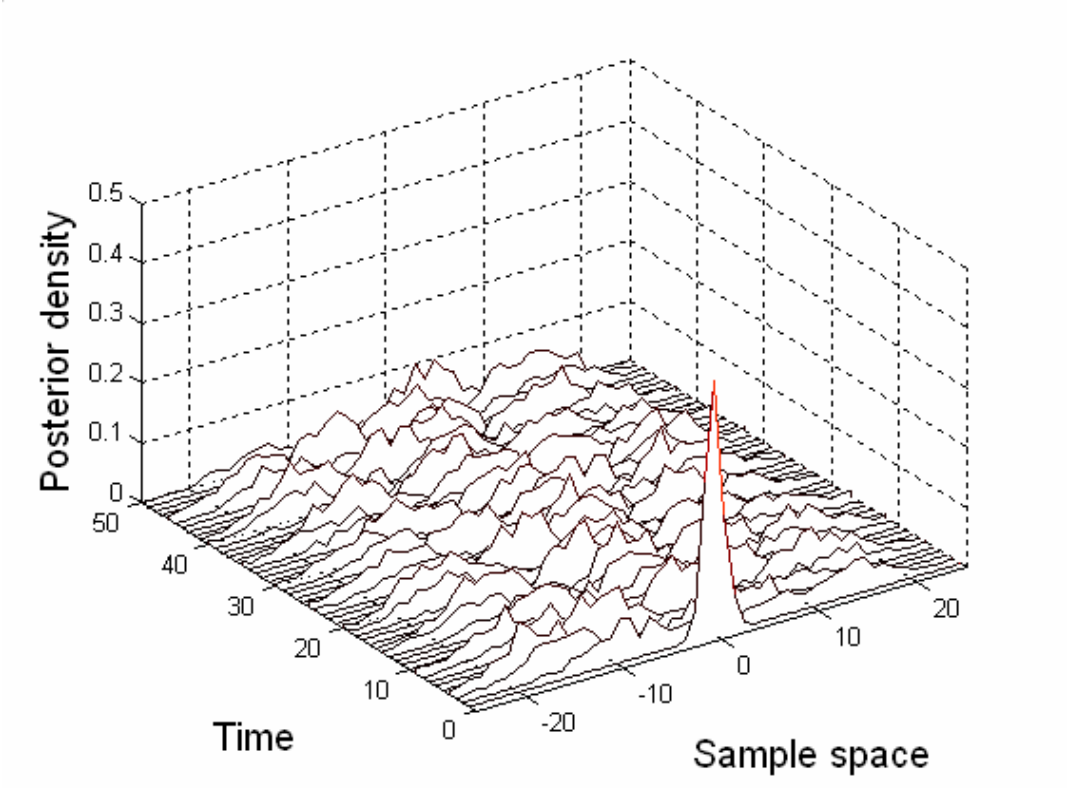


Figure 3.10. Three dimensional posterior density propagation in case of no resampling

In Figure 3.8., one step propagation of a Bootstrap particle filtering is shown. Here, first the initially drawn particles are passed through the state-transition. Then, they are shaped by the observation at $t = 2$ by the likelihood function and then resampling is applied.

In Figures 3.9 and 3.10, it is observed that the resampling procedure is a vital part that should be performed after SIS. Otherwise, desired posterior distributions cannot be estimated due to the diminishing importance weights of the particles as illustrated in Figure 3.10. It is seen that after the first iteration, importance weights become meaningless if no resampling is applied.

# 4. BAYESIAN MODELING OF NON-STATIONARY NON-GAUSSIAN PROCESSES

## 4.1. Introduction

In this chapter, we present our first contribution on the Bayesian modeling of non-stationary non-Gaussian processes. In order to provide a *unifying framework* for modeling such processes, we present a novel methodology for the modeling of TVAR *α-stable* processes which is a direct generalization of Gaussian distributed signals, sharing many common properties such as the stability and CLT (generalized version). Besides, it is well known that *α-stable* processes are widely utilized to model impulsive signals in many areas, such as radar and sonar communications, financial time-series modeling, telecommunications (Nolan and Swami, 1999) and teletraffic data modeling in computer communications (Resnick, 1997). Moreover, time evolution of these data model parameters is also an active research area. It is stated that teletraffic data can be modeled by AR *α-stable* processes and its AR coefficients may have a time-varying nature (Bates and McLaughlin, 1997). Therefore, proposing a unifying modeling methodology for non-stationary *α-stable* processes with time-varying characteristics provides an innovative approach for the solution of possible future applications in this field, since only time-invariant cases have been examined in the literature (Thavaneswaran and Peiris, 1999), according to the best of our knowledge. Possible future applications may include modeling of systems where switching between different *α-stable* processes are involved. Moreover, modeling the long-term dependence of teletraffic data would be of utmost importance in computer communications.

In our methods, we propose the utilization of Bootstrap particle filtering. As mentioned previously, functional forms of the state transitions and measurement equation need to be known to achieve satisfactory performances from any particle filter. However, functional form of the state transition equation is unknown in our problems causing a highly challenging modeling problem. Thus, different models are presented here to form appropriate importance functions.

First, we assume that the distribution parameters of the *α-stable* process are known and only TVAR coefficients are estimated by our technique (Gençağa *et al*., 2005c) which will be named as Direct Sequential Monte Carlo (DSMC) approach in the sequel. For this case, we also present an estimate of the PCRLB which bounds from below the mean squared error of the estimated coefficients for different values. This bound is also used as a benchmark to analyze the performances of the following two methods. A sub-case of this approach was presented in (Gençağa *et al*., 2005a) to model Cauchy processes.

Secondly, we extend our methodology for cases where both the TVAR and *constant* distribution parameters of $S\alpha S$ process are estimated (Gençağa et al., 2006). Here, we propose a Hybrid Sequential Monte Carlo (HSMC) method constituting a two-stage Gibbs sampling algorithm. This method is composed of a particle filter and a Hybrid Monte Carlo (HMC) method, which are used iteratively to estimate the unknown TVAR coefficients and the unknown, constant distribution parameters, respectively.

Finally, we propose a novel technique to model the most general case that can be encountered in the modeling of TVAR *α-stable* processes. In this situation, both AR coefficients and distribution parameters of an *α-stable* process change in time. Here, unknown and TVAR coefficients and distribution parameters are modeled by a single sequential Monte Carlo (SSMC) (particle filter) framework. Moreover, by this approach, skewed *α-stable* processes can also be modeled in addition to the $S\alpha S$ processes. In order to clarify the presentation, a table is presented next, where the aforementioned techniques are summarized:

Table 4.1. Proposed algorithms for non-stationary *α-stable* process modeling

| | |
|---|---|
| Algorithm1 (DSMC) | Estimation of TVAR coefficients under known distribution parameters |
| Algorithm2 (HSMC) | Estimation of both TVAR coefficients and unknown constant distribution parameters of a $S\alpha S$ process |
| Algorithm3 (SSMC) | Estimation of both TVAR coefficients and unknown time-varying distribution parameters of any *α-stable* (also skewed) process |

### 4.2. Bayesian Modeling of Non-Stationary Alpha Stable Processes

In this section, estimation of both the TVAR coefficients and the distribution parameters of an *α-stable* process are performed in order to model non-stationary impulsive signals. These signals are represented by the following equation:

$$y_t = \sum_{k=1}^{K} \phi_t(k)\, y_{t-k} + n_t \rightarrow y_t = \mathbf{y}_{t-1}^T \boldsymbol{\varphi}_t + n_t \tag{4.1}$$

where, $y_t$, $\phi_t(k)$ are known as the observation signal and autoregressive parameters, respectively. Vectors are defined as $\mathbf{y}_{t-1} = [y_{t-1}, ....., y_{t-K}]^T$ and $\boldsymbol{\varphi}_t = [\phi_1(t), ....., \phi_K(t)]^T$. Here, $n_t$ denotes the driving process with distribution $S_\alpha(\gamma, \beta, \mu)$, i.e. $n \sim S_\alpha(\gamma, \beta, \mu)$. $K$ denotes the order of the AR process. Here, the objective is to estimate the TVAR coefficients, $\boldsymbol{\varphi}_t$, which depend on time index $t$ and the distribution parameters of the α-stable process, i.e. $\alpha$, $\gamma$, $\beta$ and $\mu$.

Three different methods are proposed in the sequel:

- Algorithm1 (DSMC): Modeling of TVAR *SαS* processes in case of known distribution parameters $\alpha$ and $\gamma$. $\beta=0$ is taken, since the process is symmetric. Additionally, the location parameter is taken to be zero for the sake of simplicity.

- Algorithm2 (HSMC): Modeling of TVAR *SαS* processes in case of unknown constant distribution parameters $\alpha$ and $\gamma$.

- Algorithm3 (SSMC): Modeling of TVAR *α-stable* processes. Here, all parameters can be estimated, even if all of them are time-varying. This is the most general case that can be encountered.

### 4.2.1. Bayesian Modeling of TVAR *SαS* processes in case of known distribution parameters (DSMC)

Here, we propose a new method, which enables one to sequentially track the time-varying AR parameters of an α-stable process from the observation data. An observed TVAR *SαS* process is expressed by (4.1). This is known as the measurement (observation) equation. In order to estimate TVAR coefficients, these are modeled by a state vector, namely $\mathbf{x}_t$. For the sake of a consistent terminology, general state vector in (3.20) and the particle filtering equations given in Chapter 3 are used here by taking $\mathbf{x}_t = \boldsymbol{\varphi}_t$, whose elements correspond to the AR coefficients in this problem. Here, since there is no information regarding the transition equation of the state vector, an artificial evolution random walk is used to model unknown state-transitions, unlike the known functional forms utilized in (Andrieu and Godsill, 2000; Vermaak *et al*., 2002; Dally and Reilly, 2005). In order to approximate the optimal importance function as good as possible, discounting of old measurements can increase the performance of the algorithm, while drawing samples from the suboptimal importance function (Djuric *et al*., 2001; Djuric *et al*., 2002). Motivated by this, the state-space representation, modeling the dynamic system, can be written as follows:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t$$

which is equivalent to

$$\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_{t-1} + \mathbf{v}_t \tag{4.2.a}$$

$$y_t = \mathbf{y}_{t-1}^T \boldsymbol{\varphi}_t + n_t \tag{4.2.b}$$

where (4.1) is repeated in (4.2.b) for the sake of completeness. (4.2.a) is an artificial random walk model with $\mathbf{v} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{v_t}\right)$.

Motivated by (Djuric *et al*., 2001; Djuric *et al*., 2002), the covariance matrix $\boldsymbol{\Sigma}_{v_t}$ is sequentially estimated from the past as shown below:

$$\boldsymbol{\Sigma}_{v_t} = \boldsymbol{\Sigma}_{\mathbf{x}_{t-1}} \left( \frac{1}{\xi} - 1 \right) \tag{4.3}$$

where $\boldsymbol{\Sigma}_{\mathbf{x}_{t-1}}$ is a diagonal matrix, whose elements are variances of the particles, corresponding to the related AR coefficient at time *t-1* and $\xi$ is a real number between zero and one. This can be expressed by the following equation:

$$\boldsymbol{\Sigma}_{v_t} = diag\left(\sigma_{1,t}^2,...,\sigma_{K,t}^2\right) \tag{4.4}$$

where $\sigma_{k,t}^2 = \left(\frac{1}{\xi}-1\right)\mathrm{var}\left(x_{t-1}(k)\right), \quad k=1,2,...,K$ .

It is stated in (Djuric *et al*., 2001) that the use of such a time-varying covariance matrix provides the current state estimate to be affected more from recent data, while the effect of the previous data is reduced by the utilization of a "forgetting factor", namely $\xi$. This corresponds to making *a priori* information coming from the old measurements vaguer, compared to those coming from the recent ones (Djuric *et al*., 2001). In (Djuric *et al*., 2001), details of this selection are expressed by providing an analogy between the exponentially weighted RLS and the Bayesian methodology. In conclusion, Bootstrap filter is used with the following state transition density:

$$\begin{gathered} q\left(\mathbf{x}_t \left| \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t}\right.\right) = p\left(\mathbf{x}_t \left| \mathbf{x}_{t-1}\right.\right) = \mathcal{N}\left(\mathbf{x}_{t-1}, \boldsymbol{\Sigma}_{v_t}\right) \\ \textit{i.e.} \quad \tilde{\mathbf{x}}_t^{(i)} \sim \mathcal{N}\left(\mathbf{x}_{t-1}^{(i)}, \boldsymbol{\Sigma}_{v_t}\right) \end{gathered} \tag{4.5}$$

where $\tilde{\mathbf{x}}_t^{(i)}$ denotes the drawn samples. Here, it is observed that the optimal importance function is approximated by $p\left(\mathbf{x}_t \left| \mathbf{x}_{t-1}\right.\right)$ which is not conditional to the observation data, i.e. $\mathbf{y}_{1:t}$ . However, by using a time-varying covariance matrix $\boldsymbol{\Sigma}_{v_t}$ , information from the observation data can be obtained implicitly, since the elements of this matrix depend on the variances of the previous AR coefficients, $\mathrm{var}\left(x_{t-1}(k)\right)$, and these variances depend on the observation data.

Given a state-transition equation, which is modeled by (4.2.a) with the proposal density function of (4.5), the importance weight of each particle can be calculated by using

the *likelihood function*, given by (3.41). It is well known that the pdf of a standard *SαS* random variable, such as $n \sim S_\alpha(1,0,0)$ , can be estimated numerically by taking the IFT of its characteristic function which is shown as follows:

$$p(n|\gamma,\alpha) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-|\zeta|^\alpha\right)\exp(jn\zeta)\,d\zeta \tag{4.6}$$

where $\zeta$ denotes the dummy variable and (2.1) is used in its following form, since symmetric distributions are involved here:

$$\varphi(\zeta) = \exp\left\{-\gamma|\zeta|^\alpha\right\} \tag{4.7}$$

In order to calculate the importance weight, pertaining to the $i^{th}$ particle, (4.7) takes the following form as a result of the relationship of (4.1):

$$w_t^{(i)} = p\left(y_t \left| \tilde{\mathbf{x}}_t^{(i)}\right.\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-|\zeta|^\alpha\right)\exp\left(j\left(y_t - \mathbf{y}_t^T\tilde{\mathbf{x}}_t^{(i)}\right)\zeta\right)d\zeta \tag{4.8}$$

After the calculation of the importance weights of each particle by (4.8), for *i = 1, 2,…, N*, these weights are normalized as shown in (3.29) and resampling is performed afterwards. A pseudo-code of this method is given in the following Table.

Table 4.2. Pseudo-code of Algorithm1 (DSMC)

For i =1 to N,

1.INITIATION: Draw samples from the initial distributions of the state variables:

$$\mathbf{x}_0^{(i)} = \boldsymbol{\varphi}_0^{(i)} \sim \mathcal{N}(\mathbf{m}_\phi, \mathbf{P}_\phi)$$

where **m** and **P** denote the mean and covariance matrices of the Gaussian distributions. Note that, $\mathbf{P}_\phi$ is a diagonal matrix

Table 4.2. (continued)

For t = 1 to τ, (τ denotes the data length)

2. STATE TRANSITIONS:

Calculate the variance of each AR coefficient:

$$\text{var}(\phi_k(t-1)) = \frac{1}{N}\sum_{i=1}^{N}\left[\phi_k^{(i)}(t-1) - \frac{1}{N}\sum_{n=1}^{N}\phi_k^{(n)}(t-1)\right]^2, \qquad k = 1,...,K; \quad i,n = 1,2,...,N$$

Calculate the time-varying variances of the state-transition density and form $\mathbf{\Sigma}_{V_t}$ matrix using (4.4) as follows:

$$\sigma_{k,t}^2 = \left(\frac{1}{\xi} - 1\right)\text{var}(\phi_k(t-1)), \qquad k = 1,2,...,K$$

Draw new particles for each state variable by using the proposed state-transition equation:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{v}_t$$

which is equivalent to

$$\boldsymbol{\varphi}_t = \boldsymbol{\varphi}_{t-1} + \mathbf{v}_t$$

$$\tilde{\mathbf{x}}_t^{(i)} \sim \mathcal{N}\left(\mathbf{x}_{t-1}^{(i)}, \mathbf{\Sigma}_{v_t}\right)$$

3. CALCULATE THE IMPORTANCE WEIGHT OF EACH PARTICLE:

$$w_t^{(i)} = p\left(y_t \left| \tilde{\mathbf{x}}_t^{(i)}\right.\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(-|\zeta|^{\alpha}\right)\exp\left(j\left(y_t - \mathbf{y}_t^T\tilde{\mathbf{x}}_t^{(i)}\right)\zeta\right)d\zeta$$

4. NORMALIZE THE WEIGHTS:

$$\tilde{w}_t^i = \frac{w_t^i}{\sum_{i=1}^{N} w_t^i}$$

5. RESAMPLE AND GO TO STEP 2.

In order to analyze the performance of our method, we present an estimate of the PCRLB in the next section which will also be compared with the empirical results.

### 4.2.2. Posterior Cramer Rao Lower Bound

In this section, PCRLB is calculated, which bounds from below the mean-squared estimation errors of the AR coefficients given the true values of the distribution parameters (Tichavsky *et al*., 1998; Ristic *et al.,* 2004). This bound can also be used as a benchmark for analyzing the performances of the proposed methods where the distribution parameters are also unknowns, beside the AR coefficients.

Here, PCRLB is obtained for the dynamical system which is represented by the state-space equations of (4.2), where the driving process has a *SαS* distribution with known parameters, i.e. $n_t \sim S_\alpha(\gamma,0,0)$. It should be noted that $\mathbf{x}_t = \boldsymbol{\varphi}_t$ is utilized in (4.2) in order to keep notation consistent with the general particle filtering terminology given in (3.20). This bound is called as *posterior* CRLB, since the state dynamics is modeled with a nonzero process noise (Ristic *et al.,* 2004). CRLB is defined by the following expression (Tichavsky *et al*., 1998; Ristic *et al.,* 2004):

$$E\left\{\left(\hat{\mathbf{x}}_{t|t}-\mathbf{x}_{t|t}\right)\left(\hat{\mathbf{x}}_{t|t}-\mathbf{x}_{t|t}\right)^T\right\} \geq \mathbf{J}_t^{-1} \tag{4.9}$$

where E(.) and $\mathbf{J}_t$ denotes expectation and Fisher information matrix, respectively. $\mathbf{x}_{t|t}$ notation is used to express the filtered estimate of the state vector in (4.2). Information matrix in (4.9) can be recursively estimated by the following equation (Tichavsky *et al*., 1998; Ristic *et al.,* 2004):

$$\mathbf{J}_{t+1} = \mathbf{D}_t^{22} - \mathbf{D}_t^{21}\left(\mathbf{J}_t + \mathbf{D}_t^{11}\right)^{-1}\mathbf{D}_t^{12} \quad , \quad t>0 \tag{4.10}$$

where

$$\mathbf{D}_t^{11} = -E\left\{\nabla_{\mathbf{x}_t}\left[\nabla_{\mathbf{x}_t}\log p\left(\mathbf{x}_{t+1}\middle|\mathbf{x}_t\right)\right]^T\right\} \tag{4.11}$$

$$\mathbf{D}_t^{21} = -E\left\{\nabla_{\mathbf{x}_t}\left[\nabla_{\mathbf{x}_{t+1}} \log p\left(\mathbf{x}_{t+1}\middle|\mathbf{x}_t\right)\right]^T\right\} \tag{4.12}$$

$$\mathbf{D}_t^{12} = -E\left\{\nabla_{\mathbf{x}_{t+1}}\left[\nabla_{\mathbf{x}_t} \log p\left(\mathbf{x}_{t+1}\middle|\mathbf{x}_t\right)\right]^T\right\} = \left[\mathbf{D}_t^{21}\right]^T \tag{4.13}$$

$$\mathbf{D}_t^{22} = -E\left\{\nabla_{\mathbf{x}_{t+1}}\left[\nabla_{\mathbf{x}_{t+1}} \log p\left(\mathbf{x}_{t+1}\middle|\mathbf{x}_t\right)\right]^T\right\} - E\left\{\nabla_{\mathbf{x}_{t+1}}\left[\nabla_{\mathbf{x}_{t+1}} \log p\left(y_{t+1}\middle|\mathbf{x}_{t+1}\right)\right]^T\right\} \tag{4.14}$$

where the expectations in (4.11)-(4.13) are with respect to $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$, while in (4.14) they are with respect to $\mathbf{x}_t$ , $\mathbf{x}_{t+1}$ and $y_{t+1}$. In addition to these, operator $\nabla_{\mathbf{x}_t} = \left[\frac{\partial}{\partial x_t(1)} \cdots \frac{\partial}{\partial x_t(K)}\right]^T$ is defined. If (4.11)-(4.14) are applied to the state-space equations given in (4.2), following equations can be derived:

$$\mathbf{D}_t^{11} = E\left\{\mathbf{\Sigma}_{v_{t+1}}^{-1}\right\},\ \mathbf{D}_t^{21} = -E\left\{\mathbf{\Sigma}_{v_{t+1}}^{-1}\right\},\ \mathbf{D}_t^{12} = -E\left\{\mathbf{\Sigma}_{v_{t+1}}^{-1}\right\} \tag{4.15}$$

$$\mathbf{D}_t^{22} = E\left\{\mathbf{\Sigma}_{v_{t+1}}^{-1}\right\} - E\left\{\nabla_{\mathbf{x}_{t+1}}\left[\nabla_{\mathbf{x}_{t+1}} \log p\left(y_{t+1}\middle|\mathbf{x}_{t+1}\right)\right]^T\right\} \tag{4.16}$$

Equations in (4.15) are obtained from the process equation of (4.2), where the transition function is an identity matrix. It should be noted that the covariance matrices in (4.15) vary with time, as shown in (4.3). The second term in (4.16) can be obtained numerically, since $p\left(y_{t+1}\middle|\mathbf{x}_{t+1}\right)$ has a *SαS* distribution. This likelihood function can be numerically estimated as illustrated in (4.6). Afterwards, its derivatives can be obtained numerically.

## 4.3. Bayesian Modeling of TVAR *SαS* processes in case of unknown distribution parameters (HSMC)

The objective of this novel method, is to model an observed TVAR *SαS* process $y_t$, i.e. to estimate the vector $\mathbf{x}_t$ and the distribution parameters ($\alpha$ and $\gamma$), which are defined in (4.1). This technique is composed of two successive sections, constituting a Gibbs sampling scheme, where the TVAR coefficients are estimated by particle filter and the

distribution parameters $\alpha$ and $\gamma$ of the *SαS* process are estimated by a Hybrid Monte Carlo method. Again, for the sake of a consistent notation, the general state vector $\mathbf{x}_t$ of the particle filter will be used to denote the TVAR coefficient vector $\boldsymbol{\varphi}_t$, i.e. $\mathbf{x}_t = \boldsymbol{\varphi}_t$ will be used in this section.

If the distribution parameters are represented by a vector $\boldsymbol{\theta} = [\alpha, \gamma]^T$, two iterative stages of the proposed Gibbs sampler can be expressed as follows:

1. Sample the TVAR coefficient vector $\mathbf{x}_{0:\tau} = \boldsymbol{\varphi}_{0:\tau}$ given the current values of the distribution parameters:

$$\mathbf{x}_{0:\tau}^{(m+1)} \sim p\left(\mathbf{x}_{0:\tau} \middle| \boldsymbol{\theta}^{(m)}, y_{1:\tau}\right) \tag{4.17.a}$$

   where $\tau$ denotes the total data length and $\mathbf{x}_{0:\tau} = [\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_\tau]$. Here, *m* designates the iteration number.

2. Sample the distribution parameters given the current value of the TVAR coefficient vector:

$$\boldsymbol{\theta}^{(m+1)} \sim p\left(\boldsymbol{\theta} \middle| \mathbf{x}_{0:\tau}^{(m+1)}, y_{1:\tau}\right) \tag{4.17.b}$$

By sampling from these two conditional distributions iteratively, for a total number of *M* iterations, the algorithm is expected to provide samples from the true joint *a posteriori* distribution of two random parameters, i.e. $p\left(\mathbf{x}_{0:\tau}, \boldsymbol{\theta} \middle| y_{1:\tau}\right)$.

However, it should be noted that the AR coefficients vary in time. Thus, in order to sample $\mathbf{x}_{0:\tau}^{(m+1)} \sim p\left(\mathbf{x}_{0:\tau} \middle| \boldsymbol{\theta}^{(m)}, y_{1:\tau}\right)$, the use of particle filter is proposed. That is, during the current iteration, *m+1*, an estimate of the TVAR coefficient vector, i.e. $\mathbf{x}_{0:\tau}^{(m+1)}$ is obtained by a particle filter, using the available distribution parameters, namely $\boldsymbol{\theta}^{(m)}$.

Then, an estimate of the *SαS* innovations process is obtained as follows:

$$\hat{n}_t = y_t - \mathbf{y}_{t-1}^T \mathbf{x}_t^{(m+1)}, \qquad t=0,1,...,\tau \tag{4.18}$$

After obtaining an estimated waveform for the corresponding *SαS* innovations process, a HMC method (Godsill and Kuruoğlu, 1999) is used to sample the distribution parameters by using the estimated *SαS* innovations process. That is, in order to sample $\boldsymbol{\theta}^{(m+1)} \sim p\left(\boldsymbol{\theta} \left| \mathbf{x}_{0:\tau}^{(m+1)}, y_{1:\tau} \right.\right)$, another modified Gibbs sampler, which will be called Hybrid Monte Carlo Method (HMC) hereafter, is used for a total number of *R* iterations. The details of two successive stages of the proposed method are given as follows:

### 4.3.1. TVAR Estimation by Particle Filtering

In order to sample TVAR sequences at each algorithmic iteration *(m+1)*, the particle filtering scheme explained in Section 4.2. is used given the estimated values of the distribution parameters from the previous iteration *(m).* This sampling scheme is illustared by (4.17.a) and differs from Section 4.2. where the *true* values of the distribution parameters are used instead of their estimates. State-space formulation of (4.2) is also used here and the TVAR coefficients are drawn from the importance function given in (4.5), as in Section 4.2. However, since sampling from (4.17.a) is performed here at a specific algorithmic iteration *(m+1)*, the *estimates of the distribution parameters* are used to calculate the importance weight of each particle as shown below:

$$\begin{aligned} w_t^{(i)} &= p\left(y_t \left| \tilde{\mathbf{x}}_t^{(i)}, \boldsymbol{\theta}^{(m)} \right.\right) = p\left(y_t \left| \tilde{\mathbf{x}}_t^{(i)}, \alpha^{(m)}, \gamma^{(m)} \right.\right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\gamma^{(m)} |\zeta|^{\alpha^{(m)}}\right) \exp\left(j\left(y_t - \mathbf{y}_t^T \tilde{\mathbf{x}}_t^{(i)}\right)\zeta\right) d\zeta \end{aligned} \tag{4.19}$$

After the calculation of the importance weights of each particle by (4.19), for *i = 1, 2,…, N*, these weights are normalized as shown in (3.29) and resampling is performed

afterwards. These particle filtering steps are performed by using the same $\boldsymbol{\theta}^{(m)}$ from $t = 1$ to $t = \tau$ and the conditional posterior of the TVAR coefficients, which are denoted by (4.17.a) are obtained for the $(m+1)^{th}$ iteration of the Gibbs sampler. Then, the estimated TVAR coefficients, $\mathbf{x}_{0:\tau}^{(m+1)}$, are used as inputs in (4.18) to approximate the innovations process, which is used to find the posterior conditional distribution of $\boldsymbol{\theta}^{(m+1)}$. The Hybrid Monte Carlo method used for this purpose is elaborated in the following subsection.

### 4.3.2. Estimation of Distribution Parameters by Hybrid Monte Carlo Method

Here, the method which is developed by (Godsill and Kuruoğlu, 1999) in order to estimate the dispersion parameter of a heavy-tailed symmetric α-stable process is generalized in such a way that the shape parameter $\alpha$ can also be estimated by an additional Metropolis step. The objective of using this method is to obtain samples from the conditional posterior of the distribution parameters, i.e. $\boldsymbol{\theta}^{(m+1)} \sim p\left(\boldsymbol{\theta} \left| \mathbf{x}_{0:\tau}^{(m+1)}, y_{1:\tau} \right.\right)$, given the estimated values of the TVAR coefficients at the $(m+1)^{th}$ iteration of the two-stage Gibbs sampler. As a result of (4.18), the following equality can be written:

$$\boldsymbol{\theta}^{(m+1)} \sim p\left(\boldsymbol{\theta} \left| \mathbf{x}_{0:\tau}^{(m+1)}, y_{1:\tau} \right.\right) = p\left(\boldsymbol{\theta} \left| \hat{n}_{1:\tau} \right.\right) \tag{4.20}$$

where $\hat{n}_{1:\tau} = [\hat{n}_1, \hat{n}_2, ..., \hat{n}_\tau]$. Here, the objective is to draw samples of $\boldsymbol{\theta} = [\alpha, \gamma]^T$ from their joint posterior distributions given by (4.20). Sampling from their joint posterior distribution can be performed by drawing samples from their conditional posteriors iteratively, as shown below:

1. $\gamma^{(r+1)} \sim p\left(\gamma \left| \alpha^{(r)}, \hat{n}_{1:\tau} \right.\right)$ (4.21.a)
2. $\alpha^{(r+1)} \sim p\left(\alpha \left| \gamma^{(r+1)}, \hat{n}_{1:\tau} \right.\right)$ (4.21.b)

Iterative sampling from the conditional posteriors of (4.21.a) and (4.21.b) constitute another Gibbs sampling scheme. Here, samples of $\gamma$ are obtained by drawing samples of $\gamma^2$. Thus, (4.21.a) can be represented by $\gamma^{2(r+1)} \sim p\left(\gamma^2 \middle| \alpha^{(r)}, \hat{n}_{1:\tau}\right)$, which will be used hereafter. Sampling from $\gamma^2$ is due to the eligibility for using conjugate prior for the random variable $\gamma$, which can be chosen to be an Inverted Gamma distribution, i.e. $p\left(\gamma^2\right) = \mathcal{IG}(\eta, \kappa)$. Here $\eta$ and $\kappa$ denote the hyperparameters of the distribution (Gelman *et al*., 1995).

The conditional distributions given above do not possess any closed form analytical expressions in their most general forms. However, with the help of the "product decomposition of the stable densities" explained in Section 2.2., the first conditional distribution can be brought into a form, which can be expressed in closed form, given the value of an additional auxiliary parameter, namely $\lambda$. This is given below for the sake of completeness:

Let $z$ and $\lambda$ be independent random variables with the following distributions:

$$z \sim S_{\alpha'}(\gamma, 0, 0) \qquad \text{and} \qquad \lambda \sim S_{\alpha/\alpha'}\left\{\left(\cos\left(\frac{\pi\alpha}{2\alpha'}\right)^{\alpha'/\alpha}\right), 1, 0\right\} \tag{4.22}$$

Then $\hat{n} = z\lambda^{1/\alpha'}$ is stable with distribution $\hat{n} \sim S_\alpha(\gamma, 0, 0)$. With this notation, a Gaussian distribution with $\mathcal{N}\left(\mu, 2\gamma^2\right)$ corresponds to $S_2(\gamma, 0, \mu)$. So, if $\alpha' = 2$ is chosen and a Gaussian random variable is drawn from $z_t \sim S_2(\gamma, 0, 0)$, given that a random variable $\lambda_t$ is sampled from (4.22), at time $t$, $\hat{n}_t$ is distributed as conditionally Gaussian with distribution $\hat{n}_t \sim \mathcal{N}\left(0, 2\lambda_t\gamma^2\right)$. Moreover, a collection of $\tau$ samples of these $\hat{n}_t \sim \mathcal{N}\left(0, 2\lambda_t\gamma^2\right)$, provide an overall distribution of $\hat{n} \sim S_\alpha(\gamma, 0, 0)$. This is known as a heteroscedastic model (Godsill and Kuruoğlu, 1999) and enables one to express the distribution of a *SαS* random variable as *conditionally Gaussian* with distribution $\hat{n}_t \sim \mathcal{N}\left(0, 2\lambda_t\gamma^2\right)$ at a specific time instant $t$. Thus, by such a model, analytically inexpressible pdf of (4.21.a) can be put into a closed form. So, given the value of $\lambda_t$ at a specific time instant $t$, the likelihood can be

obtained at $t$ by $\mathcal{N}\left(\hat{n}_t \middle| 0, 2\lambda_t\gamma^2\right)$. If the likelihood for $\tau$ samples is considered, the following expression can be obtained:

$$p\left(\hat{n}_{1:\tau} \middle| \boldsymbol{\lambda}, \gamma, \alpha\right) \propto \exp\left(-\frac{1}{2\gamma^2}\sum_{t=1}^{\tau}\frac{\hat{n}_t^2}{\lambda_t}\right) \tag{4.23}$$

where $\boldsymbol{\lambda} = \left[\lambda_1, \lambda_2, ..., \lambda_\tau\right]^T$. This likelihood function in Gaussian form allows us to select conjugate prior for $\gamma^2$. It is known that if this conjugate prior is selected as Inverted Gamma, the posterior of $\gamma^2$ becomes also an Inverted Gamma distribution, thanks to the Gaussian likelihood function (Gelman *et al*., 1995; Godsill and Kuruoğlu, 1999; MacKay, 2003). So, if the *a priori* pdf of $\gamma^2$ is denoted by $p\left(\gamma^2\right) = \mathcal{IG}(\eta, \kappa)$, the posterior is obtained as follows (Godsill and Kuruoğlu, 1999):

$$p\left(\gamma^2 \middle| \alpha, \boldsymbol{\lambda}, \hat{n}_{1:\tau}\right) = \mathcal{IG}\left(\eta', \kappa'\right); \qquad \kappa' = \kappa + \frac{1}{2}\sum_{t=1}^{\tau}\frac{\hat{n}_t^2}{\lambda_t} \qquad \eta' = \eta + \tau/2 \tag{4.24}$$

As a result of these, the conditional posterior in (4.21.a) can now be expressed by a closed form distribution and sampling from this distribution is brought to an easy form. However, such an arrangement cannot be done for the posterior in (4.21.b). Thus, it is proposed to sample from this posterior by a numerical Metropolis algorithm. For this purpose, numerical likelihood functions are evaluated as in (4.19) for each data sample and then multiplied by the prior, as shown below:

$$p\left(\alpha \middle| \boldsymbol{\lambda}, \gamma, \hat{n}_{1:\tau}\right) \propto \prod_{t=1}^{\tau}\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(-\gamma\left|\zeta\right|^{\alpha}\right)\exp\left(j\hat{n}_t\zeta\right)d\zeta\right\}\mathcal{U}(\alpha|0,2) \tag{4.25}$$

which corresponds to the multiplication of the likelihood and the prior chosen as a uniform distribution for the sake of simplicity. In (4.25), $\mathcal{U}(\alpha|0,2)$ denotes the evaluation of $\alpha$ in the uniform distribution $\mathcal{U}(0,2)$ .Here, Random Walk Metropolis (Gelman *et al*., 1995;

MacKay, 2003) is utilized, where the proposal density is chosen to be a Gaussian. The steps of this Metropolis algorithm are given as follows:

1. Draw a new sample: $\alpha' = \alpha + \mathcal{N}\left(0, \sigma_\alpha^2\right)$, where $\sigma_\alpha^2$ denotes the variance of the random jumps (MacKay, 2003).

2. Calculate the acceptance ratio:

$$\varepsilon = \min\left(1, \frac{p\left(\alpha' \middle| \boldsymbol{\lambda}, \gamma, \hat{n}_{1:\tau}\right)}{p\left(\alpha \middle| \boldsymbol{\lambda}, \gamma, \hat{n}_{1:\tau}\right)}\right)$$

3. Accept this new $\alpha'$ with probability $\varepsilon$.

In addition to these, the conditional posterior of the auxiliary variable $\lambda$ should also be incorporated to complete the Gibbs sampling scheme. At a specific time instant *t*, the posterior distribution of $\lambda_t$ is given as follows, which is proportional to the multiplication of the likelihood function and its prior (Godsill and Kuruoğlu, 1999):

$$p\left(\lambda_t \middle| \hat{n}_t, \gamma, \alpha\right) \propto \mathcal{N}\left(\hat{n}_t \middle| 0, 2\lambda_t \gamma^2\right) S_{\alpha/\alpha'}\left\{\left(\cos\left(\frac{\pi\alpha}{2\alpha'}\right)^{\alpha'/\alpha}\right), 1, 0\right\} \tag{4.26}$$

where $\lambda_t$ is sampled from its *a priori* distribution, given in (4.22). Conditional posterior of $\boldsymbol{\lambda}$ is expressed as the multiplication of conditional posterior of each $\lambda_t$:

$$p\left(\boldsymbol{\lambda} \middle| \alpha, \gamma, \hat{n}_{1:\tau}\right) = \prod_{t=1}^{\tau} p\left(\lambda_t \middle| \alpha, \gamma, \hat{n}_t\right) \tag{4.27}$$

Samples from this conditional posterior can be obtained by using the Rejection Sampling scheme, since we cannot obtain a conjugate prior for these parameters. This RS scheme is elaborated next (MacKay, 2003). Here, samples are desired to be drawn from a target distribution which is represented by $p\left(\lambda_t \middle| \alpha, \gamma, \hat{n}_t\right)$. This can be expressed by (4.26), where

the effect of $\lambda_t$ is evaluated at the likelihood, i.e. $\mathcal{N}\left(\hat{n}_t \middle| 0, 2\lambda_t\gamma^2\right)$. So, instead of the posterior, likelihood can be taken as the target distribution. This likelihood is bounded from above (Godsill and Kuruoğlu, 1999):

$$\mathcal{N}\left(\hat{n}_t \middle| 0, 2\lambda_t\gamma^2\right) \leq \frac{1}{\sqrt{2\pi\hat{n}_t^2}} \exp(-1/2) \tag{4.28}$$

Thus, the steps of the rejection sampling can be given as follows (Godsill and Kuruoğlu, 1999):

For t = 1:τ, perform the following:

1. Sample $\lambda_t$ from its *a priori* distribution: $\lambda_t \sim S_{\alpha/2}\left\{\left(\cos\left(\frac{\pi\alpha}{4}\right)^{2/\alpha}\right), 1, 0\right\}$

2. Sample a random variable $u$ from the following uniform distribution:

$$u \sim \mathcal{U}\left(0, \frac{1}{\sqrt{2\pi\hat{n}_t^2}} \exp(-1/2)\right)$$

3. If $u > \mathcal{N}\left(\hat{n}_t \middle| 0, 2\lambda_t\gamma^2\right)$ go to 1.

After obtaining the necessary conditional posteriors, Gibbs sampling is performed by iterating through the following samplings, which enable us to perform the samplings indicated in (4.21):

1. $\boldsymbol{\lambda}^{(r+1)} \sim p\left(\boldsymbol{\lambda} \middle| \hat{n}_{1:\tau}, \gamma^{(r)}, \alpha^{(r)}\right)$ (4.29)

2. $\gamma^{2(r+1)} \sim p\left(\gamma^2 \middle| \alpha^{(r)}, \boldsymbol{\lambda}^{(r+1)}, \hat{n}_{1:\tau}\right)$ (4.30)

3. $\alpha^{(r+1)} \sim p\left(\alpha \middle| \boldsymbol{\lambda}^{(r+1)}, \gamma^{(r+1)}, \hat{n}_{1:\tau}\right)$ (4.31)

After iterating through these for a total of *R* iterations, these values are used as the samples of $\boldsymbol{\theta}^{(m+1)}$, which is indicated in (4.20). Later, $\boldsymbol{\theta}^{(m+1)}$ is inserted as input to the particle filtering part of the two-stage Gibbs sampler in order to sample values of the new TVAR coefficients, i.e. $\mathbf{x}_{0:\tau}^{(m+2)} \sim p\left(\mathbf{x}_{0:\tau} \middle| \boldsymbol{\theta}^{(m+1)}, y_{1:\tau}\right)$ in (4.17.a). These coupled operations are performed for a total number of *M* iterations. Finally, the average of each parameter is taken as the point estimate, which is calculated after the burn-in period (MacKay, 2003):

$$\mathbf{x}_{0:\tau} = \frac{1}{L}\sum_{m=1}^{L}\mathbf{x}_{0:\tau}^{(m)}, \gamma = \frac{1}{L}\sum_{m=1}^{L}\gamma^{(m)}, \alpha = \frac{1}{L}\sum_{m=1}^{L}\alpha^{(m)}, L < M \tag{4.32}$$

where *L* denotes the number of iterations, after the values pertaining to the burn-in period are omitted. A pseudo-code of the proposed method is given in Table 4.2.

Table 4.3. Pseudo-code of Algorithm2 (HSMC)

1. Sample from the initial distributions of the static distribution parameters:

$$\alpha^{(0)} \sim p(\alpha_0), \quad \gamma^{(0)} \sim p(\gamma_0)$$

2. Set *m = 1*,

a) Sample the TVAR coefficient vector, i.e. $\mathbf{x}_{0:\tau}^{(m)} \sim p\left(\mathbf{x}_{0:\tau} \middle| \boldsymbol{\theta}^{(m-1)}, y_{1:\tau}\right)$

Use particle filtering to draw sample of $\mathbf{x}_{0:\tau}$, given the values of

$\boldsymbol{\theta}^{(m-1)} = \left[\alpha^{(m-1)}, \gamma^{(m-1)}\right]^T$:

a1) Set *t = 0*, For *i = 1, …, N*, sample $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0)$, then set *t = 1*.

a2) For *i = 1,…, N*, sample $\tilde{\mathbf{x}}_t^{(i)} \sim p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}\right) = \mathcal{N}\left(\mathbf{x}_{t-1}, \boldsymbol{\Sigma}_{v_t}\right)$ by using (4.5)

a3) For *i = 1,…, N*, calculate the importance weights by (4.19):

$$w_t^{(i)} = p\left(y_t \middle| \tilde{\mathbf{x}}_t^{(i)}, \boldsymbol{\theta}^{(m-1)}\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(-\gamma^{(m-1)}|\zeta|^{\alpha^{(m-1)}}\right)\exp\left(j\left(y_t - \mathbf{y}_t^T\tilde{\mathbf{x}}_t^{(i)}\right)\zeta\right)d\zeta$$

Table 4.3. (continued)

a4) Normalize the importance weights: $\tilde{w}_t^{(i)} = \dfrac{w_t^{(i)}}{\sum_{j=1}^{N} w_t^{(j)}}$

a5) Resample particles

a6) Set $t = t + 1$ and go to *step a2*.

a7) When $t = \tau$, set $\mathbf{x}_{0:\tau}^{(m)} = \mathbf{x}_{0:\tau}$, i.e. $\mathbf{x}_{0:\tau}^{(m)} \sim p\left(\mathbf{x}_{0:\tau} \middle| \boldsymbol{\theta}^{(m-1)}, y_{1:\tau}\right)$

b) Obtain an estimate of the innovations process $\hat{n}_{1:\tau}$ (4.18):

$$\hat{n}_t = y_t - \mathbf{y}_{t-1}^T \mathbf{x}_t^{(m)}, \qquad t=1,...,\tau$$

c) Sample static distribution parameters (4.20), i.e.

$$\boldsymbol{\theta}^{(m)} \sim p\left(\boldsymbol{\theta} \middle| \mathbf{x}_{0:\tau}^{(m)}, y_{1:\tau}\right) = p\left(\boldsymbol{\theta} \middle| \hat{n}_{1:\tau}\right)$$

Define an auxiliary variable $\lambda$, Use Hybrid Monte Carlo method:

Set $r = 1$, For $r = 1$, $\gamma^{(r-1)} = \gamma^{(m-1)}$, $\alpha^{(r-1)} = \alpha^{(m-1)}$

c1) $\boldsymbol{\lambda}^{(r)} \sim p\left(\boldsymbol{\lambda} \middle| \hat{n}_{1:\tau}, \gamma^{(r-1)}, \alpha^{(r-1)}\right)$

c2) $\gamma^{2(r)} \sim p\left(\gamma^2 \middle| \alpha^{(r-1)}, \boldsymbol{\lambda}^{(r)}, \hat{n}_{1:\tau}\right)$

c3) $\alpha^{(r)} \sim p\left(\alpha \middle| \boldsymbol{\lambda}^{(r)}, \gamma^{(r)}, \hat{n}_{1:\tau}\right)$

c4) Set $r = r + 1$ and go to *step c1*.

c5) When $r = R$, set $\boldsymbol{\theta}^{(m)} = \left[\alpha^{(R)}, \gamma^{(R)}\right]^T$

3. Set $m = m + 1$ and go to *step 2a*.

4. When $m = M$, calculate point estimates after the burn-in period (4.32):

$$\mathbf{x}_{0:\tau} = \frac{1}{L}\sum_{m=1}^{L} \mathbf{x}_{0:\tau}^{(m)}, \gamma = \frac{1}{L}\sum_{m=1}^{L} \gamma^{(m)}, \alpha = \frac{1}{L}\sum_{m=1}^{L} \alpha^{(m)} \qquad , \quad L<M$$

For a better understanding of the HSMC method, it is graphically represented by a hierarchical Bayesian network as illustrated below:

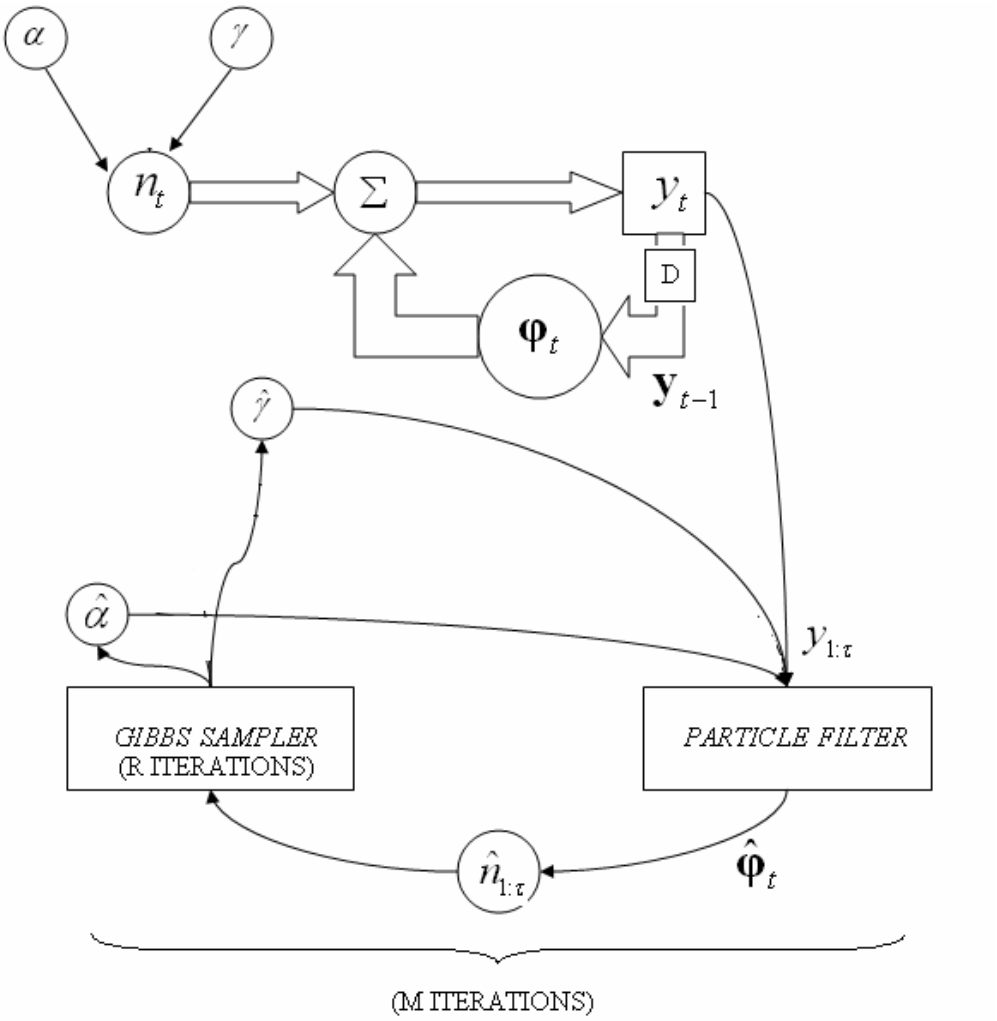


Figure 4.1. Graphical representation of the HSMC method

In this graph, circles and squares denote the unknown and observed quantities, respectively. Above, "D" denotes a delay element. As illustrated above, the nested Gibbs sampler and the particle filter are iterated for *M* times, whereas the inner Gibbs sampler (shown by the block on the lower left) has a total number of *R* iterations. Above, particle filter is used to estimate the TVAR coefficients which are given as inputs to the Gibbs sampler through the reconstruction of the innovations process estimate. Then, Gibbs sampler is used to draw new samples for the TVAR coefficients and this nested two-stage Gibbs sampler is run for a total number of M iterations.

## 4.4. Bayesian Modeling of TVAR *α-stable* processes in case of unknown distribution parameters (SSMC)

Here, a novel method is proposed to model TVAR impulsive signals, which possess skewed or symmetric *α-stable* distributions. The main contribution of this work is its ability to model both the unknown TVAR coefficients and the distribution parameters, where *all of them are time-varying*. This method is a generalization of the previously discussed method where the modeling of unknown TVAR coefficients and *constant* distribution parameters was performed merely for symmetric distributions.

In this approach, both the TVAR coefficients and the distribution parameters of *α-stable* processes are estimated. Such kinds of signals are modeled by (4.1) with a change in the distribution of the stable process, which is given as follows:

$$n_t \sim S_{\alpha_t}\left(\gamma_t, \beta_t, \mu_t\right) \tag{4.33}$$

It should be noted that all distribution parameters are unknown and time-varying in this case, constituting the most general modeling problem. It is assumed that no *a priori* information is known regarding the time-variations of the TVAR coefficients as well as the unknown time-variations of the distribution parameters. Here, the use of Bootstrap particle filter is proposed where the state vector is formed by augmenting the unknown TVAR coefficient vector with the unknown distribution parameters, which is denoted by $\mathbf{x}_t = \left[\boldsymbol{\varphi}_t, \alpha_t, \beta_t, \overline{\gamma}_t, \mu_t\right]^T$. In this modeling, $\overline{\gamma} = \log\gamma$ is used since the dispersion parameter should always take positive values. Since there is no information regarding the transition of the state variables, a random walk model is formed to model the time-evolution of the state variables, as shown below:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{V}_t \tag{4.34}$$

where the process noise vector is represented by $\mathbf{V} = \left[\mathbf{v}_\phi^T, v_\alpha, v_\beta, v_{\overline{\gamma}}, v_\mu\right]^T$ and modeled by a Gaussian distribution, i.e. $\mathbf{V} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{V_t}\right)$. In order to obtain better estimates, the

covariance of the first component of the process noise is modeled to be time-varying, which enables discounting of old measurements during the learning process of the TVAR coefficients. On the other hand, we propose to model variances of the last four terms of the process noise vector by *constants*, since no closed form expressions can be obtained for these terms providing the discounting of old measurements as in (Djuric *et al*., 2001). Following this brief information, the covariance matrix of the process noise is proposed to be in the following form: $\mathbf{\Sigma}_{V_t} = diag\left(\sigma_{1,t}^2,...,\sigma_{K,t}^2,\sigma_{\alpha,t}^2,\sigma_{\beta,t}^2,\sigma_{\bar{\gamma},t}^2,\sigma_{\mu,t}^2\right)$, where diag(.) indicates a diagonal matrix. Here, the first $K$ components of the main diagonal correspond to the variances of the elements of vector $\mathbf{v}_x$, while the last four components denote the variances of $v_\alpha$, $v_\beta$, $v_{\bar{\gamma}}$ and $v_\mu$, respectively. The elements of this covariance matrix are chosen as shown below, according to the aforementioned discussions:

$$\begin{aligned}
&\sigma_{k,t}^2 = \left(\frac{1}{\xi} - 1\right)\mathrm{var}(\phi_k(t-1)), \qquad k = 1,2,...,K \\
&\sigma_{\alpha,t}^2 = \text{constant} \\
&\sigma_{\beta,t}^2 = \text{constant} \\
&\sigma_{\bar{\gamma},t}^2 = \text{constant} \\
&\sigma_{\mu,t}^2 = \text{constant}
\end{aligned} \tag{4.35}$$

Here, $\xi$ denotes the forgetting factor which is chosen between 0 and 1 and $\mathrm{var}(\phi_k(t-1))$ represents the variance of the particles corresponding to the k[th] AR coefficient at time (t-1). As a result of the procedure that is given by (4.34) and (4.35), the state transition model is obtained; resulting in the following state-space formulation of the problem:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{V}_t$$

which is equivalent to (4.36.a)

$$\begin{bmatrix} \boldsymbol{\varphi}_t \\ \alpha_t \\ \beta_t \\ \bar{\gamma}_t \\ \delta_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varphi}_{t-1} \\ \alpha_{t-1} \\ \beta_{t-1} \\ \bar{\gamma}_{t-1} \\ \delta_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_t^{\phi} \\ v_t^{\alpha} \\ v_t^{\beta} \\ v_t^{\bar{\gamma}} \\ v_t^{\delta} \end{bmatrix}$$

$$y_t = \mathbf{y}_{t-1}^T \boldsymbol{\varphi}_t + n_t \quad , \quad n_t \sim S(\alpha_t, \beta_t, \sigma_t, \delta_t) \tag{4.36.b}$$

After forming a state transition model as explained above, particles corresponding to each state variable can be drawn sequentially by using (4.36.a). For the sake of completeness, the general importance weight calculation of (3.35) is shown below:

$$\tilde{w}_t^{(i)} \propto \tilde{w}_{t-1}^{(i)} \frac{p\left(\mathbf{y}_t \left| \mathbf{x}_t^{(i)} \right.\right) p\left(\mathbf{x}_t^{(i)} \left| \mathbf{x}_{t-1}^{(i)} \right.\right)}{q\left(\mathbf{x}_t^{(i)} \left| \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t} \right.\right)} \tag{4.37}$$

Here, Bootstrap particle filtering is used again. Therefore, $q\left(\mathbf{x}_t \left| \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t} \right.\right) = p\left(\mathbf{x}_t \left| \mathbf{x}_{t-1} \right.\right)$ is used as the importance function. Particle sampling is expressed by $\mathbf{x}_t^{(i)} \sim \mathcal{N}\left(\mathbf{x}_{t-1}^{(i)}, \mathbf{\Sigma}_{V_t}\right)$. Given a state-transition model, which is proposed by the procedure described in (4.34) through (4.36.a) with the proposal density function of $p\left(\mathbf{x}_t \left| \mathbf{x}_{t-1} \right.\right)$, the importance weight of each particle can be calculated by using the likelihood function. It is well known that the pdf of an α-stable random variable, denoted by (4.33), can be estimated numerically by taking the IFT of its characteristic function which is shown as follows:

$$\begin{aligned} & p\left(n_t | \alpha_t, \beta_t, \gamma_t, \mu_t\right) \\ & = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{ j\mu_t \zeta - \gamma_t |\zeta|^{\alpha_t} \left[1 + j\beta_t sign(\zeta) \omega(\zeta, \alpha_t)\right] \right\} \exp\left( jn\zeta \right) d\zeta \end{aligned} \tag{4.38}$$

In order to calculate the importance weight, pertaining to the i[th] particle, likelihood function takes the following form as a result of the relationship of (4.1):

$$\begin{aligned} & w_t^{(i)} = p\left( y_t \left| \mathbf{x}_t^{(i)} \right.\right) \\ & = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{ j\mu_t^{(i)} \zeta - \gamma_t^{(i)} |\zeta|^{\alpha_t^{(i)}} \left[1 + j\beta_t^{(i)} sign(\zeta) \omega(\zeta, \alpha_t^{(i)})\right] \right\} \exp\left( j\left( y_t - \mathbf{y}_t^T \boldsymbol{\varphi}_t^{(i)} \right) \zeta \right) d\zeta \end{aligned} \tag{4.39}$$

After the calculation of the importance weight of each particle, these weights are normalized and resampling is performed afterwards. A pseudo-code of the proposed method is given below:

Table 4.4. Pseudo-code of Algorithm3 (SSMC)

For i =1 to N,

1. INITIATION: Draw samples from the initial distributions of the state variables:

$\boldsymbol{\varphi}_0^{(i)} \sim \mathcal{N}(\mathbf{m}_\phi, \mathbf{P}_\phi),\ \alpha_0^{(i)} \sim \mathcal{U}(0,2],\ \beta_0^{(i)} \sim \mathcal{U}[-1,1],\ \overline{\gamma}_0^{(i)} \sim \mathcal{N}(m_{\overline{\gamma}}, P_{\overline{\gamma}}) \rightarrow \gamma_0^{(i)} = \exp\left(\overline{\gamma}_0^{(i)}\right),$

$\mu_0^{(i)} \sim \mathcal{N}(m_\mu, P_\mu)$

where $\mathcal{U}(.)$ denotes uniform distribution, while **m** and **P** denote the mean and covariance matrices of the Gaussian distributions. Note that, $\mathbf{P}_\phi$ is a diagonal matrix whereas $\mathbf{P}_{\overline{\gamma}}$ and $\mathbf{P}_\delta$ are positive scalars.

For t = 1 to τ, (τ denotes the data length)

2. STATE TRANSITIONS:

Calculate the variance of each AR coefficient:

$$\text{var}(\phi_k(t-1)) = \frac{1}{N}\sum_{i=1}^{N}\left[\phi_k^{(i)}(t-1) - \frac{1}{N}\sum_{n=1}^{N}\phi_k^{(n)}(t-1)\right]^2, \quad k = 1,...,K; \quad i,n = 1,2,...,N$$

Calculate the time-varying variances of the state-transition density and form $\boldsymbol{\Sigma}_{V_t}$ matrix using (4.35) as follows:

$$\sigma_{k,t}^2 = \left(\frac{1}{\xi} - 1\right)\text{var}(\phi_k(t-1)), \quad k = 1,2,...,K$$

$$\sigma_{\alpha,t}^2 = \text{constant}, \sigma_{\beta,t}^2 = \text{constant}, \sigma_{\overline{\gamma},t}^2 = \text{constant}, \sigma_{\mu,t}^2 = \text{constant}$$

Draw new particles for each state variable by using the proposed state-transition equation:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{V}_t$$

which is equivalent to

$$\begin{bmatrix} \boldsymbol{\varphi}_t \\ \alpha_t \\ \beta_t \\ \overline{\gamma}_t \\ \delta_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varphi}_{t-1} \\ \alpha_{t-1} \\ \beta_{t-1} \\ \overline{\gamma}_{t-1} \\ \delta_{t-1} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_t^\phi \\ v_t^\alpha \\ v_t^\beta \\ v_t^{\overline{\gamma}} \\ v_t^\delta \end{bmatrix}$$

$$\mathbf{x}_t^{(i)} \sim \mathcal{N}\left(\mathbf{x}_{t-1}, \boldsymbol{\Sigma}_{V_t}\right), \quad i = 1,2,...,N$$

Table 4.4. (continued)

where the following condition must be satisfied during the transition of states, namely $\alpha_t$ and $\beta_t$:

$$p(\alpha_t|\alpha_{t-1}) = \mathcal{N}(\alpha_{t-1}, \sigma^2_{\alpha,t})\mathbb{I}_{(0,2]}(\alpha_t), \quad p(\beta_t|\beta_{t-1}) = \mathcal{N}(\beta_{t-1}, \sigma^2_{\beta,t})\mathbb{I}_{[-1,1]}(\beta_t)$$

where I denotes the indicator function: $\mathbb{I}_{[a,b]}(x) = \begin{cases} 1, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}$

3. CALCULATE THE IMPORTANCE WEIGHT OF EACH PARTICLE:

$$w_t^{(i)} = p\left(y_t \middle| \mathbf{x}_t^{(i)}\right)$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left\{j\mu_t^{(i)}\zeta - \gamma_t^{(i)}|\zeta|^{\alpha_t^{(i)}}\left[1 + j\beta_t^{(i)} sign(\zeta)\omega(\zeta, \alpha_t^{(i)})\right]\right\} \exp\left(j\left(y_t - \mathbf{y}_t^T \boldsymbol{\varphi}_t^{(i)}\right)\zeta\right) d\zeta$$

4. NORMALIZE THE WEIGHTS:

$$\tilde{w}_t^i = \frac{w_t^i}{\sum_{i=1}^{N} w_t^i}$$

5. RESAMPLE AND GO TO STEP 2.

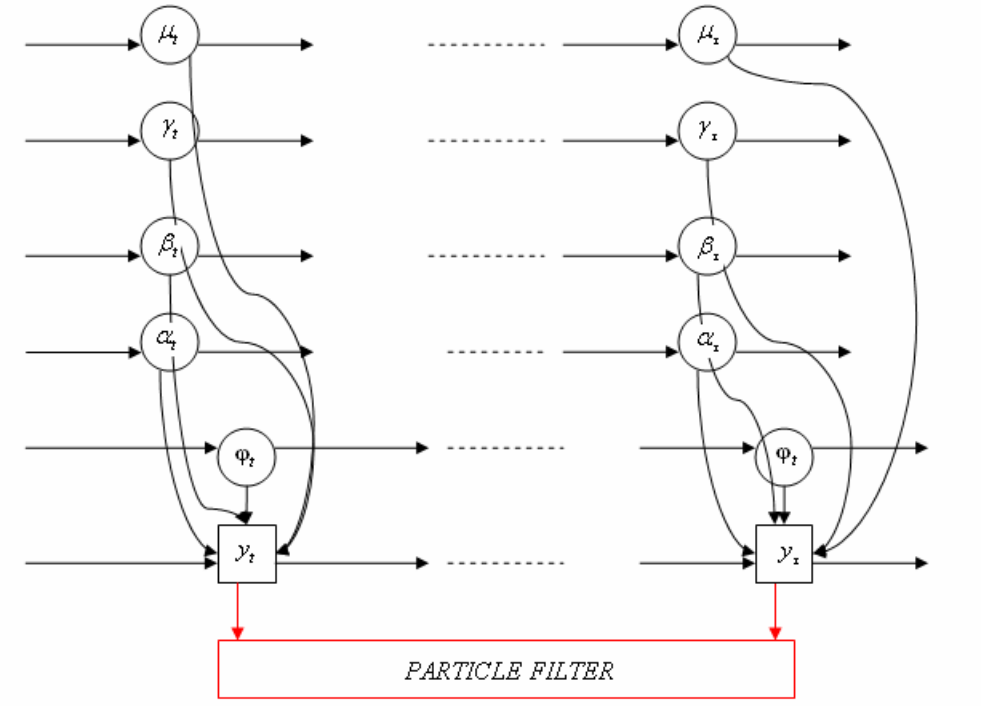


Figure 4.2. Graphical representation of the SSMC methodology and its DBN

To bring a better insight to the SSMC technique, both the Dynamic Bayesian Network (DBN) and the particle filtering algorithm are illustrated in the above hierarchical graph, where circles and squares denote the unknown and known parameters, respectively. Here, every parameter is modeled to be time-varying, unlike the previous case.

## 4.5. Experiments

In this section, experiments justifying the performances of the proposed methods are illustrated. Experiments are given in order pertaining to each method discussed above, which show the performances of the DSMC, HSMC and SSMC methods, respectively.

### 4.5.1. DSMC method

In these simulations, a synthetically generated first order AR process is used, which can be given in the following form:

$$y(t) = \phi(t)\, y(t-1) + n(t) \tag{4.40}$$

where the AR coefficients are time-varying and represented by $\phi(t)$. The driving process n(t) is generated from various α- stable distributions, as explained below. In all cases the distributions are symmetric (β = 0) and standard (γ = 1, δ = 0). For each experiment, 20 ensembles are used in order to estimate the Normalized Mean Square Errors (NMSE), which can be given as follows:

$$NMSE(t) = \frac{\sum_{i=1}^{20} (\hat{x}_i(t) - x(t))^2}{\sum_{i=1}^{20} |\mathbf{x}|^2} = \frac{\sum_{i=1}^{20} (\hat{x}_i(t) - x(t))^2}{\sum_{i=1}^{20} \sum_{t=1}^{1000} x^2(t)} \tag{4.41}$$

where *i* denotes the related ensemble and $\hat{x}_i(t) = \hat{\phi}_i(t)$ and $x(t) = \phi(t)$ denote the Minimum Mean Square Estimate (MMSE) and the original AR coefficients, respectively. In all

experiments, 100 particles and 1000 time samples are used. Two experiments are conducted, where the time-variations of the AR coefficients are taken to be as follows in order to consider as many cases as possible that can be encountered in the physical world.

a) AR coefficient is taken to be 0.99 until the 500$^{th}$ sample, where it changes abruptly to 0.95. This is examined for 4 different α parameters (Figure 4.3).

b) AR coefficient is taken to be changing sinusoidally with time. The effect of various α parameters is examined and the estimated AR trajectories and their instantaneous NMSE are plotted as a function of time (Figure 4.4).

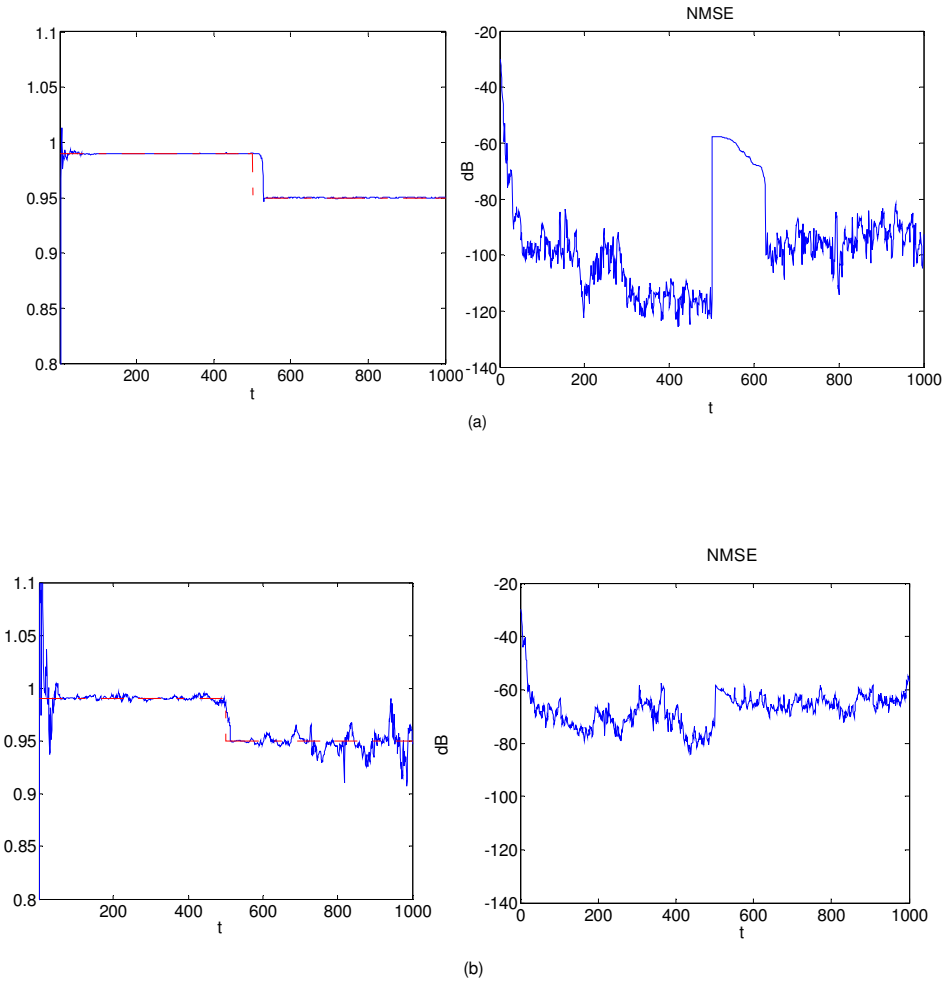


Figure 4.3. Performance of Algorithm1: Estimation of the abruptly varying AR parameter for different α values: a) α = 0.5, b) α = 1

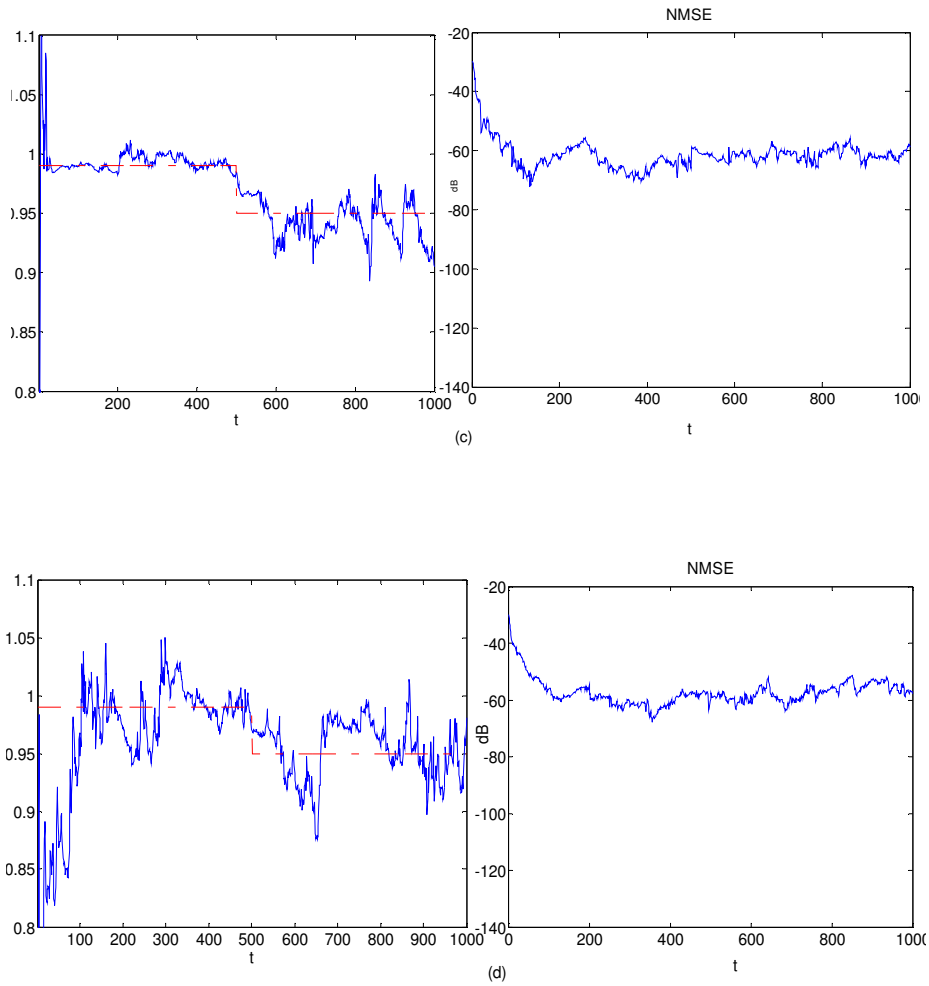


Figure 4.3. (continued) Performance of Algorithm1: Estimation of the abruptly varying AR parameter for different α values: c) α = 1.5, d) α = 2

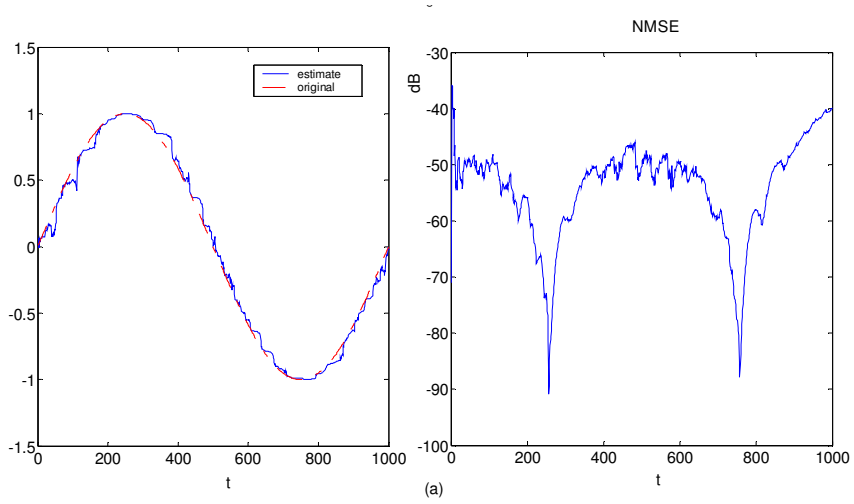


Figure 4.4. Performance of Algorithm1: Estimation of the sinusoidally varying AR parameter for different α values: a) α = 0.5

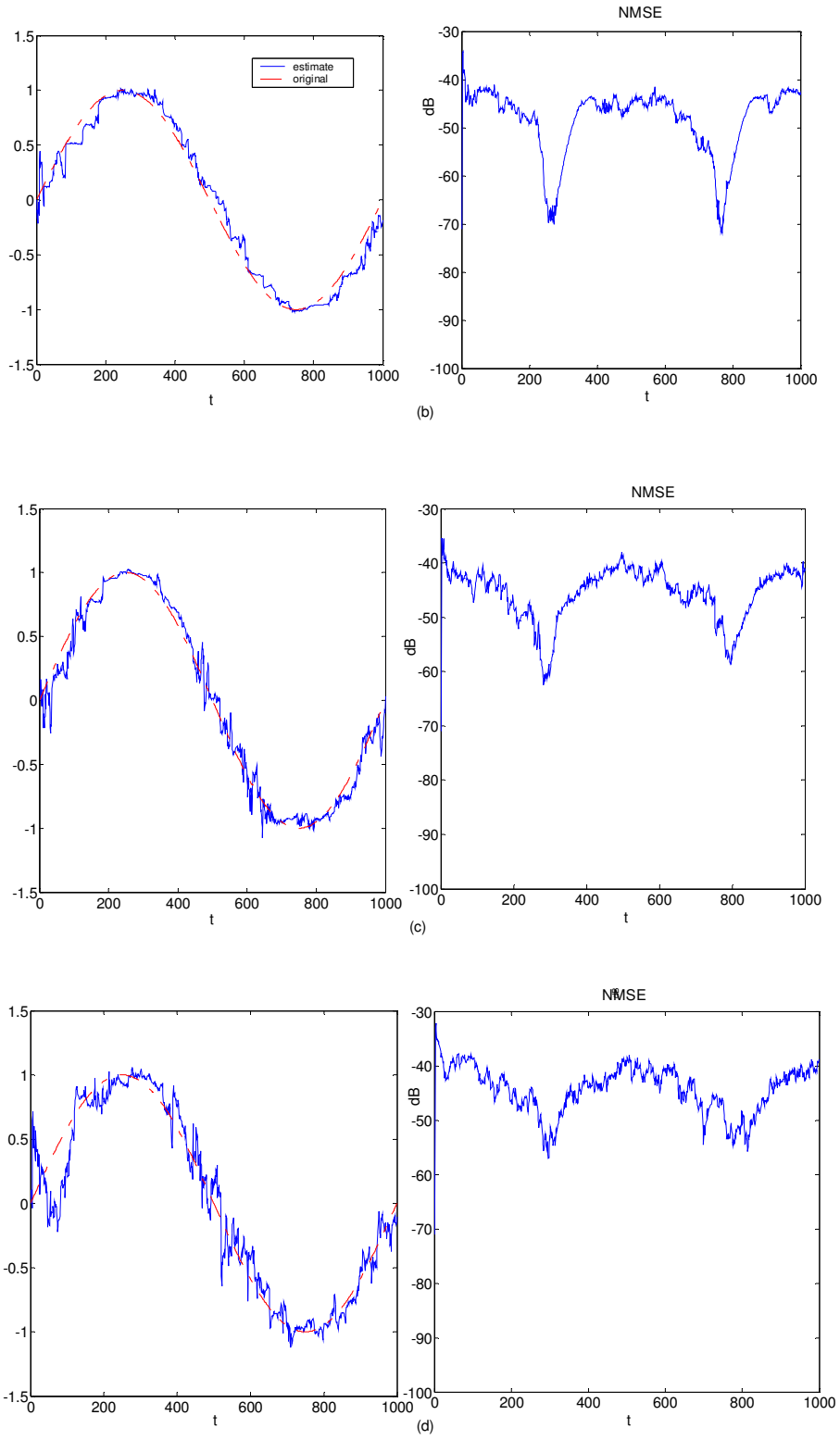


Figure 4.4. (continued) Performance of Algorithm1: Estimation of the sinusoidally varying AR parameter for different α values: b) α = 1, c) α = 1.5, d) α = 2

The performance of the DSMC method is tested for several values of the α parameter as shown above and it is observed to perform very well. It is seen that the quality of the MMSE of the AR coefficients increases as the value of the α decreases, that is, as the process becomes more heavy-tailed. This is illustrated in Figure 4.4. This point is also justified by the PCRLB estimates which will be elaborated next to provide a benchmark for the performance analysis of the proposed techniques. Moreover, in Figure 4.4., it is also noted that the NMSE value at the peaks of the AR waveform decreases significantly. This is due to the slow variation of the AR coefficients throughout these regions. Tracking performance in case of an abruptly changing AR coefficient is shown in Figure 4.3. When Figures 4.3. and 4.4. are compared, it is seen that the quality of the estimates increase as the time variation of the AR coefficients decrease.

### 4.5.2. Posterior Cramer Rao Lower Bound

Here, PCRLB is estimated for twelve different AR coefficients, which are given by [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]. For each AR coefficient, 100 realizations of the related first order AR process are generated by filtering 100 different realizations of a *SαS* innovations process with distribution $n \sim S_{1.5}(1.5,0,0)$. The length of each process is taken to be $\tau = 1000$. Then, particle filter, whose state-space equations are defined by (4.2), is run for each; given the true values of the distribution parameters. Afterwards, the expectations in (4.15) and (4.16) are estimated from the ensemble, for each AR coefficient and $\mathbf{J_t}$ is calculated, recursively by (4.10). Thus, $\mathbf{J_t}$ is obtained for $t = 1{:}\ \tau$. The scalars corresponding to the PCRLB for each AR coefficient is obtained by averaging the values of $\mathbf{J_t}$ in their steady-state regions, which corresponds to the numerical version of (88) and (89) in (Tichavsky *et al.*, 1998) for the case discussed here. In this work, the variance of the state-transition distribution is time-varying, which is taken to be $\sigma_{v_t}^2 = \sigma_{x_{t-1}}^2 \left( \frac{1}{\xi} - 1 \right)$ for a first order AR process. In Figure 4.5.a, the PCRLB of this particle filter is illustrated by a solid line. In addition to this, PCRLB's for the particle filter with state-transition distributions having constant variances are also shown on the same figure, for different values. In order to show the effect of using a time-varying state-transition variance on the estimation of a distribution parameter, PCRLB for another ideal scenario is

shown in Figure 4.5.b, *where the performance of the shape parameter α* is examined, given the true values of the AR coefficient and the dispersion parameter $\gamma$.

In Figure 4.5.a, if the curves corresponding to constant process noise variances are compared with those obtained when a Gaussian AR process is observed (Figure 1 in (Tichavsky *et al.*, 1998)), it can be observed that samples from an α-stable process provide us with more information when compared to a Gaussian one with the same AR coefficient. Therefore, it can be concluded that more accurate estimates of the AR coefficients can be obtained as the process becomes more heavy-tailed, which has also been shown in Figure 4.4.

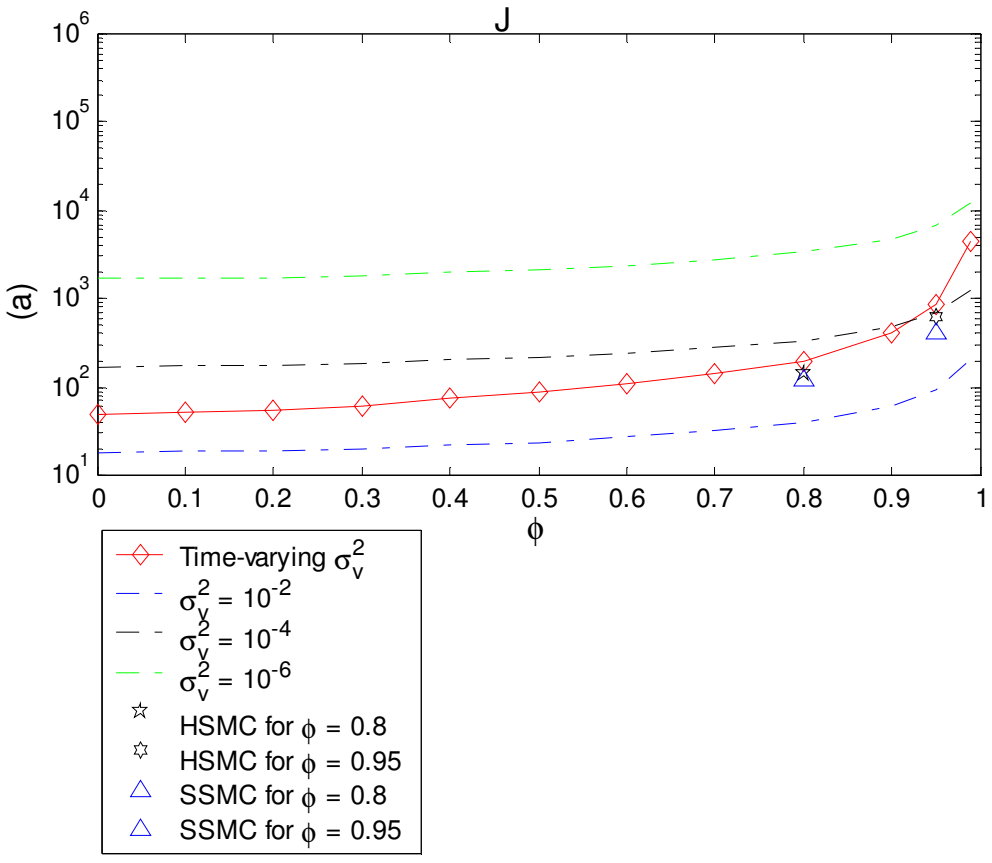


Figure 4.5. a) Fisher information (inverse of PCRLB ) for the ideal scenario ($\phi$ is estimated given the true values of the distribution parameters)

Here, the performances of the HSMC and SSMC are illustrated by "stars" and "triangles", respectively. These points represent the reciprocals of the variance estimates of the AR coefficients $\phi = 0.8$ and $\phi = 0.95$

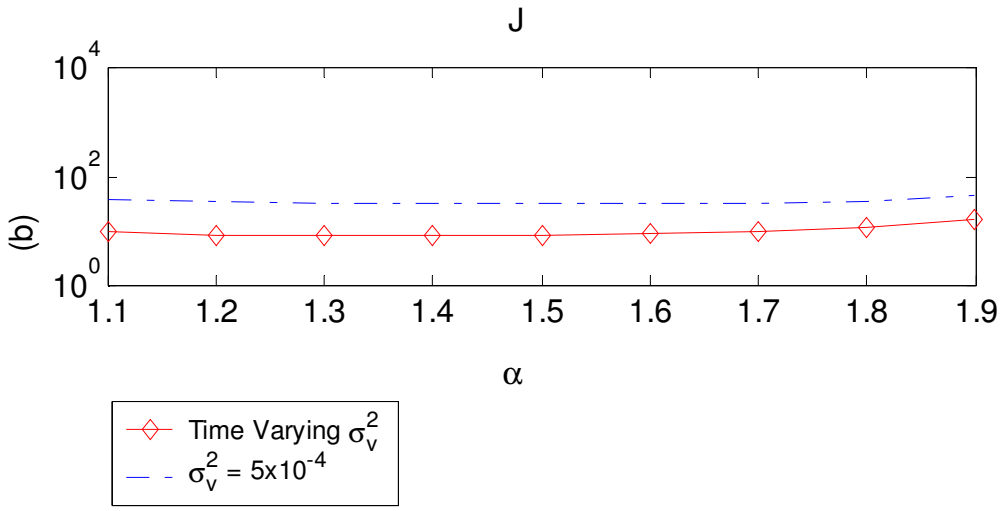


Figure 4.5. b) Fisher information (inverse of PCRLB) for the ideal scenario 2 ($\alpha$ is estimated given the true values of the AR coefficient and $\gamma$)

### 4.5.3. HSMC method

In this experiment, an ensemble of 100 realizations is generated synthetically from a first order TVAR *SαS* process with the following values:

$$y_t = \mathbf{y}_{t-1}^T \boldsymbol{\phi}_t + n_t, \quad n \sim S_\alpha(\gamma,0,0) \tag{4.42}$$

$$\phi_t = \begin{cases} 0.95, & t \le \tau/2 \\ 0.8, & \tau/2 \le t \le \tau \end{cases}, \quad \alpha = 1.5, \quad \gamma = 1.5, \quad \forall t \tag{4.43}$$

where $\tau$ = *1000* is taken for the length of each TVAR process realization.

It should be noted that, $\mathbf{x}_t = \boldsymbol{\varphi}_t$ notation is utilized in this algorithm. Here, (4.3) takes the following form in case of a first order AR model:

$$\sigma_{v_t}^2 = \sigma_{x_{t-1}}^2 \left( \frac{1}{\xi} - 1 \right) \tag{4.44}$$

where $\xi = 0.9$ is used in the simulations. For each realization in the ensemble, the total number of iterations used in the coupled two-stage algorithm is chosen to be *M = 10* which is observed to be sufficient for a steady-state performance. The coupled two-stage algorithm constitutes a Gibbs sampling scheme and it has a nested structure, since the Hybrid Monte Carlo method part of this overall algorithm is also a modified Gibbs sampler. Here, iterations of the overall two-stage Gibbs sampler and the Hybrid Monte Carlo are denoted by *m* and *r*, respectively. In simulations, there are *M* TVAR estimates of length $\tau$ and *M* estimates of distribution parameters of length *R* where *R* is taken 100.

In Figures 4.6.a through 4.6.c, the mean of 100 realizations are illustrated for the estimations of the TVAR coefficient, shape and dispersion parameter, respectively. Corresponding variances are illustrated in Figures 4.6.d through 4.6.f. All of these quantities related to this algorithm are plotted in solid line throughout the figures. From the empirical examinations, it can be concluded that after eight iterations of the overall Gibbs sampler, with 100 iterations run by the Hybrid Monte Carlo method in each Gibbs iteration, steady state values of the related parameters can be sampled. Therefore, the scalar mean and variance values of the TVAR coefficients and the distribution parameters, given in Table 4.5, are obtained by taking the averages at *m = 9* and *m = 10*. In all simulations, *N = 100* particles are used and the stability region of the TVAR coefficients is used as *a priori* information, during the sampling from the *a priori* state transition distributions, i.e. $x_t^{(i)} \sim p\left(x_t \left| x_{t-1}^{(i)}\right.\right)\mathbb{I}_{(-1,1)}$ is used for a first order AR process, where $\mathbb{I}_{(-1,1)} = \begin{cases} 1, & x_t^{(i)} \in (-1,1) \\ 0, & x_t^{(i)} \notin (-1,1) \end{cases}$.

Above, the experiment of estimating the TVAR coefficients and static parameters ($\alpha$ and $\gamma$) is performed by method proposed in Section 4.3. As mentioned previously, this method is capable of estimating static distribution parameters of SαS distributions only. On the other hand, the method proposed in Section 4.4. is developed for the estimation of *time-varying* distribution parameters of general *α-stable* distributions without the need of symmetry constraint. To compare the performances of both methods, the latter algorithm is also tested on the above scenario, explained by (4.42) and (4.43). Following values are chosen to model the artificial state-transition model of the latter algorithm, i.e. equations (4.34) and (4.35) for the above scenario:

$$\sigma_{k,t}^2 = \left(\frac{1}{\xi} - 1\right) \mathrm{var}(\phi_k(t-1)), k = 1 \quad , \xi = 0.9, \sigma_{\alpha,t}^2 = 5\mathrm{x}10^{-4}, \sigma_{\bar{\gamma},t}^2 = 5\mathrm{x}10^{-4} \tag{4.45}$$

In order to compare both methods an ensemble of 100 realizations is used and the initial starting points of both algorithms are set to the same values for a better comparison. This can be better expressed as follows: For each realization, both algorithms are started with the following initial values: $\phi_0^e \sim \mathcal{U}(-1,1)$, $\alpha_0^e = 1.8$, $\gamma_0^e \sim \mathcal{IG}(1,1)$, where *e* denotes the realization number in the ensemble, i.e. *e = 1, 2, …, E = 100.*

The mean of the estimated TVAR coefficient, $\alpha$ and $\gamma$ are superimposed onto Figures 4.6.a to 4.6.c, respectively; whereas the same superimposition is performed for the variances in Figures 4.6.d to 4.6.f. The average values of each mean and variance is tabulated in Table 4.6, which is performed by averaging the related values in the steady-state.

In Figure 4.6.e, the transient part of the variance curve starts from smaller values and increases for the single sequential Monte Carlo method. This is due to the similar convergence structures of each realization of $\alpha$, during the transient parts of their learning curves.

Table 4.5. Steady-state performance of the HSMC method: Mean and variance estimates of the AR coefficients and the distribution parameters

| Parameter | Original Value | Mean | Variance | Accepted steady-state region |
|---|---|---|---|---|
| $\phi$ | 0.95 | 0.9333 | 0.0016 | $300 \le t \le 500$ in *m = 9 & m = 10* |
| | 0.8 | 0.7842 | 0.0069 | $800 \le t \le 1000$ in *m = 9 & m = 10* |
| α | 1.5 | 1.4735 | 0.0045 | *m = 9 & m = 10* |
| γ | 1.5 | 1.4487 | 0.0032 | *m = 9 & m = 10* |

Table 4.6. Steady-state performance of SSMC method: Mean and variance estimates of the AR coefficients and the distribution parameters

| Parameter | Original Value | Mean | Variance | Accepted steady-state region |
|---|---|---|---|---|
| $\phi$ | 0.95 | 0.9308 | 0.0024 | $300 \le t \le 500$ |
| | 0.8 | 0.7844 | 0.0081 | $800 \le t \le 1000$ |
| α | 1.5 | 1.5063 | 0.0266 | $800 \le t \le 1000$ |
| γ | 1.5 | 1.5193 | 0.0350 | $800 \le t \le 1000$ |

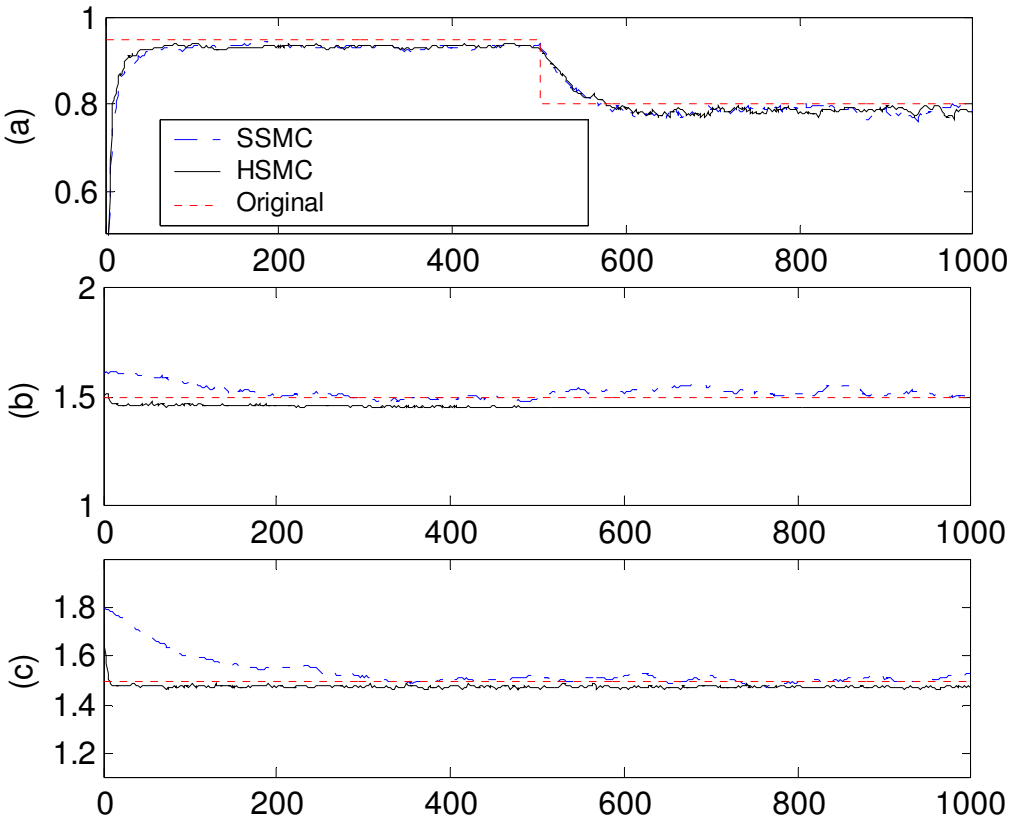


Figure 4.6. Performance of SSMC and HSMC methods: a) Ensemble mean of TVAR coefficient vector $\mathbf{x}_{0:\tau}^{(M)}$ at $m = 10$ and $\mathbf{x}_{0:\tau}$, b) Ensemble mean of $\alpha^{(m)}$, $m = 1,\ldots, 10$ and $\alpha_{0:\tau}$, c) Ensemble mean of $\gamma^{(m)}$, $m = 1,\ldots, 10$ and $\gamma_{0:\tau}$

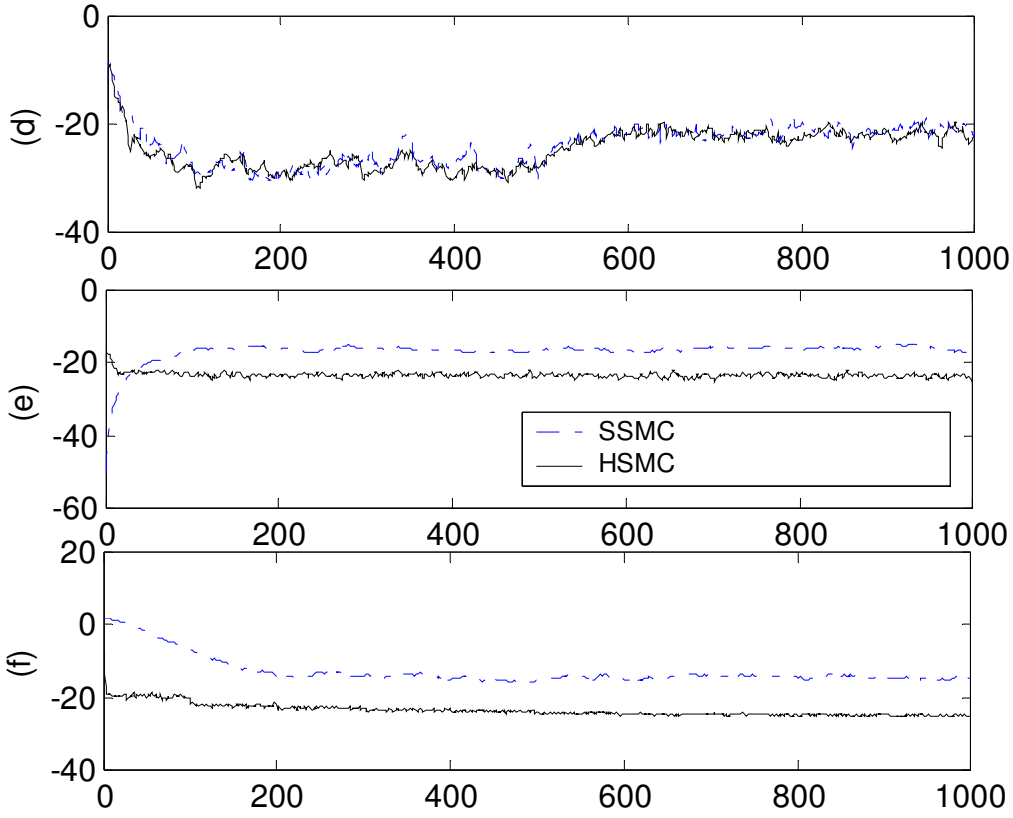


Figure 4.6. (continued): d) Ensemble variance of TVAR coefficient vector $\mathbf{x}_{0:\tau}^{(M)}$ at $m = 10$ and $\mathbf{x}_{0:\tau}$, e) Ensemble variance of $\alpha^{(m)}$, $m = 1,\ldots, 10$ and $\alpha_{0:\tau}$, f) Ensemble variance of $\gamma^{(m)}$, $m = 1,\ldots, 10$ and $\gamma_{0:\tau}$. (y-axes of (d), (e), (f) are expressed in terms of dB)

Above, solid and dashed lines correspond to the curves of the HSMC method and the SSMC technique, respectively. In figures (b), (c), (e) and (f), two indices are used to denote the x-axis: For dashed lines, x-axis denotes $t = 1,\ldots,\tau = 1000$, parameter has no superscripts. For solid lines, x-axis denotes the iteration number $m$ of the overall Gibbs sampler, where values of each $m$ are concatenated for $m = 1,\ldots,M = 10$, parameter has superscript. Each $m$ consists of a total of $R = 100$ iterations of the inner Gibbs sampler, i.e. values corresponding from $r = 1,\ldots, R = 100$, resulting in $MxR = 1000$, which sets the same scale as $\tau = 1000$. In figures (a) through (c), dotted lines show the true values of the parameters.

Here, the methods proposed in Sections 4.3. and 4.4. are compared for the case of unknown TVAR coefficients and constant distribution parameters of a $S\alpha S$ process. The

performances of both algorithms are illustrated in Figure 4.6. and Tables 4.5. and 4.6. When these tables, Figures 4.6.e and 4.6.f. are examined, it is seen that the performances of the AR coefficients are similar. However, HSMC method provides distribution parameters with estimation variances which are one order of magnitude less than those obtained by the SSMC method. So, it is of more advantage to prefer HMSC to SSMC, if the modeled TVAR α-stable process is *symmetric* with *constant* distribution parameters. On the other hand, if the modeled process is not symmetric and its distribution parameters are also *time-varying* in addition to the TVAR coefficients, then the latter technique should be preferred due to its online estimation capability to model the changes of the parameters in time.

For HSMC method, the inverse of the variance estimates of the AR coefficients are 144.9275 and 625 for $x_t = 0.8$ and $x_t = 0.95$, respectively. Similarly, 123.4568 and 416.67 are obtained for the related coefficients when SSMC technique is applied. These values are represented by "stars" and "triangles" in Figure 4.5.a. From these results, it can be concluded that the performance of the proposed methods are close to the PCRLB of the ideal scenario, where the true values of the static distribution parameters are *known*. This is a successful result, since the performance of the estimated AR coefficients are not affected so much, despite the additional burden of the estimation of the unknown distribution parameters.

Moreover, from Figure 4.5.a, it is seen that the information gathered from the observed process increases faster as the process becomes more correlated for the method utilizing a *time-varying variance*. Thus, a better approximation of the optimal importance function is obtained during the estimation of the AR coefficients.

In order to show the effect of using a time-varying state-transition variance on the estimation of a distribution parameter, PCRLB for another ideal scenario is shown in Figure 4.5.b, where the performance of the shape parameter $\alpha$ is examined, given the true values of the AR coefficient and the dispersion parameter $\gamma$. It is observed that the constant state-transition variance value used in the previous section ($5x10^{-4}$) provides a better PCRLB when compared with that of the particle filter using a time-varying variance for the transition of $\alpha$.

### 4.5.4. SSMC method

In this section, different experiments with respect to the time-variation of the AR coefficients and the distribution parameters are performed. 20 realizations of first order TVAR α-stable processes are synthetically generated by (4.1) and (4.33) and ensemble averaged results are illustrated in the following figures for the estimations of the AR coefficients and the distribution parameters, for each scenario.

In these experiments, the objective is to estimate the time waveforms of the following state variables: $\mathbf{x}_t = \left[x_1(t), x_2(t), x_3(t), x_4(t), x_5(t)\right]^T = \left[\phi_t, \alpha_t, \beta_t, \gamma_t, \mu_t\right]^T$ . To measure the estimation performance numerically, the Normalized Mean Square Error (NMSE) of each state variable is also estimated by the following equation:

$$NMSE(t) = \frac{\sum_{i=1}^{20}\left(\hat{x}_{j,i}(t) - x_{j,i}(t)\right)^2}{\sum_{i=1}^{20}\sum_{t=1}^{\tau} x_{j,i}^2(t)}, \qquad j = 1,2,3,4,5 \tag{4.46}$$

where $\hat{x}$ and $i$ denote the estimate and the member index of the waveform in the ensemble, respectively. In all experiments, $\tau$ represents the length of the observed data, which is taken to be 3000. The details of the experiments are illustrated in Table 4.7. with the code of each experiment. Following experiments are illustrated:

a) Sequential Modelling of TVAR SαS Processes

In this section, the proposed method is used to model TVAR SαS processes. Thus, parameter $\beta$ is taken to be zero during the experiments performed here. The objective of performing this scenario is to provide a solution for possible future applications where time-correlations of SαS processes vary in time. The possible application areas include time-varying impulsive signal modeling in wireless communications and time-varying financial data models used by corporations.

- Experiment A1:

In this experiment, the TVAR coefficient, the shape and dispersion parameters of the α-stable process are taken to be piecewise constants in time, where all of them change their

values at t = $\tau$/2. By such a scenario, switching between different two processes or signal changes in wireless communications can be simulated. This is illustrated below:

$$\phi_t = \begin{cases} 0.9 & t < \tau/2 \\ 0.5 & t \geq \tau/2 \end{cases}, \alpha_t = \begin{cases} 1.5 & t < \tau/2 \\ 1.1 & t \geq \tau/2 \end{cases}, \gamma_t = \begin{cases} 2 & t < \tau/2 \\ 5 & t \geq \tau/2 \end{cases}, \beta = 0, \delta = 0 \qquad (4.47)$$

Following values are used here:

$$\sigma_{k,t}^2 = \left(\frac{1}{\xi} - 1\right) \mathrm{var}(\phi_k(t-1)), k = 1 \quad , \xi = 0.9, \sigma_{\alpha,t}^2 = 5\mathrm{x}10^{-4}, \sigma_{\bar{\gamma},t}^2 = 5\mathrm{x}10^{-4} \qquad (4.48)$$

In Figure 4.7., estimates of $\phi_t$, $\alpha_t$ and $\gamma_t$ are shown, respectively. Then, their corresponding NMSE curves are illustrated.

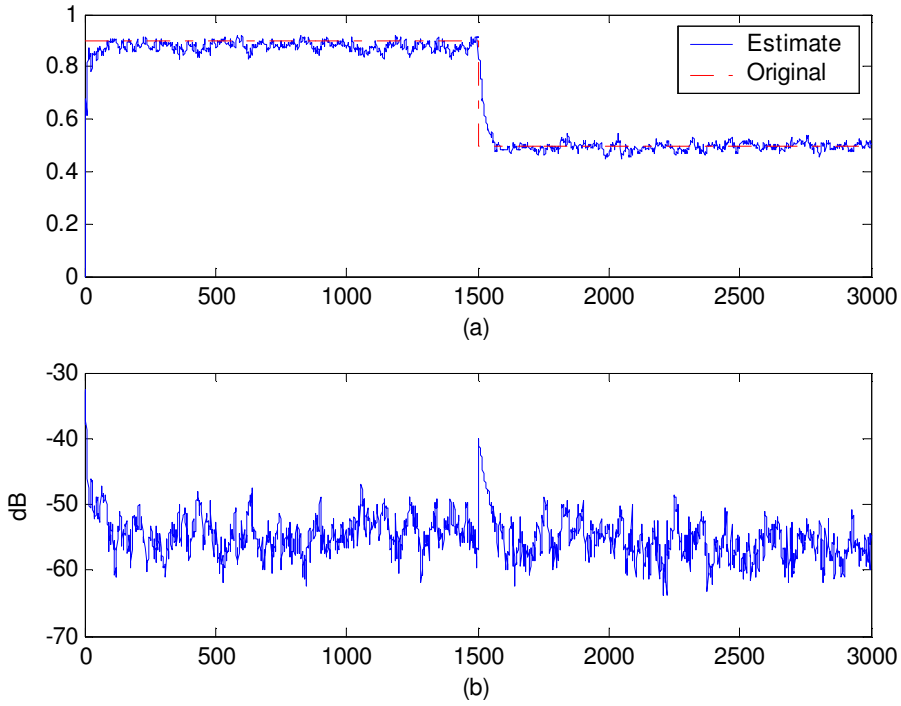


Figure 4.7. Experiment A1: Estimation of time-varying AR coefficient and distribution parameters of *SαS* process: a) Estimation of TVAR coefficient, b) NMSE curve of TVAR estimate

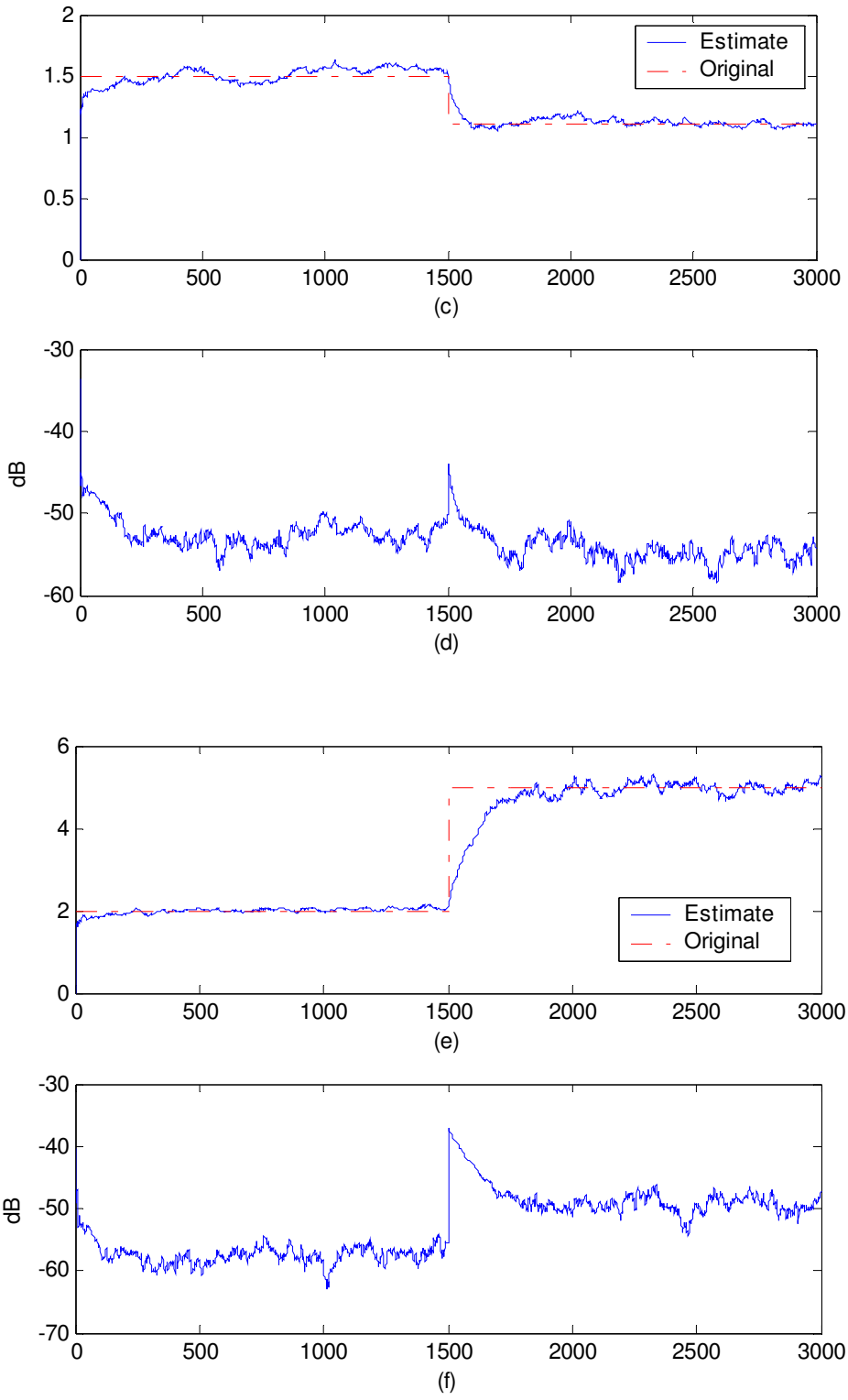


Figure 4.7. (continued): c) Estimation of the shape parameter, d) NMSE curve of the shape parameter estimation, e) Estimation of the dispersion parameter, f) NMSE curve of the dispersion parameter estimation

- Experiment A2:

In this experiment, AR coefficient and the shape parameter are chosen to be piecewise constant as in Experiment 1, while the variation of the dispersion parameter is considered to be sinusoidally changing in time, as shown below:

$$\gamma_t = 1.5 + \sin\left(\frac{2\pi t}{\tau}\right) \tag{4.49}$$

By considering such a scenario, modeling of both abrupt and smooth variations is simulated on different parameters that can be encountered in signal changes due to movement in wireless communications and time-varying financial time-series models of corporations. In Figure 4.8., estimates of $\phi_t$, $\alpha_t$ and $\gamma_t$, and their corresponding NMSE curves are illustrated.

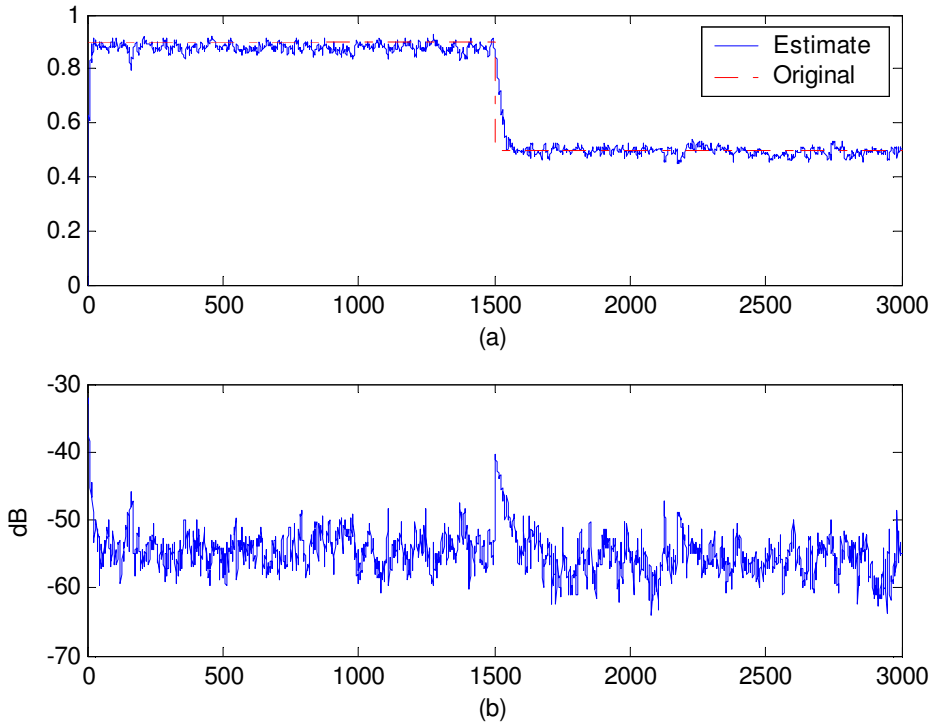


Figure 4.8. Experiment A2: Estimation of time-varying AR coefficient and distribution parameters of *SαS* process a) Estimation of TVAR coefficient, b) NMSE curve of TVAR estimate

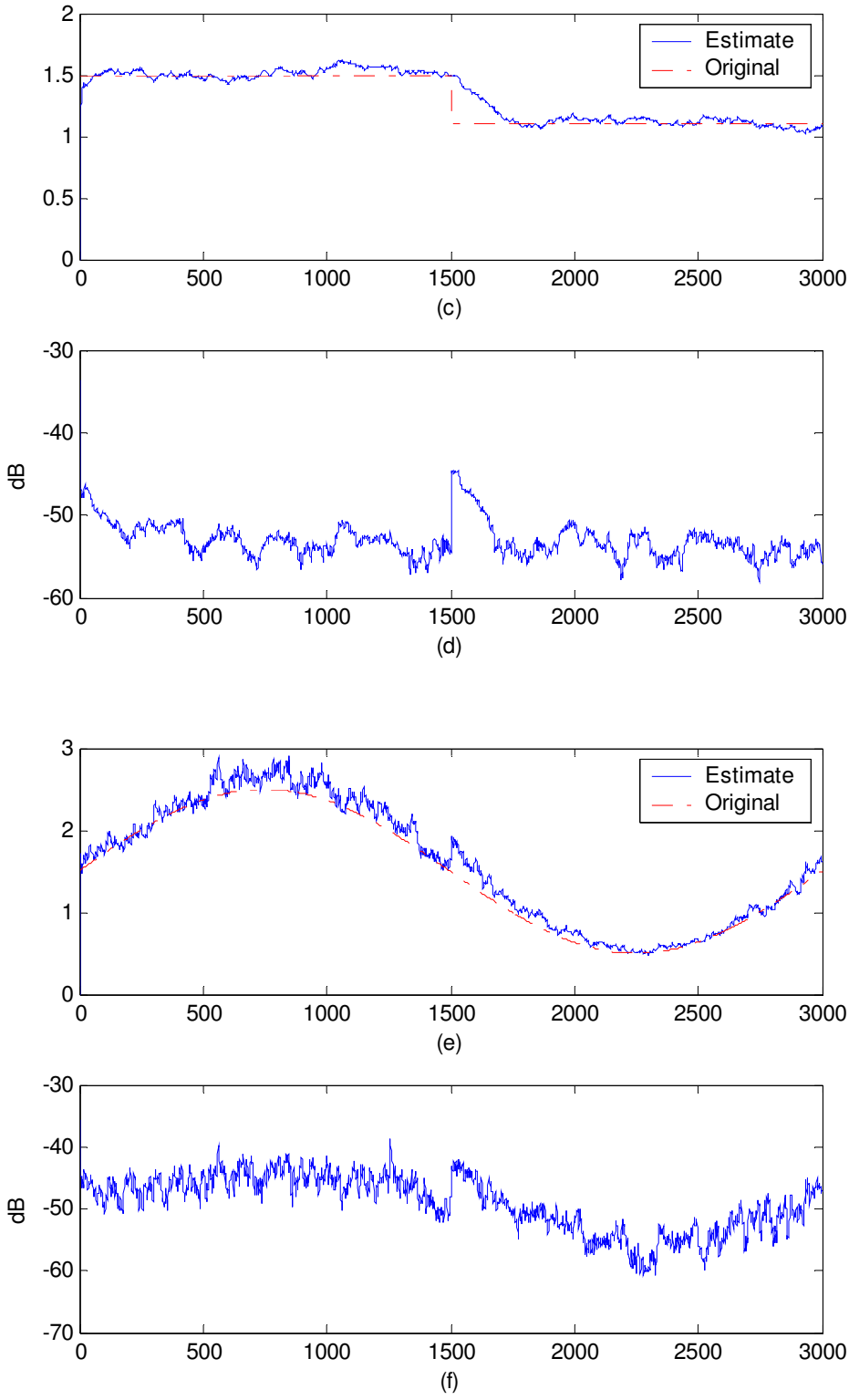


Figure 4.8. (continued): c) Estimation of the shape parameter, d) NMSE curve of the shape parameter estimation, e) Estimation of the dispersion parameter, f) NMSE curve of the dispersion parameter estimation

b) Sequential Modelling of TVAR Skewed α-stable Processes

In this section, the proposed method is used to model TVAR Skewed α-stable processes. Thus, parameter $\beta$ is not zero during the experiments and changes in time. This scenario can be utilized for the modeling of teletraffic data in future applications, since it is believed that such a data may be time-varying beside its skewed distribution.

- Experiment B1:

In this experiment, the TVAR coefficient, the shape and dispersion parameters of the α-stable process are as in Experiment A2. Additionally, the skewness parameter is also taken to be time-varying as shown below:

$$\beta_t = \begin{cases} 0.5 & t < \tau/2 \\ -0.5 & t \geq \tau/2 \end{cases} \tag{4.50}$$

In Figure 4.9., estimates of $\phi_t$, $\alpha_t$, $\beta_t$ and $\gamma_t$, and their NMSE curves are illustrated.

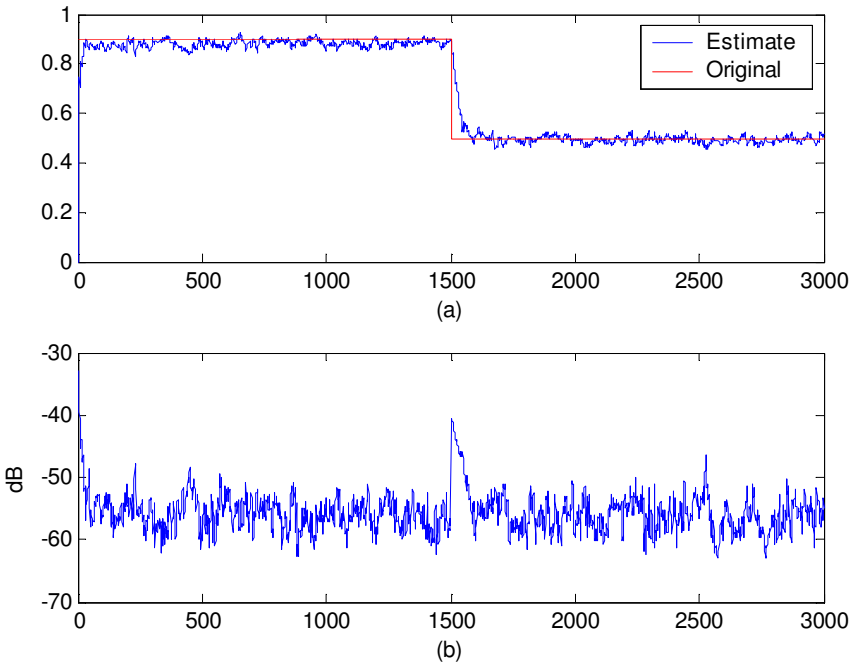


Figure 4.9. Experiment B1: Estimation of time-varying AR coefficient and distribution parameters of *Skewed α-stable* process a) Estimation of TVAR coefficient, b) NMSE curve of TVAR estimate

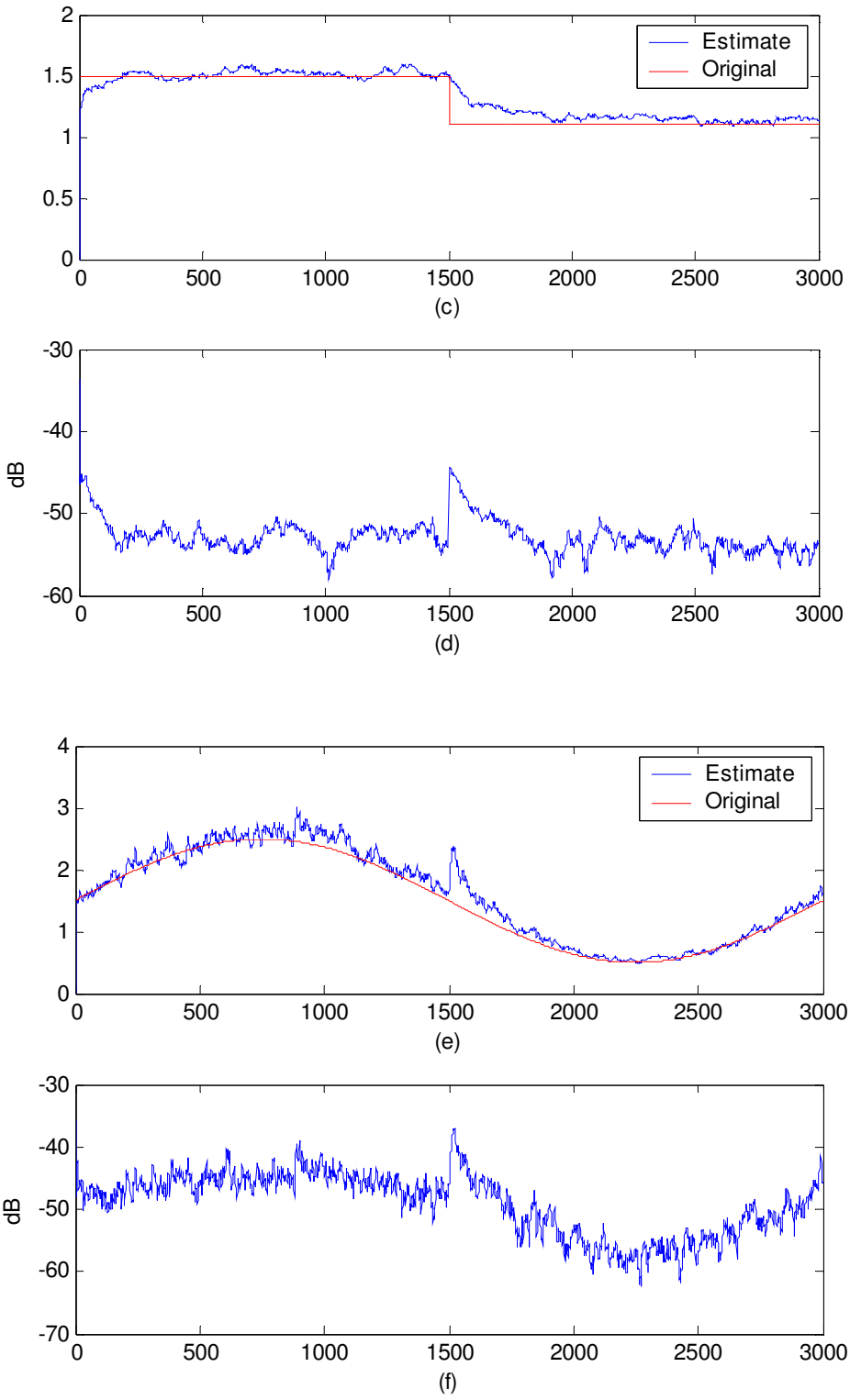


Figure 4.9. Experiment B1: Estimation of time-varying AR coefficient and distribution parameters of *Skewed α-stable* process, c) Estimation of the shape parameter, d) NMSE curve of the shape parameter estimation, e) Estimation of the dispersion parameter, f) NMSE curve of the dispersion parameter estimation

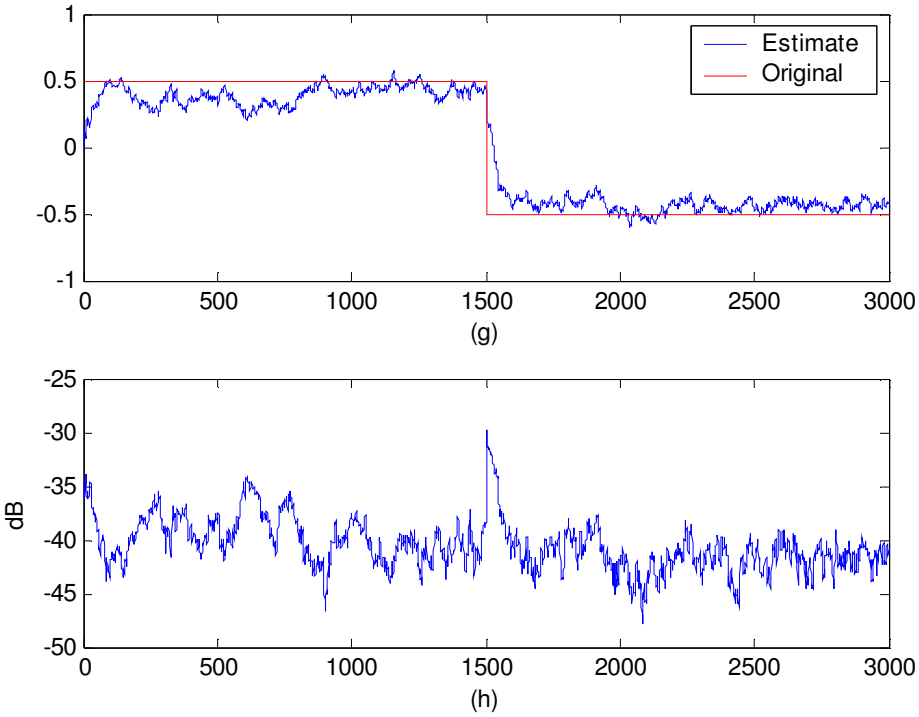


Figure 4.9. (continued): g) Estimation of the skewness parameter, h) NMSE curve of the skewness parameter estimation

- Experiment B2:

In this experiment, the TVAR coefficient, the shape, dispersion and skewness parameters of the α-stable process are taken as in the previous experiment. Moreover, the location parameter is also modeled here, where it is taken to be a ramp. This scenario is considered for cases where the distribution of the process changes location in time due to the movement in wireless communications, beside the other time-varying parameters that can vary as a result of time-varying channels. This scheme is illustrated below:

$$\mu = \frac{1}{\tau} t \tag{4.51}$$

In Figure 4.10., estimates of $\phi_t$, $\alpha_t$, $\beta_t$, $\gamma_t$ and $\mu_t$, and their corresponding NMSE curves are illustrated.

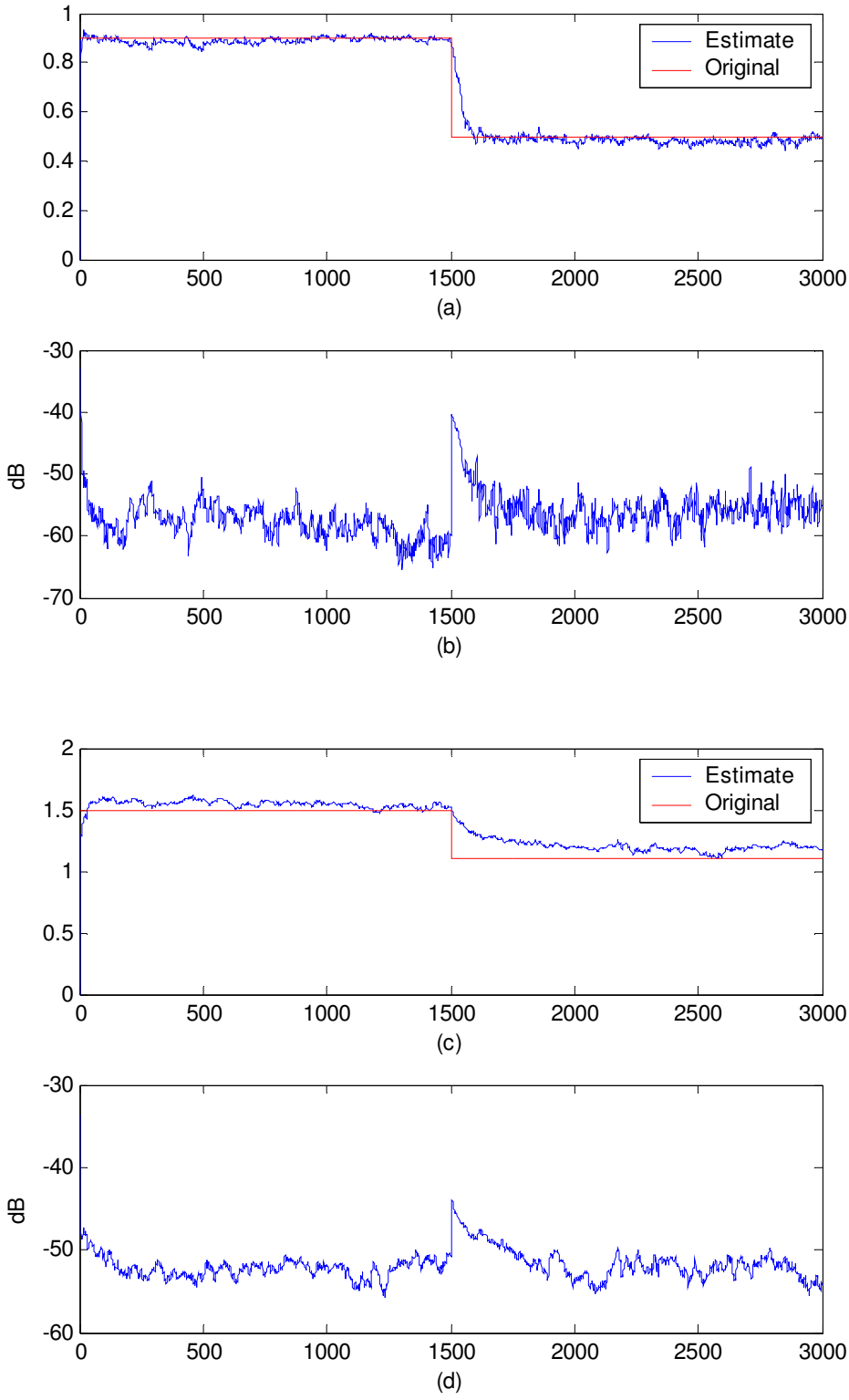


Figure 4.10. Experiment B2: Estimation of time-varying AR coefficient and distribution parameters of *Skewed α-stable* process a) Estimation of TVAR coefficient, b) NMSE curve of TVAR estimate, c) Estimation of the shape parameter, d) NMSE curve of the shape parameter estimation

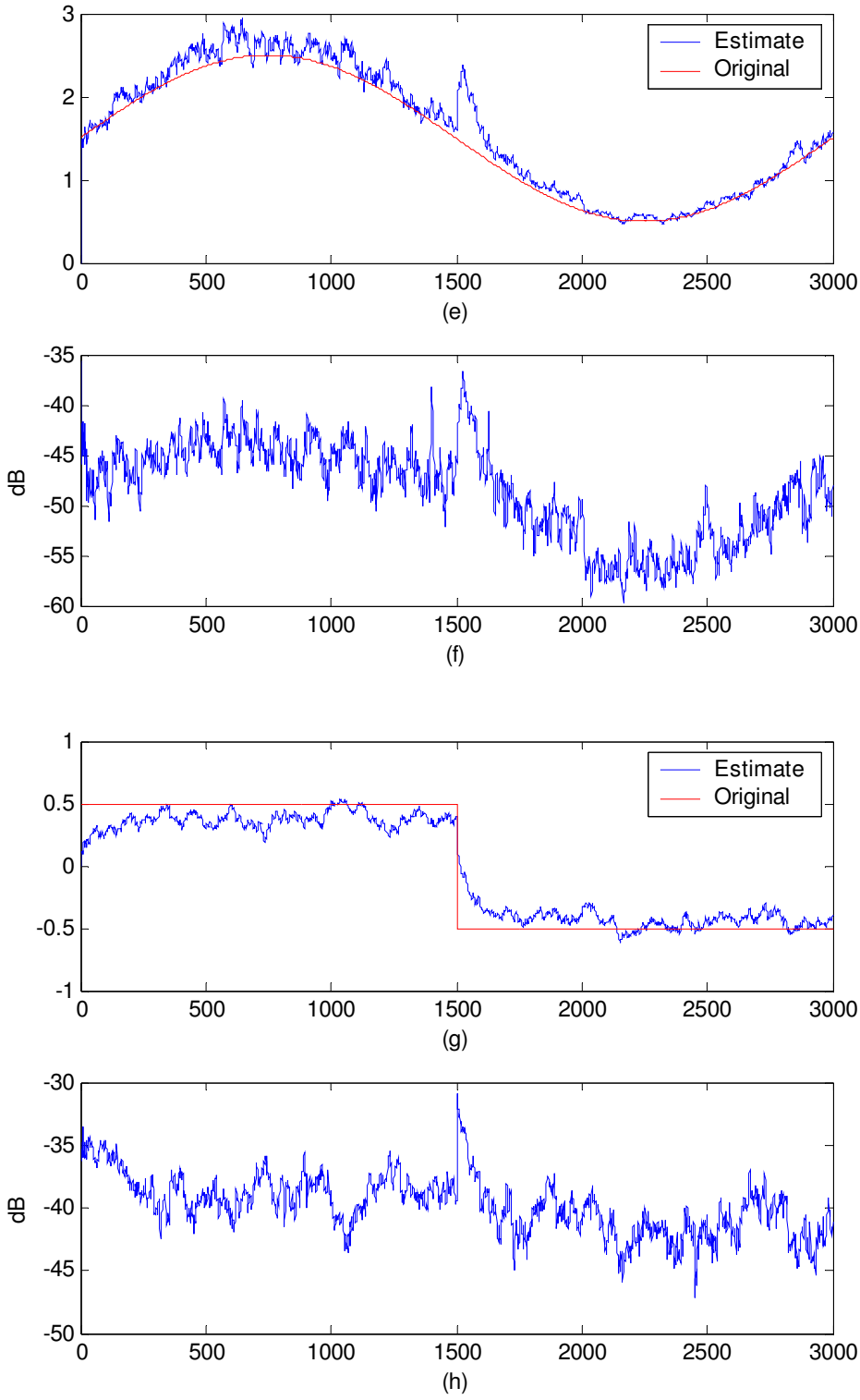


Figure 4.10. (continued): e) Estimation of the dispersion parameter, f) NMSE curve of the dispersion parameter estimation, g) Estimation of the skewness parameter, h) NMSE curve of the skewness parameter estimation

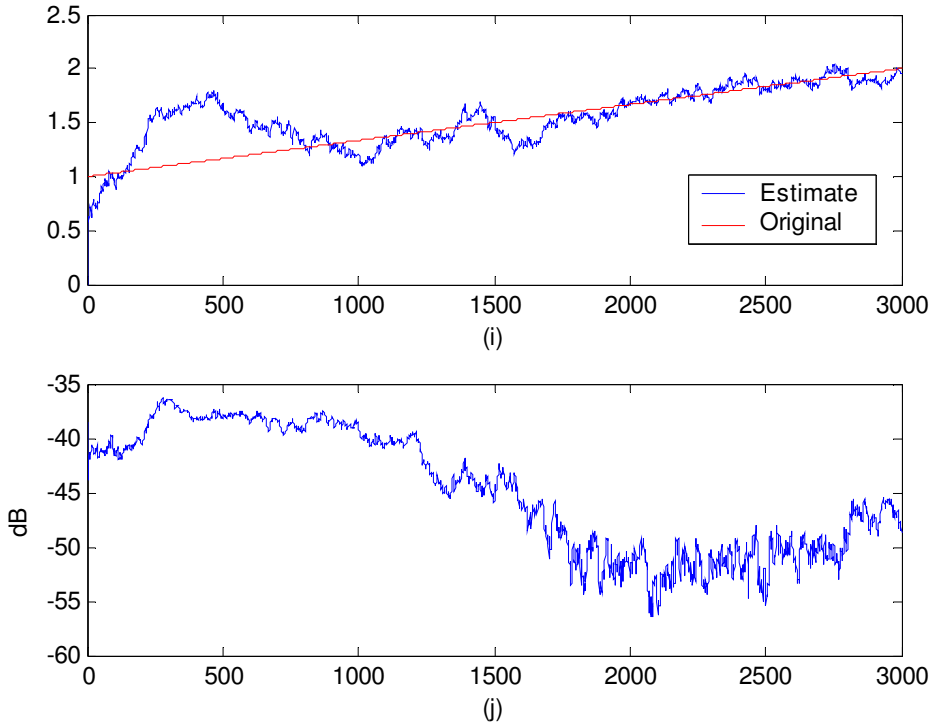


Figure 4.10. (continued): i) Estimation of the location parameter, j) NMSE curve of the location parameter estimation

- Experiment B3:

In this experiment, the TVAR coefficient, the shape, skewness and location parameters of the α-stable process are taken as in the previous case. However, the variation of the dispersion parameter is taken to be piecewise continuous as illustrated below:

$$\gamma_t = \begin{cases} 2 & t < \tau/2 \\ 1 & t \geq \tau/2 \end{cases} \tag{4.52}$$

In Figure 4.11., estimates of $\phi_t$, $\alpha_t$, $\beta_t$, $\gamma_t$ and $\mu_t$, and their corresponding NMSE curves are illustrated.

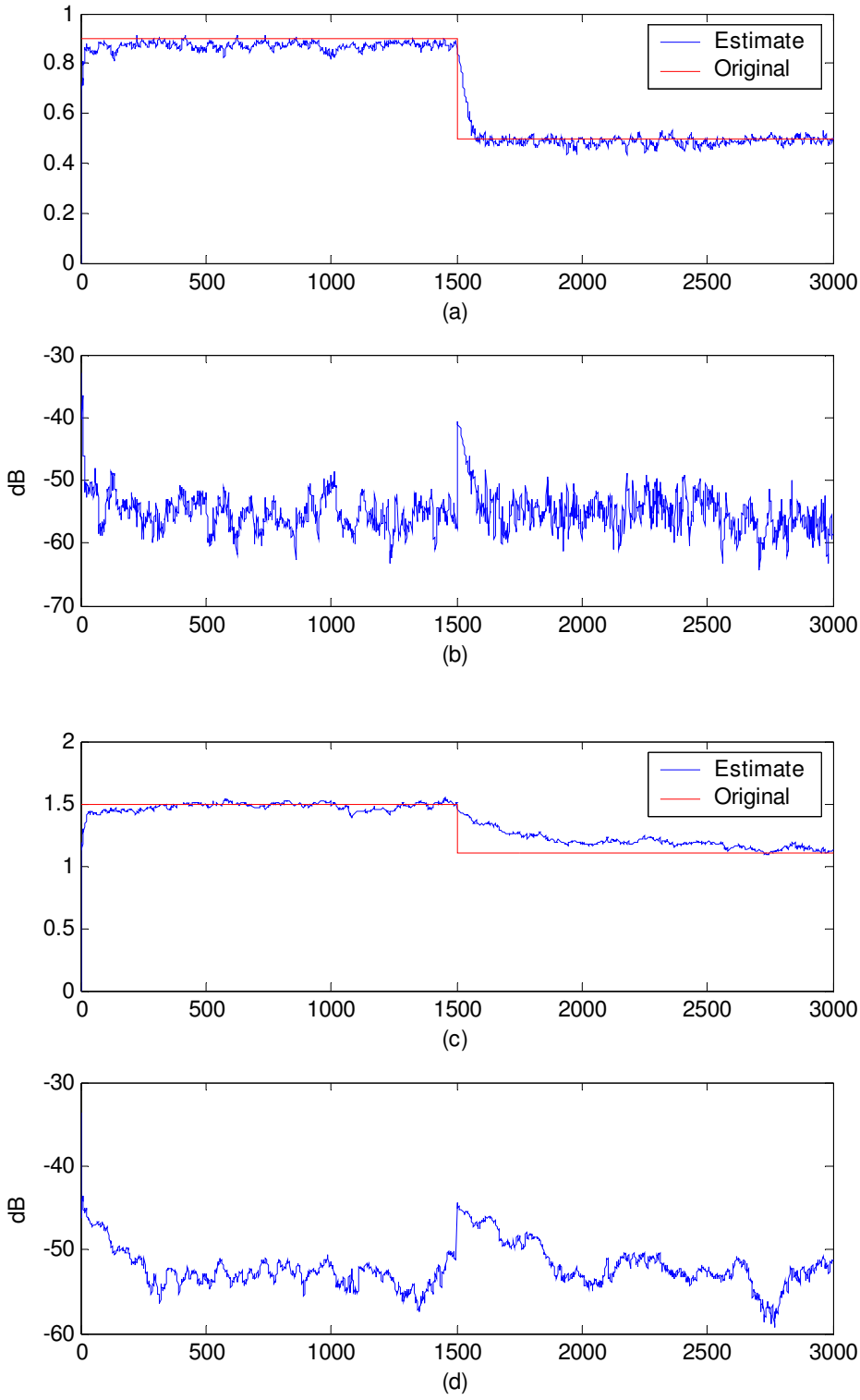


Figure 4.11. Experiment B3: Estimation of time-varying AR coefficient and distribution parameters of *Skewed α-stable* process a) Estimation of TVAR coefficient, b) NMSE curve of TVAR estimate, c) Estimation of the shape parameter, d) NMSE curve of the shape parameter estimation

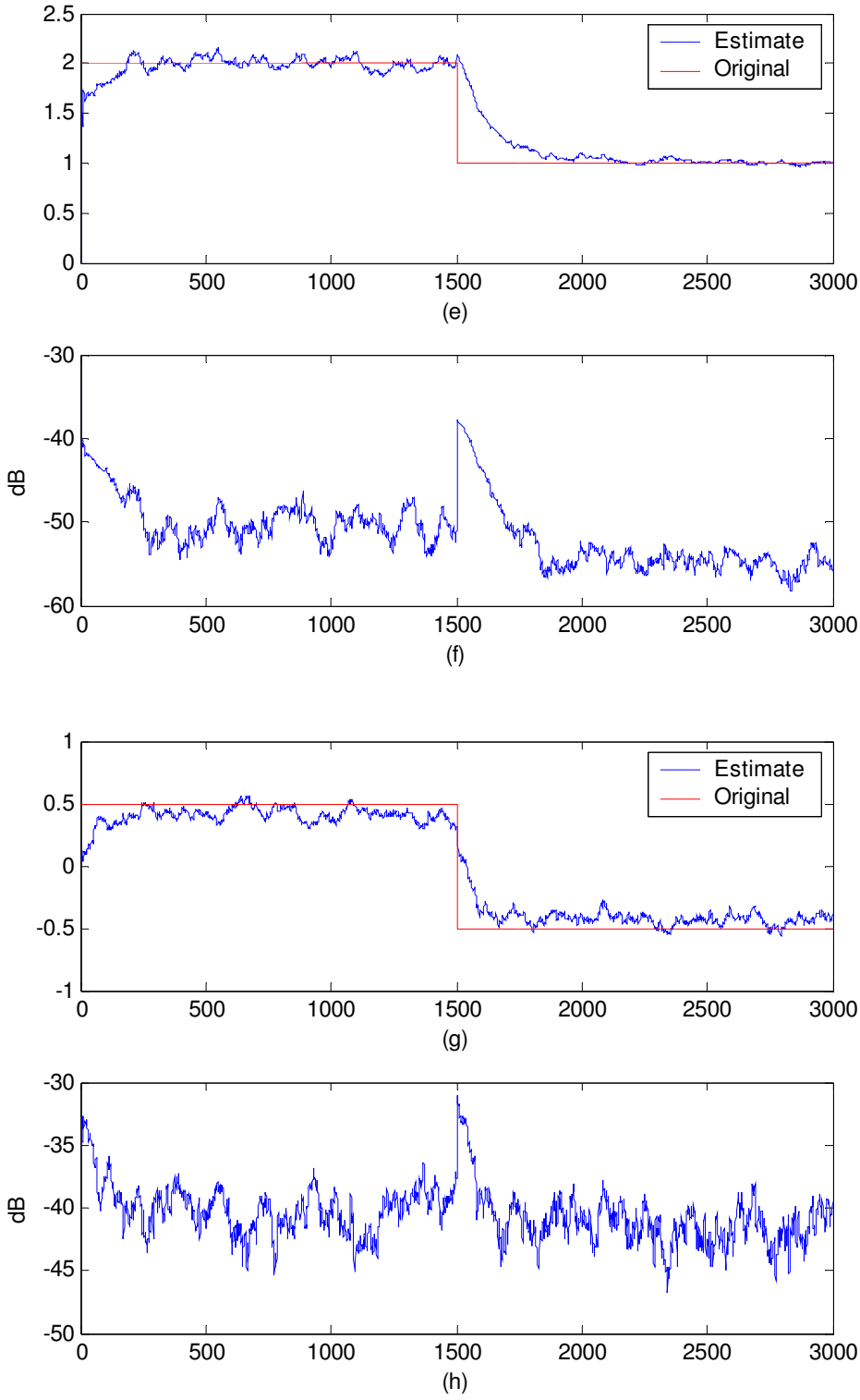


Figure 4.11. (continued): e) Estimation of the dispersion parameter, f) NMSE curve of the dispersion parameter estimation, g) Estimation of the skewness parameter, h) NMSE curve of the skewness parameter estimation

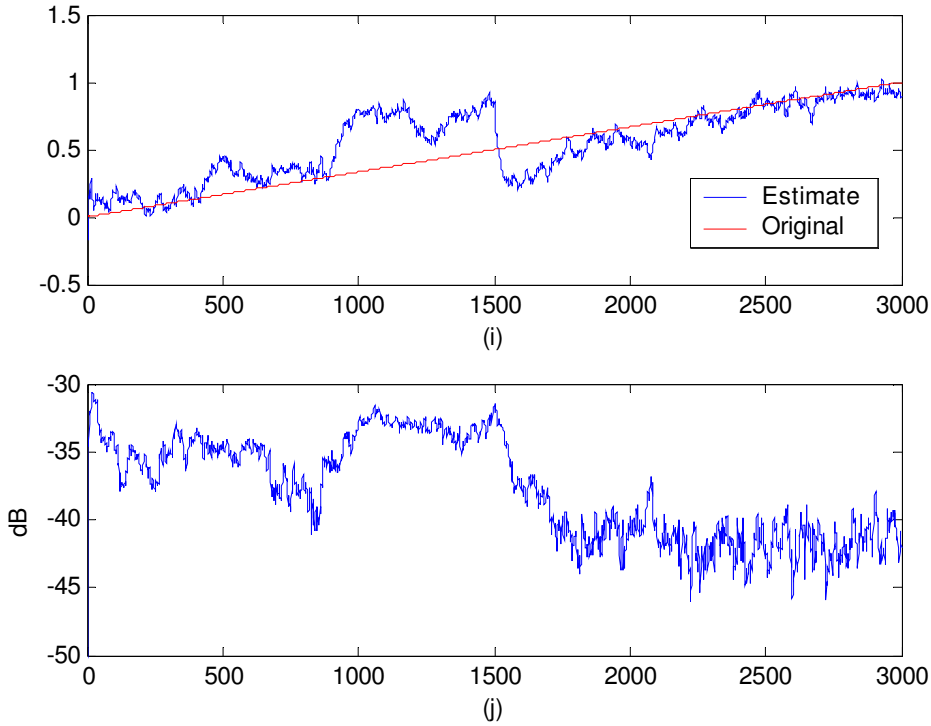


Figure 4.11. (continued): i) Estimation of the location parameter, j) NMSE curve of the location parameter estimation

c) Sequential Modelling of TVAR SαS Processes (Sinusoidal AR variation)

In this section, the proposed method is used to model TVAR SαS processes. Thus, parameter $\beta$ is taken to be zero during the experiments performed here.

- Experiment C1:

In this experiment, AR coefficient is taken to be sinusoidally varying in time, as shown below:

$$x_t = \sin\left(\frac{2\pi t}{\tau}\right) \tag{4.53}$$

while the variation of the shape and dispersion parameters are selected to be as in Experiment B2. In Figure 4.12., estimates of $\phi_t$, $\alpha_t$ and $\gamma_t$, and their corresponding NMSE curves are illustrated.

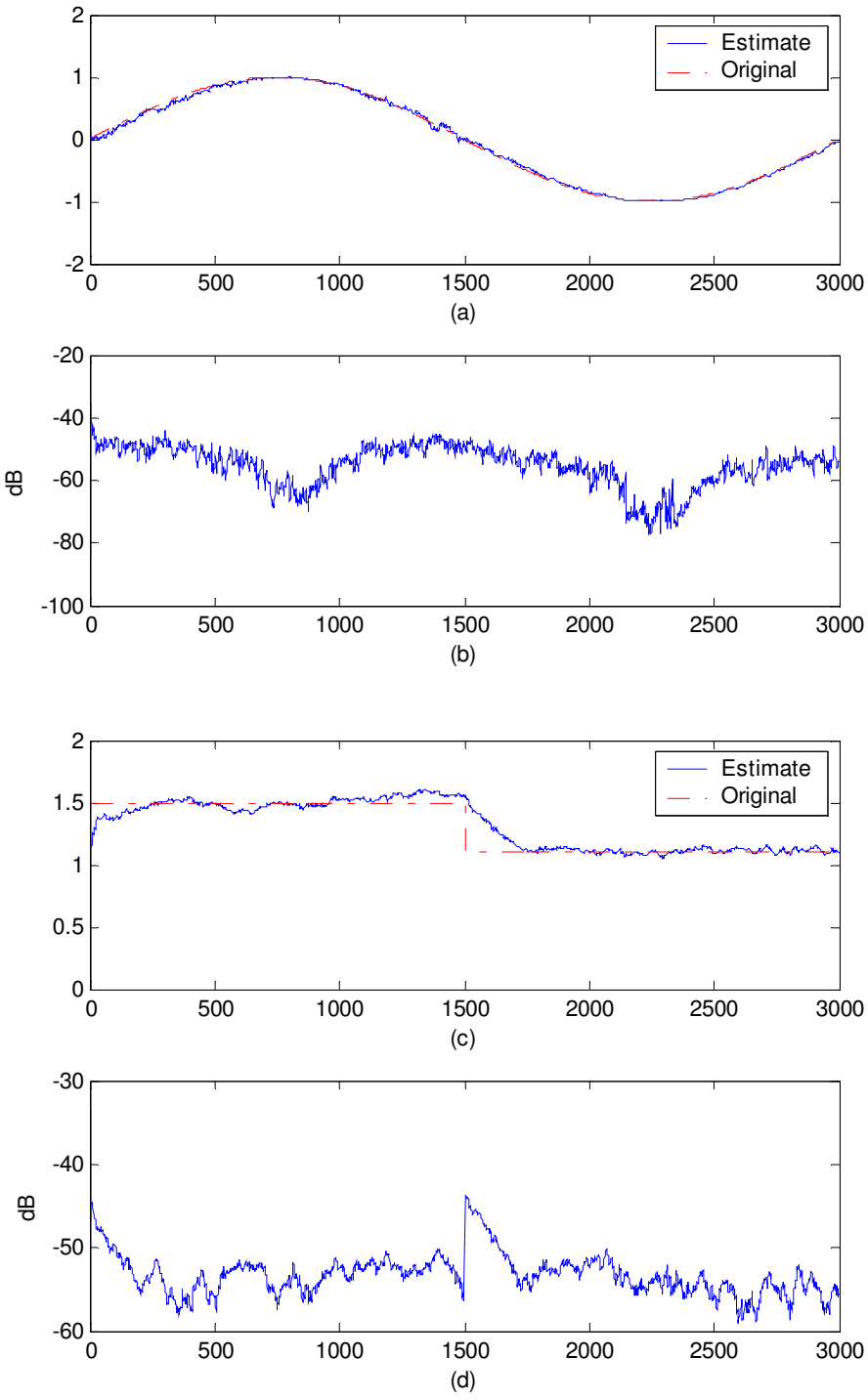


Figure 4.12. Experiment C1: Estimation of time-varying AR coefficient and distribution parameters of *SαS* process a) Estimation of TVAR coefficient, b) NMSE curve of TVAR estimate, c) Estimation of the shape parameter, d) NMSE curve of the shape parameter estimation

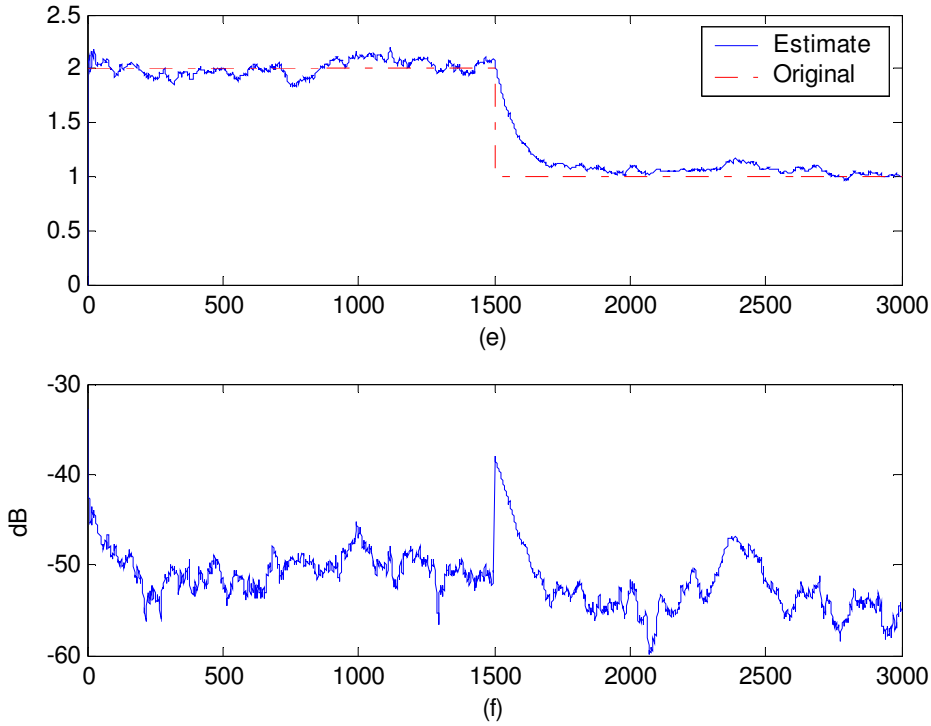


Figure 4.12. (continued): e) Estimation of the dispersion parameter, f) NMSE curve of the dispersion parameter estimation

## 4.6. Discussion

In this chapter, firstly, a new method is proposed to estimate the time-varying AR processes, which are driven by symmetric-α-stable processes with known distribution parameters. The performance of the method is tested for several values of the α parameter and it is observed to perform very well. This is also justified by the PCRLB estimates.

Then, the methods proposed in Sections 4.3. and 4.4. are compared for the case of unknown TVAR coefficients and constant distribution parameters of a *SαS* process. It is seen that the performances of the AR coefficients are similar. However, HSMC method provides distribution parameters with estimation variances which are one order of magnitude less than those obtained by the SSMC method. On the other hand, due to the online modeling capability of the SSMC technique, it is superior to HSMC when time-varying distribution parameters are to be modeled. Finally, the successful performance of the SSMC method has been demonstrated on different combinations of time-varying

parameters which can be encountered in possible future applications of non-stationary α-stable process modeling. Successful performances on almost all possible scenarios have shown that this technique provides a unifying methodology to model non-stationary α-stable processes.

Table 4.7. Experimental Scenarios

| Exp | AR | α | β | γ | μ |
|---|---|---|---|---|---|
| A1 | $\phi_t = \begin{cases} 0.9 & t < \tau/2 \\ 0.5 & t \geq \tau/2 \end{cases}$ $\sigma_{1,t}^2 = \left(\frac{1}{\xi} - 1\right) \mathrm{var}(\phi(t-1))$ | $\alpha_t = \begin{cases} 1.5 & t < \tau/2 \\ 1.1 & t \geq \tau/2 \end{cases}$ $\sigma_{\alpha,t}^2 = 5\mathrm{x}10^{-4}$ | 0 | $\gamma_t = \begin{cases} 2 & t < \tau/2 \\ 5 & t \geq \tau/2 \end{cases}$ $\sigma_{\bar{\gamma},t}^2 = 5\mathrm{x}10^{-4}$ | 0 |
| A2 | $\phi_t = \begin{cases} 0.9 & t < \tau/2 \\ 0.5 & t \geq \tau/2 \end{cases}$ $\sigma_{1,t}^2 = \left(\frac{1}{\xi} - 1\right) \mathrm{var}(\phi(t-1))$ | $\alpha_t = \begin{cases} 1.5 & t < \tau/2 \\ 1.1 & t \geq \tau/2 \end{cases}$ $\sigma_{\alpha,t}^2 = 5\mathrm{x}10^{-4}$ | 0 | $\gamma_t = 1.5 + \sin\left(\frac{2\pi t}{\tau}\right)$ $\sigma_{\bar{\gamma},t}^2 = 5\mathrm{x}10^{-3}$ | 0 |
| B1 | $\phi_t = \begin{cases} 0.9 & t < \tau/2 \\ 0.5 & t \geq \tau/2 \end{cases}$ $\sigma_{1,t}^2 = \left(\frac{1}{\xi} - 1\right) \mathrm{var}(\phi(t-1))$ | $\alpha_t = \begin{cases} 1.5 & t < \tau/2 \\ 1.1 & t \geq \tau/2 \end{cases}$ $\sigma_{\alpha,t}^2 = 5\mathrm{x}10^{-4}$ | $\beta_t = \begin{cases} 0.5 & t < \tau/2 \\ -0.5 & t \geq \tau/2 \end{cases}$ $\sigma_{\beta,t}^2 = 5\mathrm{x}10^{-3}$ | $\gamma_t = 1.5 + \sin\left(\frac{2\pi t}{\tau}\right)$ $\sigma_{\bar{\gamma},t}^2 = 5\mathrm{x}10^{-3}$ | 0 |
| B2 | $\phi_t = \begin{cases} 0.9 & t < \tau/2 \\ 0.5 & t \geq \tau/2 \end{cases}$ $\sigma_{1,t}^2 = \left(\frac{1}{\xi} - 1\right) \mathrm{var}(\phi(t-1))$ | $\alpha_t = \begin{cases} 1.5 & t < \tau/2 \\ 1.1 & t \geq \tau/2 \end{cases}$ $\sigma_{\alpha,t}^2 = 5\mathrm{x}10^{-4}$ | $\beta_t = \begin{cases} 0.5 & t < \tau/2 \\ -0.5 & t \geq \tau/2 \end{cases}$ $\sigma_{\beta,t}^2 = 5\mathrm{x}10^{-3}$ | $\gamma_t = 1.5 + \sin\left(\frac{2\pi t}{\tau}\right)$ $\sigma_{\bar{\gamma},t}^2 = 5\mathrm{x}10^{-3}$ | $\mu = \frac{1}{\tau} t$ $\sigma_{\mu,t}^2 = 5\mathrm{x}10^{-3}$ |
| B3 | $\phi_t = \begin{cases} 0.9 & t < \tau/2 \\ 0.5 & t \geq \tau/2 \end{cases}$ $\sigma_{1,t}^2 = \left(\frac{1}{\xi} - 1\right) \mathrm{var}(\phi(t-1))$ | $\alpha_t = \begin{cases} 1.5 & t < \tau/2 \\ 1.1 & t \geq \tau/2 \end{cases}$ $\sigma_{\alpha,t}^2 = 5\mathrm{x}10^{-4}$ | $\beta_t = \begin{cases} 0.5 & t < \tau/2 \\ -0.5 & t \geq \tau/2 \end{cases}$ $\sigma_{\beta,t}^2 = 5\mathrm{x}10^{-3}$ | $\gamma_t = \begin{cases} 2 & t < \tau/2 \\ 1 & t \geq \tau/2 \end{cases}$ $\sigma_{\bar{\gamma},t}^2 = 5\mathrm{x}10^{-4}$ | $\mu = \frac{1}{\tau} t$ $\sigma_{\mu,t}^2 = 5\mathrm{x}10^{-3}$ |
| C1 | $\phi_t = \sin\left(\frac{2\pi t}{\tau}\right)$ $\sigma_{1,t}^2 = \left(\frac{1}{\xi} - 1\right) \mathrm{var}(\phi(t-1))$ | $\alpha_t = \begin{cases} 1.5 & t < \tau/2 \\ 1.1 & t \geq \tau/2 \end{cases}$ $\sigma_{\alpha,t}^2 = 5\mathrm{x}10^{-4}$ | 0 | $\gamma_t = \begin{cases} 2 & t < \tau/2 \\ 1 & t \geq \tau/2 \end{cases}$ $\sigma_{\bar{\gamma},t}^2 = 5\mathrm{x}10^{-4}$ | 0 |

# 5. BAYESIAN MODELING OF CROSS-CORRELATED NON-STATIONARY NON-GAUSSIAN PROCESSES

## 5.1. Introduction

The previous chapter brings solution to the estimation of linear, non-stationary non-Gaussian processes. In order to model time-structured processes, linear autoregressive modeling has been performed due to its wide application areas in the literature (Lütkepohl, 1993; Hamilton, 1994). This section generalizes this idea for multivariate cases in order to model relationships between different AR processes. In the literature, these relationships between different cross-correlated processes has been modeled by VAR structures in many areas such as chemical processes (Hsu, 1997), functional magnetic resonance imaging (Sato *et al*., 2006), multichannel EEG data analysis (Möller *et al*., 2001) and mobile communication channels (Jachan and Matz, 2005). Here, we extend our previous particle filtering based methodology so that these relationships can also be modeled. These AR processes are expressed in terms of a VAR model as explained in Chapter 2. In the modeling of *stationary* VAR processes, cross-correlations between AR processes are incorporated by the non-diagonal elements of an AR matrix and/or by the cross-correlated driving noise processes (Lütkepohl, 1993; Hamilton, 1994). This stationary modeling has been demonstrated by (2.8) and (2.9) for a bivariate case. Then, it is generalized so that any stationary VAR process of order $K$ can be modeled by (2.10) and (2.11). In literature, these processes are estimated successfully by ML methods (Lütkepohl, 1993; Hamilton, 1994) and LS approaches (Lütkepohl, 1993; Hsu, 1997; Neumaier and Schneider, 2001). A major shortcoming of these approaches is their inability to model time-varying VAR processes where the elements of the AR matrix and/or driving process statistics do change in time. Thus, such non-stationarities cannot be handled by the above methods. Such a model is shown below for the bivariate case:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix} \tag{5.1}$$

where the AR coefficient matrix and driving noise statistics is given as follows:

$$\mathbf{\Phi}_{1,t} = \left(\boldsymbol{\varphi}_{1,t}^T, \boldsymbol{\varphi}_{2,t}^T\right)^T = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{bmatrix} \quad \text{and} \quad \mathbf{n}_t = \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix}, \quad \mathbf{n}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_{n_t}\right) \tag{5.2}$$

It should be noted that both $\mathbf{\Phi}_{1,t}$ and $\mathbf{\Sigma}_{n_t}$ change over time, unlike the model in (2.8) and (2.9). This bivariate model can be generalized for a time-varying VAR($K$) by the following equation:

$$\mathbf{y}_t = \mathbf{\Phi}_{1,t}\mathbf{y}_{t-1} + \mathbf{\Phi}_{2,t}\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_{K,t}\mathbf{y}_{t-K} + \mathbf{n}_t \tag{5.3}$$

where $\mathbf{y}_t = \left[y_{1,t}, y_{2,t}, \ldots, y_{d,t}\right]^T$, $\mathbf{n}_t = \left[n_{1,t}, n_{2,t}, \ldots, n_{d,t}\right]^T$ and $\mathbf{\Phi}_{\mathrm{j,t}}$ denote the VAR($K$) process, driving noise and ($d$ x $d$) matrix of AR coefficients for $j = 1, 2, \ldots, K$, respectively. Similar to (5.2), noise is usually modeled by the following Gaussian form in the literature:

$$\mathbf{n}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_{n_t}\right) \tag{5.4}$$

It is observed in (5.3) and (5.4) that the AR coefficient matrices and the noise covariance are generally modeled to be time dependent in this structure.

Although different approaches have been proposed to model time-varying VAR($K$) processes (Lütkepohl, 1993; Möller *et al*., 2001; Jachan and Matz, 2005; Sato *et al*., 2006), in all of these references driving process was modeled by a *Gaussian* distribution. To the best of our knowledge, modeling of non-stationary, cross-correlated VAR($K$) processes with *non-Gaussian* driving noise innovations is an open research area. This chapter brings a solution to this problem with the proposition of a novel method that can estimate time-varying VAR processes by the utilization of particle filtering. Ability to model these non-Gaussian cases presents a unifying property of our particle filtering methodology. Moreover, this section will be used as a bridge to model non-stationary mixtures of cross-correlated AR processes in the following chapter.

## 5.2. Modeling of Non-Stationary Cross Correlated TVAR Processes

In general, a cross correlated time-varying VAR processes of any order, *K*, can be modeled by vector autoregressions as shown in (5.3). In the literature, the driving process is generally demonstrated by (5.4). However, in this chapter, we model the driving process by a mixture of Gaussians, which is a special case of sub-Gaussian α-stable processes (Kuruoğlu *et al*., 1998). For the sake of completeness, the time-varying VAR(*K*) model of (5.3) is rewritten below followed by the non-Gaussian model of the driving process:

$$\mathbf{y}_t = \mathbf{\Phi}_{1,t}\mathbf{y}_{t-1} + \mathbf{\Phi}_{2,t}\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_{K,t}\mathbf{y}_{t-K} + \mathbf{n}_t \tag{5.5}$$

where $\mathbf{y}_t = \left[ y_{1,t}, y_{2,t}, \ldots, y_{d,t} \right]^T$, $\mathbf{n}_t = \left[ n_{1,t}, n_{2,t}, \ldots, n_{d,t} \right]^T$ and $\mathbf{\Phi}_{\mathrm{j,t}}$ denote the VAR(*K*) process and (*d* x *d*) matrix of AR coefficients for $j = 1, 2, \ldots, K$, respectively. Subscript *t* indicates that the AR coefficients are time-varying. Here, the driving process is represented by $\mathbf{n}_{\mathrm{t}}$, which is modeled to be a mixture of Gaussians as shown below:

$$\mathbf{n}_t \sim \sum_{l=1}^{N_n} p_l \mathcal{N}\left(\mathbf{m}_l, \mathbf{\Sigma}_l\right) \tag{5.6}$$

where $\mathbf{n}_t$ is represented by a mixture of $N_n$ Gaussians with $p_l, \mathbf{m}_l, \mathbf{\Sigma}_l$ denoting the probability, mean and covariance matrices of each Gaussian component, respectively. Here, we assume that these statistical properties are known and time-invariant. Non-stationarity of the time-varying VAR(*K*) process only arises due to the time-varying nature of its AR coefficient matrices, $\mathbf{\Phi}_{\mathrm{j,t}}$.

In this section, in order to model non-stationary VAR processes given by (5.5), the use of the Bootstrap particle filtering scheme (Table 3.3) is proposed with the following definition of the state vector:

$$\mathbf{x}_t = \left[ x_{1,t}, x_{2,t}, \ldots, x_{K',t} \right]^T = \left[ \tilde{\mathbf{\Phi}}_{1,t}^T, \tilde{\mathbf{\Phi}}_{2,t}^T, \ldots, \tilde{\mathbf{\Phi}}_{K,t}^T \right]^T \tag{5.7}$$

where $\tilde{\mathbf{\Phi}}_{j,t} = vec\left(\mathbf{\Phi}_{j,t}\right)$ denoting the column vector of size ($d^2$ x *1*) which is composed of concatenating the columns of the $j^{th}$ AR matrix. Thus, $\mathbf{x}_t$ has a dimension of ( $K'$ x *1*), where $K' = Kd^2$. Here, there is no *a priori* information regarding the time-variations of the state variables as in Chapter 4. We only know the functional form of the measurement equation given by 5.5. Therefore, in order to use a particle filter, we have to model an artificial state-transition equation as in Section 4.3. Thus, process equation of the state vector of (5.7) is modeled by an artificial random walk which is driven by a zero-mean Gaussian noise process having a time-varying covariance matrix. So, state-space representation for this problem can be given by the following equations:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{V}_t$$

$$\mathbf{x}_t = \begin{bmatrix} \tilde{\mathbf{\Phi}}_{1,t} \\ \tilde{\mathbf{\Phi}}_{2,t} \\ \vdots \\ \tilde{\mathbf{\Phi}}_{K,t} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{\Phi}}_{1,t-1} \\ \tilde{\mathbf{\Phi}}_{2,t-1} \\ \vdots \\ \tilde{\mathbf{\Phi}}_{K,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \\ \vdots \\ v_{K',t} \end{bmatrix} \tag{5.8.a}$$

$$\mathbf{y}_t = \mathbf{\Phi}_{1,t}\mathbf{y}_{t-1} + \mathbf{\Phi}_{2,t}\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_{K,t}\mathbf{y}_{t-K} + \mathbf{n}_t \tag{5.8.b}$$

where $\mathbf{x}_t$ is elaborated in (5.7) and time-varying covariance matrix of the process noise, $\mathbf{V}_t$, is given by the following equation:

$$\mathbf{\Sigma}_{\mathbf{V}_t} = \mathbf{\Sigma}_{\mathbf{x}_{t-1}}\left(\frac{1}{\xi} - 1\right) \tag{5.9}$$

where

$$\mathbf{\Sigma}_{\mathbf{V}_t} = diag\left(\sigma_{1,t}^2, ..., \sigma_{K',t}^2\right) \tag{5.10}$$

In (5.10), each variance is calculated as follows:

$$\sigma_{k,t}^2 = \left(\frac{1}{\xi} - 1\right)\mathrm{var}\left(x_{t-1}(k)\right), \qquad k = 1, 2, ..., K' \tag{5.11}$$

Here, particles are drawn from *a priori* state-transition pdf which is determined by the artificial model of (5.8.a), i.e. $q\left(\mathbf{x}_t \left| \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t} \right.\right) = p\left(\mathbf{x}_t \left| \mathbf{x}_{t-1} \right.\right)$ is used implying that $\mathbf{x}_t^{(i)} \sim \mathcal{N}\left(\mathbf{x}_t^{(i)}, \mathbf{\Sigma}_{V_t}\right)\mathbb{I}_{\overline{\mathbf{x}}}$ where the indicator function is defined as $\mathbb{I}_{\mathbf{x}} = \begin{cases} 1, & \mathbf{x}_t^{(i)} \in \overline{\mathbf{x}} \\ 0, & \mathbf{x}_t^{(i)} \notin \overline{\mathbf{x}} \end{cases}$. The indicator function is used to ensure that the drawn particles take place within the stability condition $\overline{\mathbf{x}}$ of the VAR(*K*) process. After drawing each particle, their importance weights are calculated by the likelihood function, since Bootstrap particle filtering is used (Table 3.3):

$$\begin{aligned} w_t^{(i)} &= p\left(\mathbf{y}_t \left| \mathbf{x}_t^{(i)} \right.\right) \\ p\left(\mathbf{y}_t \left| \mathbf{x}_t^{(i)} \right.\right) &= \mathcal{N}\left(\mathbf{y}_t; \mathbf{m}_t^{(i)}, \mathbf{\Sigma}_{\mathbf{n}_t}\right) \end{aligned} \tag{5.12}$$

where the mean of the likelihood function can be calculated from the drawn particles as follows:

$$\mathbf{m}_t^{(i)} = \mathbf{\Phi}_{1,t}^{(i)}\mathbf{y}_{t-1} + \mathbf{\Phi}_{2,t}^{(i)}\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_{K,t}^{(i)}\mathbf{y}_{t-K} \tag{5.13}$$

After the calculation of the importance weights, resampling is performed. A pseudo-code of the algorithm is given in Table 5.1.

Table 5.1. Pseudo-code of the proposed method to model cross-correlated non-stationary non-Gaussian processes

For *i =1 to N*,

1. INITIATION:

   Draw samples from the initial distributions of the state variables: $\mathbf{x}_t^{(i)} \sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$,

   where **m** and **P** denote the mean and covariance matrices of the Gaussian distribution. Note that, $\mathbf{P}_0$ is a diagonal matrix.

For *t = 1 to* $\tau$ , ($\tau$ denotes the data length)

Table 5.1. (continued)

2. STATE TRANSITIONS:

Calculate the variance of each AR coefficient:

Calculate time-varying variances of the state-transition density and form $\mathbf{\Sigma}_{\mathbf{V}_t}$ matrix:

$$\sigma_{k,t}^2 = \left(\frac{1}{\xi} - 1\right) \mathrm{var}\left(x_{t-1}(k)\right), \qquad k = 1, 2, ..., K'$$

Draw new particles for each state variable by using the proposed state-transition equation:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{V}_t$$

$$\mathbf{x}_t = \begin{bmatrix} \tilde{\mathbf{\Phi}}_{1,t} \\ \tilde{\mathbf{\Phi}}_{2,t} \\ \vdots \\ \tilde{\mathbf{\Phi}}_{K,t} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{\Phi}}_{1,t-1} \\ \tilde{\mathbf{\Phi}}_{2,t-1} \\ \vdots \\ \tilde{\mathbf{\Phi}}_{K,t-1} \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \\ \vdots \\ v_{K',t} \end{bmatrix}$$

$$\mathbf{x}_t^{(i)} \sim \mathcal{N}\left(\mathbf{x}_t^{(i)}, \mathbf{\Sigma}_{V_t}\right) \mathbb{I}_{\bar{\mathbf{x}}}$$

3. CALCULATE THE IMPORTANCE WEIGHT OF EACH PARTICLE:

$$w_t^{(i)} = p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right)$$

$$p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) = \mathcal{N}\left(\mathbf{y}_t; \mathbf{m}_t^{(i)}, \mathbf{\Sigma}_{\mathbf{n}_t}\right)$$

where

$$\mathbf{m}_t^{(i)} = \mathbf{\Phi}_{1,t}^{(i)} \mathbf{y}_{t-1} + \mathbf{\Phi}_{2,t}^{(i)} \mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_{K,t}^{(i)} \mathbf{y}_{t-K}$$

4. NORMALIZE THE WEIGHTS:

$$\tilde{w}_t^i = \frac{w_t^i}{\sum_{i=1}^{N} w_t^i}$$

5. RESAMPLE AND GO TO STEP 2.

## 5.3. Experiments

In this section, performance of the proposed method is illustrated by computer simulations where the estimation of a non-stationary cross correlated VAR(1) process is performed. First, performance is illustrated by taking a Gaussian driving process. Then, its more general case is presented where mixtures of Gaussians drive the VAR process.

**Experiment 1. Gaussian distributed driving process**

Here, time-varying AR coefficients of a bivariate VAR(1) process are estimated. Total data length is taken to be $\tau = 1000$. This process is given by the following equation:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix} \tag{5.14}$$

where the cross correlation between the two AR processes is generated by the following TVAR coefficients:

$$\begin{aligned} \phi_{11,t} &= \begin{cases} 0.5 & t < 500 \\ -0.5 & t \geq 500 \end{cases}, \quad \phi_{12,t} = \begin{cases} 0.1 & t < 500 \\ -0.1 & t \geq 500 \end{cases} \\ \phi_{21,t} &= \begin{cases} 0.4 & t < 500 \\ -0.4 & t \geq 500 \end{cases}, \quad \phi_{22,t} = \begin{cases} 0.5 & t < 500 \\ -0.5 & t \geq 500 \end{cases} \end{aligned} \tag{5.15}$$

and the driving process is distributed as follows:

$$\mathbf{n}_t \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 0.1 & 0 \\ 0 & 0.3 \end{bmatrix}\right) \tag{5.16}$$

Here, driving process statistics is assumed to be known and the particle filtering method given in Table 5.1. is used with $N = 100$ particles, $\xi = 0.9$ and the initial state distribution of $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ with dimension (*4 x 1*). 100 realizations are used for ensemble averaging and the following estimates of the TVAR coefficients are obtained:

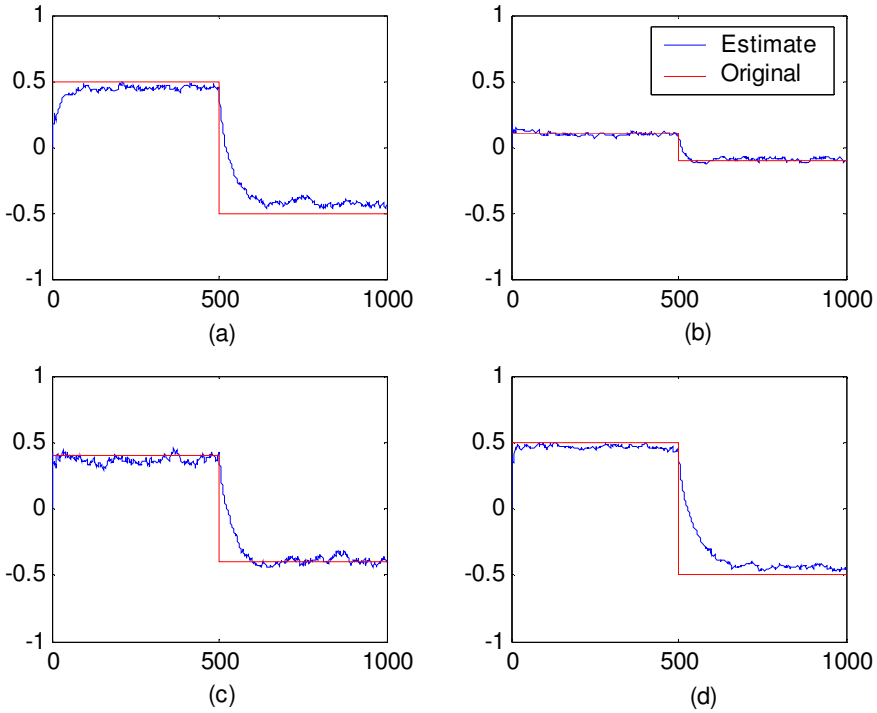


Figure 5.1. Estimates of TVAR coefficients of a bivariate VAR(1) process with Gaussian driving process *a)* $\phi_{11}(t)$, *b)* $\phi_{12}(t)$, *c)* $\phi_{21}(t)$, *d)* $\phi_{22}(t)$

Processes are illustrated below:

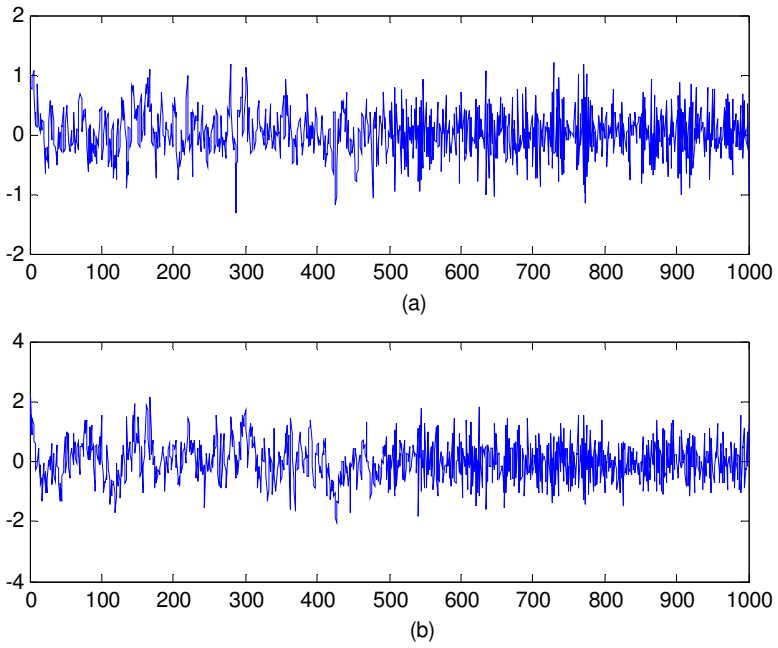


Figure 5.2. TVAR processes for Gaussian driving processes *a)* $y_{1,t}$, *b)* $y_{2,t}$

**Experiment 2. Mixture of Gaussian distributed driving process**

Here, time-varying AR coefficients of a bivariate VAR(1) process are estimated. Total data length is taken to be $\tau = 1000$. The observed processes and the time-variation of the AR coefficients are taken as the same in (5.14) and (5.15), respectively. Only, the distribution of the driving process is chosen to be as follows:

$$\mathbf{n}_t \sim 0.5\mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) + 0.5\mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}\right) \tag{5.17}$$

Here, driving process statistics is assumed to be known and the particle filtering method given in Table 5.1. is used with $N = 100$ particles, $\xi = 0.9$ and the initial state distribution of $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ with dimension (*4 x 1*). 100 realizations are used for ensemble averaging and the following estimates of the TVAR coefficients are obtained:

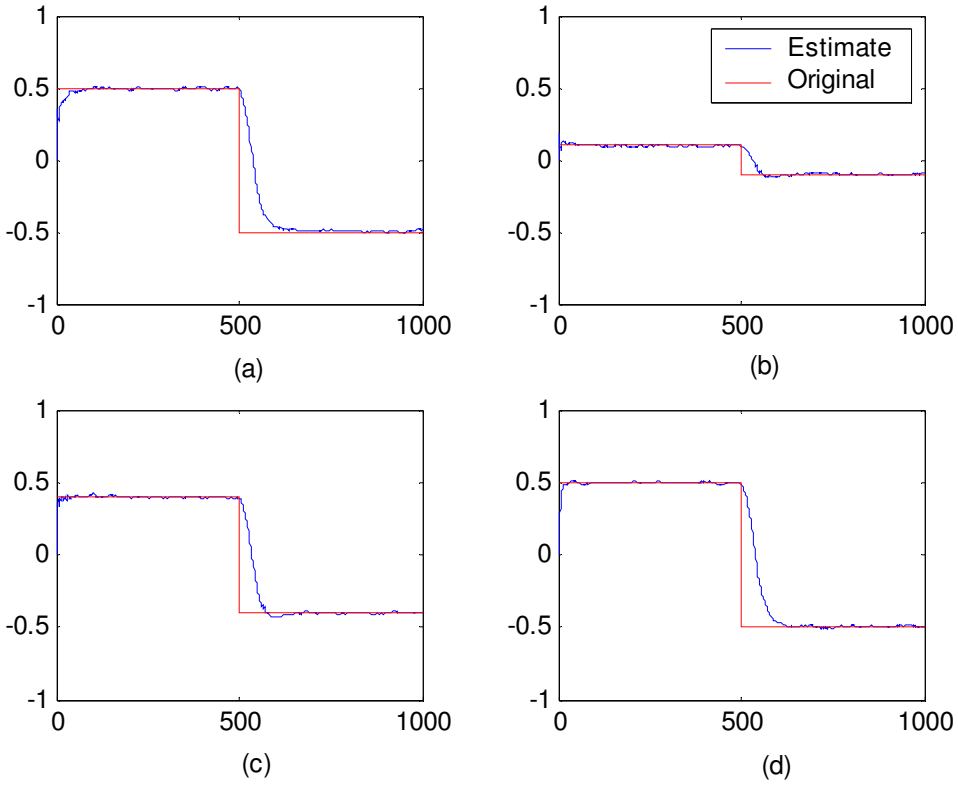


Figure 5.3. Estimates of TVAR coefficients of a bivariate VAR(1) process with mixture of Gaussians driving process *a)* $\phi_{11}(t)$, *b)* $\phi_{12}(t)$, *c)* $\phi_{21}(t)$, *d)* $\phi_{22}(t)$

The waveforms of the cross-correlated non-Gaussian processes are illustrated in the following figure:

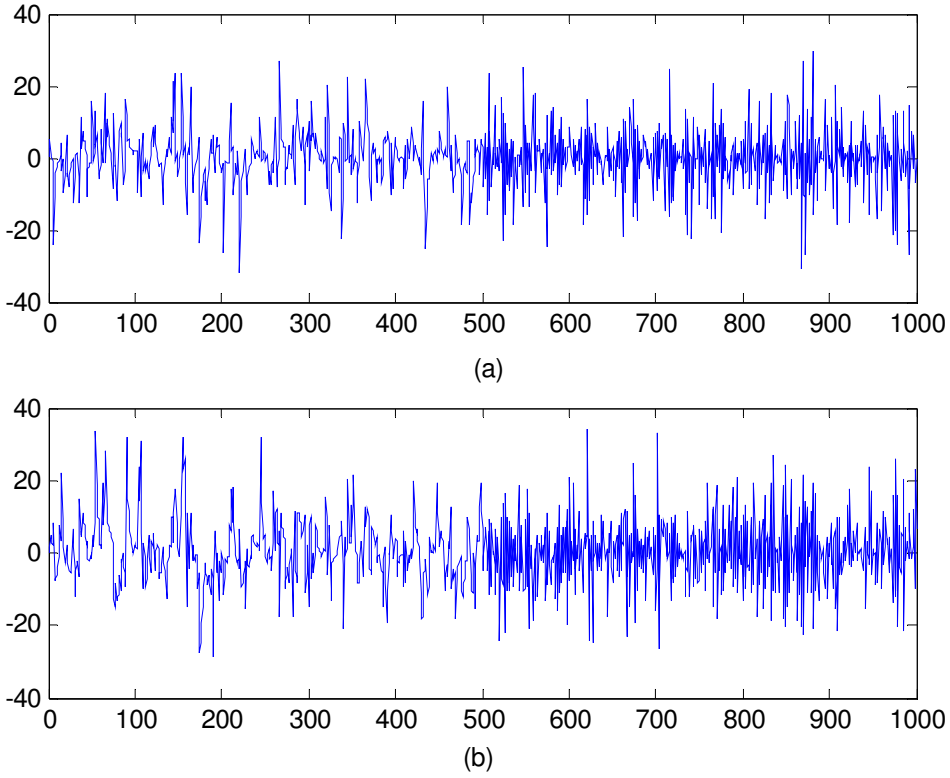


Figure 5.4. TVAR processes for non-Gaussian driving processes *a*) $y_{1,t}$, *b*) $y_{2,t}$

## 5.4. Discussion

In this chapter, we extended our particle filtering methodology so that the cross-correlated non-stationary non-Gaussian processes can be modeled. For this purpose, we made use of the VAR structured cross-correlated processes. As a result, our modeling scheme has been generalized for cross-correlated multivariate cases which are highly encountered in many application areas which are introduced in Chapter 1. To the best of our knowledge, although different techniques have been developed in the literature to model time-varying VAR processes, our method is the only one that can be used in case of non-Gaussian driving processes. The performance of the algorithm has been demonstrated both for Gaussian and mixture of Gaussian distributed driving processes. It is observed from Figures 5.1. and 5.3. that the technique is successful in both cases. Moreover, it can be seen that the quality of the estimated AR coefficients increases as the process becomes

more heavy tailed. This result is in accordance with our observations in the previous chapter.

# 6. BAYESIAN MODELING OF NON-STATIONARY MIXTURES OF CROSS-CORRELATED PROCESSES

## 6.1. Introduction

Having solved the problem of estimating non-stationary cross correlated AR processes successfully, a more unifying model is presented here, where only mixtures of these signals are observed. This brings an additional complexity to the problem, since the AR signals cannot be observed directly. They are unobservable latent processes in such cases. This problem can be interpreted as a Dependent Component Analysis (DCA) model where the underlying AR processes are treated as the sources to be separated. In the literature, separation of independent TVAR sources from their mixtures was studied by (Andrieu and Godsill, 2000) to separate audio signals and by (Dally and Reilly, 2005) to recover signals mixed by convolutive channels, such as those encountered in wireless communications and reverberating enclosures. In these works, model based separation algorithms have been developed by using particle filters. Despite their successful separation performances, the *independence assumption* of the sources makes the use of these algorithms impossible in the DCA problem. This chapter brings a novel solution to the estimation of the mixtures of cross-correlated processes by extending our particle filtering methodology, i.e. non-stationary DCA problem is solved here. Here, different AR sources are treated as a VAR process with cross correlations between them. This problem is formulated below for a bivariate situation:

$$\begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} s_{1,t-1} \\ s_{2,t-1} \end{bmatrix} + \begin{bmatrix} \upsilon_{1,t} \\ \upsilon_{2,t} \end{bmatrix} \tag{6.1}$$

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix} \begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix} \tag{6.2}$$

where (6.1) corresponds to (5.1) in case of time-invariant AR coefficients. Here, an instantaneous mixture of these cross-correlated processes, $s_{1,t}$ and $s_{2,t}$, is observed as

$\mathbf{y}_t = \left[ y_{1,t}, y_{2,t} \right]^T$. Thus, this is a more challenging problem than that of Chapter 5, since the cross-correlated sources ($s_{1,t}$ and $s_{2,t}$) cannot be observed directly. It should be noted that the AR coefficient matrix, $\mathbf{\Phi}_1 = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$, in (6.1) is taken to be time-invariant, while the mixing matrix, $\mathbf{A}_t = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix}$ is modeled to be time-varying. According to our observations, we noted that only one of these two matrices can be time-varying so that the modeling could be performed successfully. Otherwise, an ambiguity arises between these two matrices. This phenomenon is due to the unknown time-variation of these matrices. Here, it is assumed that time-variation model of the mixing matrix elements and the AR coefficients are not known *a priori*, unlike the approaches of (Andrieu and Godsill, 2000) and (Dally and Reilly, 2005). Thus, the proposed method brings a successful solution to a more challenging problem.

In (6.2), modeling of a mixture of cross-correlated AR processes is illustrated. However, in order to approach the problem from a simpler point of view, modeling of cross-correlated but temporally independent processes is examined, first. This modeling problem can be shown as follows for a bivariate case:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix} \begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix} \tag{6.3}$$

where $s_{1,t}$ and $s_{2,t}$ do not regress on their past values. However, they possess spatial cross-correlation in the following form:

$$\mathbf{s}_t = \begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} \sim \mathcal{N}(\mathbf{m}_s, \mathbf{\Sigma}_\mathbf{s}), \quad E\left[ (\mathbf{s}_t - \mathbf{m}_s)(\mathbf{s}_t - \mathbf{m}_s)^T \right] = \mathbf{\Sigma}_\mathbf{s} \tag{6.4}$$

where $\mathbf{\Sigma}_s$ is a constant and positive definite non-diagonal matrix. In order to model such mixtures, we will assume that the statistical properties of the sources are known *a priori*. In this case, we will name our particle filtering technique as the "Mixture Modeling of Cross-

Correlated Sources (MCS)". On the other hand, we will name our technique as "Mixture Modeling of Cross-Correlated AR Sources (MCARS)", where mixtures of cross-correlated AR processes are modeled as shown in (6.1) and (6.2). Here, we remove the assumption that the source statistics are known.

Unlike the cases in the preceding chapters, a mixture of different cross-correlated processes are observed in the measurement equation here and not the process samples. Therefore, it is extremely tedious to form an artificial state-transition model for such a case, i.e. it is very difficult to propose an importance function that can approximate the optimal one.

In MCS, we overcome this difficulty assuming that we have strong beliefs regarding the statistics of the sources. In this method, we propose a *hybrid algorithm* where the mixing matrix and the sources are estimated by a particle filter and a MCMC algorithm, respectively (Gençağa *et al*., 2005b). By using strong beliefs about the source distributions, we only need to model the state-transition equation pertaining to the elements of the mixing matrix. Therefore, we reduce the dimension of the states by eliminating the sources from the state vector. Moreover, the importance weights of the remaining states (mixing matrix elements) can be calculated by using a Monte Carlo integration, where samples of sources need to be drawn. This can be accomplished easily, since we assume that the source distributions are almost known *a priori* which allows us to draw samples. Then, resampling is applied and the MMSE estimates of the mixing matrix elements are successfully found. However, waveforms of the sources cannot be obtained by this particle filter, since they have been integrated out. Thus, the mixing matrix estimate, found by the particle filter, is used as the a priori value of the mixing matrix which is necessary to apply the MCMC scheme for extracting the source waveforms. However, in order to apply the batch MCMC algorithm, non-stationary data (total length of $\tau$) is divided into sub-blocks assumed to be stationary. In conclusion, in each of these small blocks, particle filtering is used to estimate the slowly varying mixing matrix and this estimate is used as the prior required to extract the sources by the MCMC method. The hierarchical Dynamic Bayesian Network (DBN) of this model is illustrated below. Here, DBN is shown in black, whereas the procedures applied for modeling are given in red colors. Here, observation data of

length $\tau$ is divided into $M$ sub-blocks. In each sub-block, it is assumed that the mixing matrix changes very slowly.

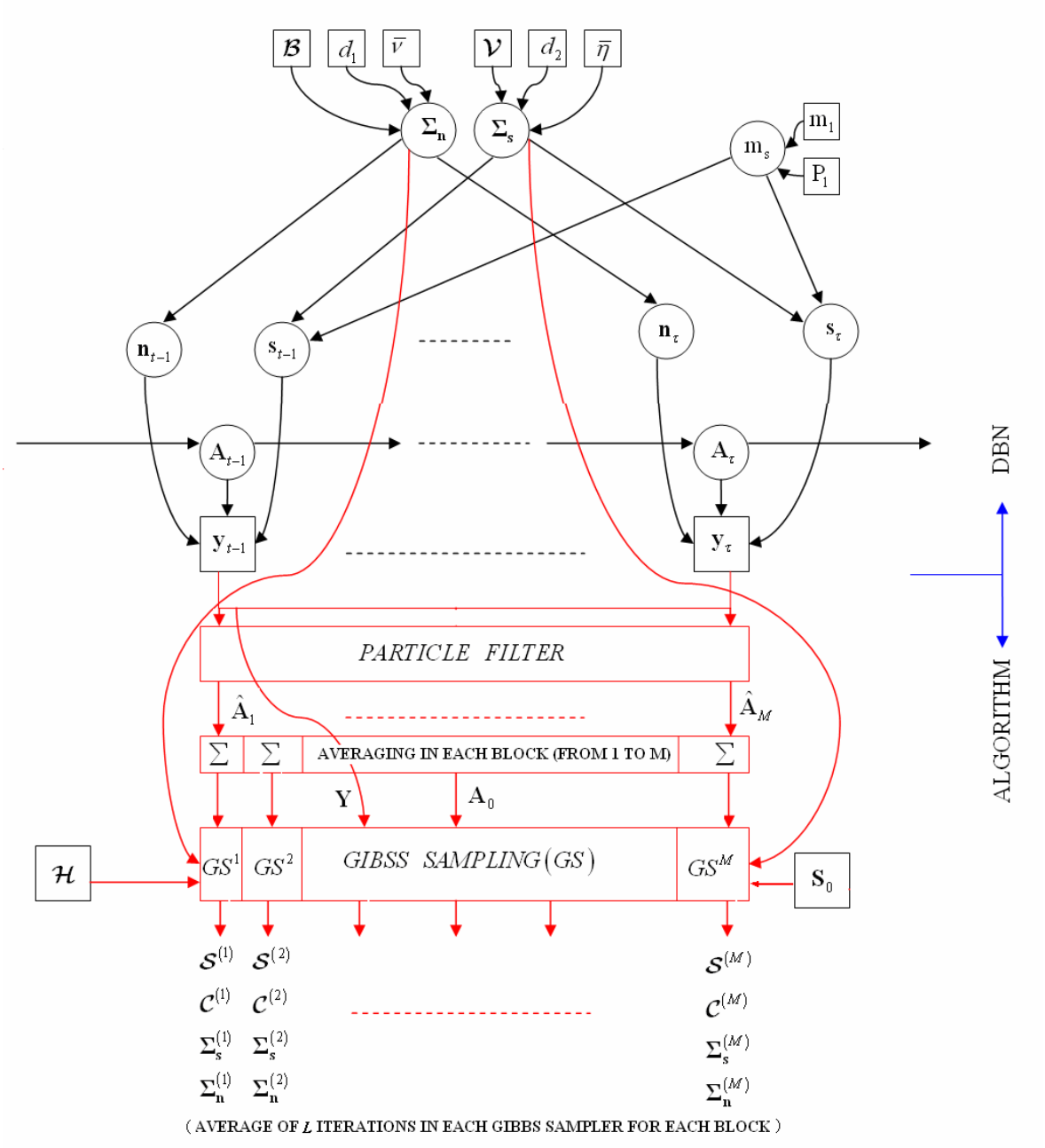


Figure 6.1. DBN and Modeling scheme of MCS technique

In Figure 6.1., a hierarchical model is represented where squares and circles denote the known and unknown parameters, respectively.

In order to model the most general scenario, mixtures of cross-correlated AR sources are considered next. Moreover, in order to propose a unifying methodology for all possible

scenarios, local stationarity assumption of the data is dropped here. Thus, utilization of the previous hybrid method becomes insufficient in this case. Therefore, in cases where we do not have strong beliefs about the source statistics (unlike the preceding case in MCS), we need to model an artificial state-transition equation for all unknown states and use a single sequential Monte Carlo framework. That is why, a hierarchical importance function is utilized here by the appropriate state-transition equation modeling. Unlike the introductory scenario of MCS technique, to consider time structures of the sources, they are modeled by cross-correlated VAR structures as described in Chapters 2 and 5. However, this time their mixtures are observed instead of their original samples, unlike the direct observation of the autoregressive processes in the preceding chapters. This difficulty avoids the use of previously proposed artificial state-transition equations, since we cannot obtain information from the observations directly, which was performed by the discounting of old measurements in the preceding chapters to approximate the optimal importance function. Therefore, we propose to use a *hybrid importance function* here, where a combination of the current values and the previous MAP estimates of the sources are utilized. By this technique, information from the observation data is considered, since it affects the previous MAP estimation of the sources. Moreover, current values of the sources used in this mixture are obtained by proposing an artificial state-transition equation where all parameters regarding the sources (their VAR model parameters), mixing matrix elements and the unknown observation noise process statistics are treated as state variables. Additionally, hierarchical sampling from these transitions is necessary to obtain successful results. After drawing particles, their importance weights are calculated and then resampling is performed. We will name this second technique as the "Mixture Modeling of Cross-Correlated AR Sources (MCARS)" from now on. In order to clarify these, two methods are summarized in the following table:

Table 6.1. Proposed techniques for modeling non-stationary mixtures of cross-correlated processes

| | |
|---|---|
| MCS | Modeling of non-stationary mixtures of cross-correlated processes (Temporally uncorrelated sources ) |
| MCARS | Modeling of non-stationary mixtures of cross-correlated AR processes (Sources possess a VAR structure in time) |

Below, DBN and the modeling scheme are illustrated, where squares and circles represent the known (or observed) and unknown parameters, respectively.

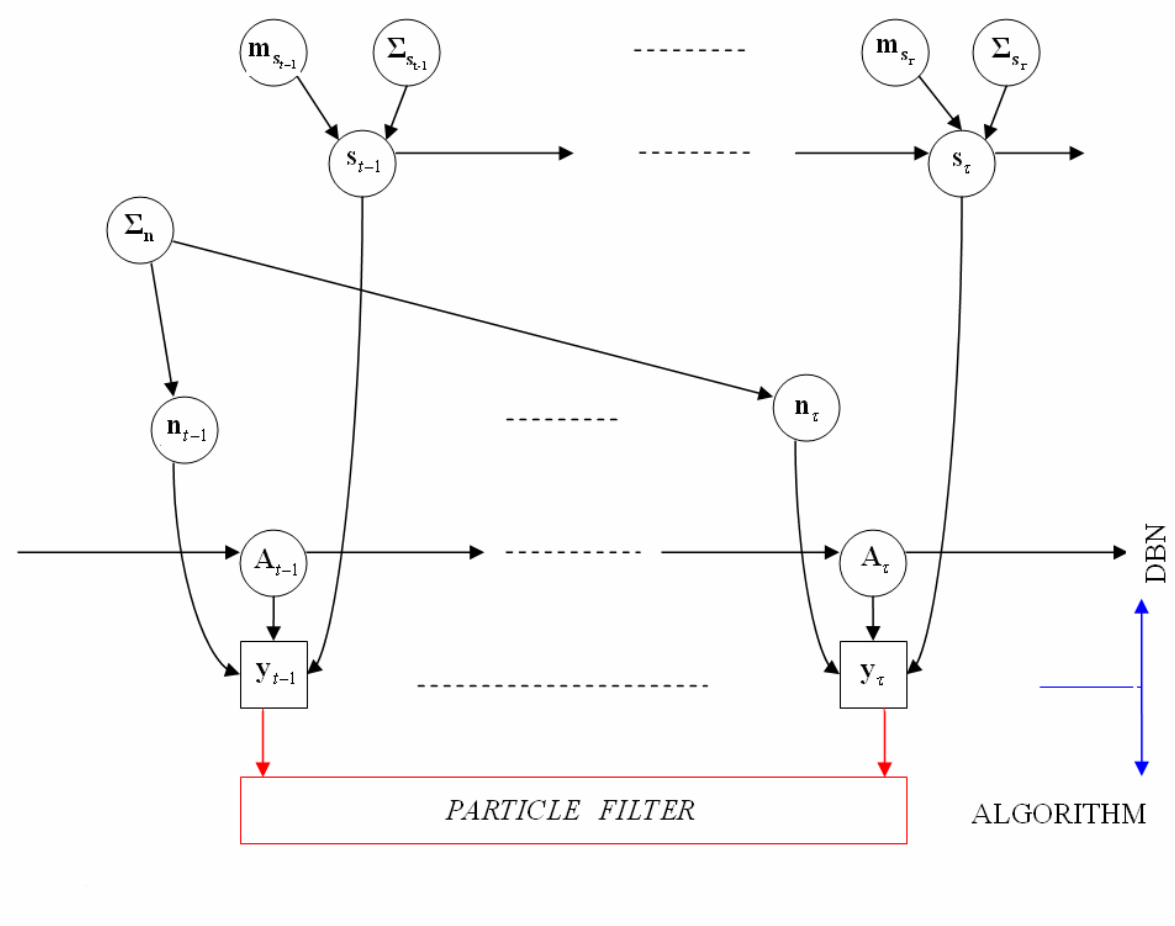


Figure 6.2. DBN and modeling scheme of MCARS technique

Above, it should be noted that cross-correlated source vectors, $\mathbf{s}_{t-1}$ through $\mathbf{s}_\tau$, possess a time structure, unlike the cross-correlated source vectors shown in Figure 6.1.

Next, both algorithms are presented. Then, they are justified with computer simulations. The successful performance of the proposed techniques serves as a promising result for future developments in this area.

First, we present a general model to formulate a non-stationary mixture of cross-correlated processes. Here, the observed signal is equal to the multiplication of two unknowns: mixing matrix and the source vector, as shown in the following mixture signal:

$$\mathbf{y}_t = \mathbf{m} + \mathbf{A}_t \mathbf{s}_t + \mathbf{n}_t \tag{6.5}$$

where $\mathbf{y}_t, \mathbf{m}, \mathbf{A}_t, \mathbf{s}_t, \mathbf{n}_t$ denote the mixture, overall constant mean, mixing matrix, source and the noise matrices, respectively. These matrices are represented as follows: $\mathbf{y}_t = \left(y_1(t),..., y_{d_1}(t)\right)^T$ , $\mathbf{m} = \left(m_1,..., m_{d_1}\right)^T$ , $\mathbf{s}_t = \left(s_1(t),..., s_{d_2}(t)\right)^T$ , $\mathbf{n}_t = \left(n_1(t),..., n_{d_1}(t)\right)^T$ $\mathbf{A}_t = \left(\mathbf{a}_1^T(t), \mathbf{a}_2^T(t).....,\mathbf{a}_{d_1}^T(t)\right)^T$ . Here, source and noise processes are taken to be cross correlated and spatio-temporally independent, respectively:

$$E\left[\left(\mathbf{s}_{t-t_1} - E\left[\mathbf{s}_{t-t_1}\right]\right)\left(\mathbf{s}_{t-t_2} - E\left[\mathbf{s}_{t-t_2}\right]\right)^T\right] = \mathbf{\Sigma}_{\mathbf{s},(t_2-t_1)} \tag{6.6.a}$$

$$E\left[\left(\mathbf{s}_{t-t_1} - E\left[\mathbf{s}_{t-t_1}\right]\right)\left(\mathbf{s}_{t-t_2} - E\left[\mathbf{s}_{t-t_2}\right]\right)^T\right] = \begin{cases} \mathbf{\Sigma}_{\mathbf{s}}, & t_1 = t_2 \\ \mathbf{0}, & t_1 \neq t_2 \end{cases} \tag{6.6.b}$$

$$E\left[\left(\mathbf{n}_{t-t_1} - E\left[\mathbf{n}_{t-t_1}\right]\right)\left(\mathbf{n}_{t-t_2} - E\left[\mathbf{n}_{t-t_2}\right]\right)^T\right] = \begin{cases} \mathbf{\Sigma}_{\mathbf{n}}, & t_1 = t_2 \\ \mathbf{0}, & t_1 \neq t_2 \end{cases} \tag{6.7}$$

where $\mathbf{\Sigma}_{\mathbf{s},(t_2-t_1)}$ and $\mathbf{\Sigma}_{\mathbf{s}}$ are positive definite, non-diagonal covariance matrices. (6.6.a) and (6.6.b) are valid in cases of MCARS and MCS scenarios, respectively. Covariance matrix of noise, however, has a constant and diagonal form.

In (6.5), it is seen that the observation signal is equal to the product of a mixing matrix and a source vector, assuming that **m = 0** without loss of generality. Should the estimation of $\mathbf{A}_t$ and $\mathbf{s}_t$ is performed by a Bootstrap particle filter, the elements of these two quantities are to be modeled as states, i.e. $\mathbf{x}_t = \{\mathbf{A}_t, \mathbf{s}_t\}$. Then the importance weight of each particle set $\mathbf{x}_t^{(i)}$ needs to be elaborated according to the following *likelihood function* (Table 3.3):

$$\begin{aligned} w_t^{(i)} &= p\left(\mathbf{y}_t \mid \mathbf{x}_t^{(i)}\right) \\ p\left(\mathbf{y}_t \mid \mathbf{x}_t^{(i)}\right) &= \mathcal{N}\left(\mathbf{y}_t; \mathbf{A}_t^{(i)}\mathbf{s}_t^{(i)}, \mathbf{\Sigma}_{\mathbf{n}}\right) \end{aligned} \tag{6.8}$$

where the mean and the covariance matrices are represented by $\mathbf{A}_t^{(i)}\mathbf{s}_t^{(i)}$ and $\mathbf{\Sigma_n}$, respectively. In (6.8), it should be noted that both the mixing matrix and the source vector are unknown, leading to an ambiguity between the two, product of which can provide the observed signal value even if their individual values may be wrong. That is, there is no constraining parameter in DCA unlike the *past values of the observation signal* in the direct AR modeling scheme of the previous chapters, i.e. the observation does not regress on its past samples. Therefore, choice of an importance function is a highly challenging problem here. Before discussing this difficult problem, a simpler case (scenario of MCS), is discussed next.

## 6.2. Modeling non-stationary Mixtures of Cross-Correlated Sources (MCS)

In the preceding section, the ambiguity problem in DCA has been mentioned. In this sub-problem, it is assumed that we have strong beliefs about the statistics of the sources. This assumption avoids the ambiguity problem and allows the evaluation of the importance weights corresponding to the particles of $\mathbf{A}_t$ states via the integration of the other state variables (sources). The possibility of performing such an integration is due to the exploitation of *a priori* information. By this simplification, approximation of the optimal importance function can be performed for the particles of $\mathbf{A}_t$ states.

Having reduced the construction of an importance function for $\mathbf{x}_t = \{\mathbf{A}_t, \mathbf{s}_t\}$ to that of $\mathbf{A}_t$ only, particle filtering is performed to estimate this state variable. On the other hand, source waveforms cannot be obtained by this particle filtering since they are integrated out. So, an additional approach is utilized here to separate the source waveforms having found the mixing matrix elements by particle filtering. It is assumed that the mixing matrix elements are slowly varying and they can be taken to be as constants in small data blocks. By this approach, the MCMC method proposed in (Rowe, 2003) can be used in these data blocks which are assumed to be stationary. The mixing matrix $\hat{\mathbf{A}}_t$ estimated by the particle filter forms the *informative prior* for the mixing matrix that is needed in the MCMC algorithm. As a result, a *coupled hybrid method* is proposed where the mixing matrix is estimated by particle filtering, whereas the sources are extracted by a MCMC method. The

MCMC method proposed in (Rowe, 2003) is explained next, since it is utilized for the estimation of sources in *stationary* assumed small data blocks. Then, the utilization of this MCMC approach within the hybrid method will be introduced.

### 6.2.1. Stationary Source Separation (MCMC part of MCS)

The non-stationary mixture model of (6.5) is expressed as follows when stationary cases are considered:

$$\mathbf{y}_t = \mathbf{m} + \mathbf{A}\mathbf{s}_t + \mathbf{n}_t \tag{6.9}$$

where the time-variation of the mixing matrix is removed and it is considered as an unknown *constant* matrix. Moreover, sources are assumed to be *stationary* in this sub-problem. (6.9) represents a mixture model at a given time instant *t*. In order to estimate the unknowns of (6.9), $\tau'$ number of observations are collected and the following matrix notation is formed:

$$\mathbf{Y} = \mathbf{e}_{\tau'}\mathbf{m}^T + \mathcal{S}\mathbf{A}^T + \mathbf{N} \tag{6.10}$$

where $\mathbf{Y} = \left(\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{\tau'}\right)^T$, $\mathcal{S} = \left(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_{\tau'}\right)^T$, $\mathbf{N} = \left(\mathbf{n}_1, \mathbf{n}_2, ..., \mathbf{n}_{\tau'}\right)^T$. Here, $\mathbf{e}_{\tau'}$ denotes a $\tau'$-dimensional vector of ones (Rowe, 2003). By augmenting the $\mathbf{m}$ and the $\mathbf{e}_{\tau'}$ vectors, (6.10) is put into a more compact form as follows:

$$\mathbf{Y} = \mathcal{Z}\mathcal{C}^T + \mathbf{N} \tag{6.11}$$

where $\mathcal{C} = \left(\mathbf{m}, \mathbf{A}\right)$ and $\mathcal{Z} = \left(\mathbf{e}_{\tau'}, \mathcal{S}\right)$. Then, the likelihood function is given as follows:

$$p(\mathbf{Y} \mid \mathcal{C}, \mathcal{Z}, \mathbf{\Sigma}_n) \propto \left|\mathbf{\Sigma}_n\right|^{-\tau'/2} \exp\left(-\frac{1}{2} tr\left(\mathbf{Y} - \mathcal{Z}\mathcal{C}^T\right)\mathbf{\Sigma}_n^{-1}\left(\mathbf{Y} - \mathcal{Z}\mathcal{C}^T\right)^T\right) \tag{6.12}$$

where $\mathbf{\Sigma}_n$ denotes the diagonal covariance matrix of the noise vector and *tr*(.) represents the trace operator. Distribution in (6.12) is known as a Matrix-Normal distribution (Rowe, 2003). For the model of (6.11), the following conjugate prior distributions can be used for

the model parameters (Rowe, 2003), where $\mathcal{MN}$ and $\mathcal{IW}$ stand for Matrix Normal and Inverted Wishart distributions, respectively. In the following equations, $\mathbf{S}_0, \mathbf{C}_{0,} \bar{\eta}, \bar{v}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{H}}$ denote the hyperparameters, through which, *a priori* information is exploited.

$$p(\boldsymbol{\mathcal{S}}, \boldsymbol{\Sigma}_{\mathbf{s}}, \boldsymbol{\mathcal{C}}, \boldsymbol{\Sigma}_{\mathbf{n}}) = p(\boldsymbol{\mathcal{S}} \mid \boldsymbol{\Sigma}_{\mathbf{s}})\, p(\boldsymbol{\Sigma}_{\mathbf{s}})\, p(\boldsymbol{\mathcal{C}} \mid \boldsymbol{\Sigma}_{\mathbf{n}})\, p(\boldsymbol{\Sigma}_{\mathbf{n}}) \tag{6.13.a}$$

$$p(\boldsymbol{\mathcal{S}} \mid \boldsymbol{\Sigma}_{\mathbf{s}}) \propto \left|\boldsymbol{\Sigma}_{\mathbf{s}}\right|^{-\tau'/2} \exp\left( -\frac{1}{2} tr\left(\boldsymbol{\mathcal{S}} - \mathbf{S}_0\right) \boldsymbol{\Sigma}_{\mathbf{s}}^{-1} \left(\boldsymbol{\mathcal{S}} - \mathbf{S}_0\right)^T \right) \rightarrow \mathcal{MN} \tag{6.13.b}$$

$$p(\boldsymbol{\Sigma}_{\mathbf{s}}) \propto \left|\boldsymbol{\Sigma}_{\mathbf{s}}\right|^{-\bar{\eta}/2} \exp\left( -\frac{1}{2} tr \boldsymbol{\Sigma}_{\mathbf{s}}^{-1} \boldsymbol{\mathcal{V}} \right) \rightarrow \mathcal{IW}(\boldsymbol{\mathcal{V}}, d_2, \bar{\eta}), \tag{6.13.c}$$

$$p(\boldsymbol{\Sigma}_{\mathbf{n}}) \propto \left|\boldsymbol{\Sigma}_{\mathbf{n}}\right|^{-\bar{v}/2} \exp\left( -\frac{1}{2} tr \boldsymbol{\Sigma}_{\mathbf{n}}^{-1} \boldsymbol{\mathcal{B}} \right) \rightarrow \mathcal{IW}(\boldsymbol{\mathcal{B}}, d_1, \bar{v}), \tag{6.13.d}$$

$$p(\boldsymbol{\mathcal{C}} \mid \boldsymbol{\Sigma}_{\mathbf{n}}) \propto \left|\boldsymbol{\mathcal{H}}\right|^{-d_2/2} \left|\boldsymbol{\Sigma}_{\mathbf{n}}\right|^{-(d_1+1)/2} \exp\left( -\frac{1}{2} tr \boldsymbol{\Sigma}_{\mathbf{n}}^{-1} \left(\boldsymbol{\mathcal{C}} - \mathbf{C}_0\right) \boldsymbol{\mathcal{H}}^{-1} \left(\boldsymbol{\mathcal{C}} - \mathbf{C}_0\right)^T \right) \rightarrow \mathcal{MN} \tag{6.13.e}$$

Above, $\boldsymbol{\Sigma}_{\mathbf{s}}$ denotes the covariance matrix of the sources and it is not constrained to be diagonal and all of its elements are left free, to model the dependencies between the sources. After some algebra performed on the conjugate priors and the likelihood function, the posterior conditionals are obtained to be utilized in the Gibbs sampling (Rowe, 2003). By means of cycling through the posterior conditionals given below, Gibbs sampling can be performed in order to estimate the model parameters (Rowe, 2003).

$$p(\boldsymbol{\mathcal{C}} \mid \boldsymbol{\mathcal{S}}, \boldsymbol{\Sigma}_{\mathbf{s}}, \boldsymbol{\Sigma}_{\mathbf{n}}, \mathbf{Y}) \propto \exp\left( -\frac{1}{2} tr \boldsymbol{\Sigma}_{\mathbf{n}}^{-1} (\boldsymbol{\mathcal{C}} - \tilde{\boldsymbol{\mathcal{C}}})(\boldsymbol{\mathcal{H}}^{-1} + \boldsymbol{\mathcal{Z}}^T \boldsymbol{\mathcal{Z}})(\boldsymbol{\mathcal{C}} - \tilde{\boldsymbol{\mathcal{C}}})^T \right), \tag{6.14.a}$$

$$\tilde{\boldsymbol{\mathcal{C}}} = \left(\mathbf{C}_0 \boldsymbol{\mathcal{H}}^{-1} + \mathbf{Y}^T \boldsymbol{\mathcal{Z}}\right)\left(\boldsymbol{\mathcal{H}}^{-1} + \boldsymbol{\mathcal{Z}}^T \boldsymbol{\mathcal{Z}}\right)^{-1} \tag{6.14.b}$$

$$p(\boldsymbol{\Sigma}_{\mathbf{n}} \mid \boldsymbol{\mathcal{S}}, \boldsymbol{\Sigma}_{\mathbf{s}}, \boldsymbol{\mathcal{C}}, \mathbf{Y}) \propto \left|\boldsymbol{\Sigma}_{\mathbf{n}}\right|^{-(\tau'+\bar{v}+d_1+1)/2} \exp\left( -\frac{1}{2} tr \boldsymbol{\Sigma}_{\mathbf{n}}^{-1} \boldsymbol{\mathcal{G}} \right), \tag{6.14.c}$$

$$\boldsymbol{\mathcal{G}} = \left(\mathbf{Y} - \boldsymbol{\mathcal{Z}}\boldsymbol{\mathcal{C}}^T\right)^T \left(\mathbf{Y} - \boldsymbol{\mathcal{Z}}\boldsymbol{\mathcal{C}}^T\right) + \left(\boldsymbol{\mathcal{C}} - \mathbf{C}_0\right) \boldsymbol{\mathcal{H}}^{-1} \left(\boldsymbol{\mathcal{C}} - \mathbf{C}_0\right)^T + \boldsymbol{\mathcal{B}} \tag{6.14.d}$$

$$p(\boldsymbol{\Sigma}_{\mathbf{s}} \mid \boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{S}}, \boldsymbol{\Sigma}_{\mathbf{n}}, \mathbf{Y}) \propto \left|\boldsymbol{\Sigma}_{\mathbf{s}}\right|^{-(\tau'+\bar{\eta})/2} \exp\left( -\frac{1}{2} tr \boldsymbol{\Sigma}_{\mathbf{s}}^{-1} \left[ \left(\boldsymbol{\mathcal{S}} - \mathbf{S}_0\right)^T \left(\boldsymbol{\mathcal{S}} - \mathbf{S}_0\right) + \boldsymbol{\mathcal{V}} \right] \right), \tag{6.14.e}$$

$$p(\boldsymbol{\mathcal{S}} \mid \boldsymbol{\mathcal{C}}, \boldsymbol{\Sigma}_{\mathbf{s}}, \boldsymbol{\Sigma}_{\mathbf{n}}, \mathbf{Y}) \propto \exp\left( -\frac{1}{2} tr\left(\boldsymbol{\mathcal{S}} - \tilde{\boldsymbol{\mathcal{S}}}\right)\left(\boldsymbol{\Sigma}_{\mathbf{s}}^{-1} + \mathbf{A}^T \boldsymbol{\Sigma}_{\mathbf{n}}^{-1} \mathbf{A}\right)\left(\boldsymbol{\mathcal{S}} - \tilde{\boldsymbol{\mathcal{S}}}\right)^T \right), \quad (6.14.f)$$

$$\tilde{\boldsymbol{\mathcal{S}}} = \left[ \mathbf{S}_0 \boldsymbol{\Sigma}_{\mathbf{s}}^{-1} + \left(\mathbf{Y} - \mathbf{e}_\tau \mathbf{m}^T\right) \boldsymbol{\Sigma}_{\mathbf{n}}^{-1} \mathbf{A} \right] \left(\boldsymbol{\Sigma}_{\mathbf{s}}^{-1} + \mathbf{A}^T \boldsymbol{\Sigma}_{\mathbf{n}}^{-1} \mathbf{A}\right)^{-1} \quad (6.14.g)$$

### 6.2.2. Non-stationary Mixing Matrix Estimation (Particle Filtering part of MCS)

In the preceding section, the MCMC technique for source separation is explained, where *the model parameters are assumed to be constants within the observed data block*, i.e. *stationary mixtures* are involved. Here, in order to handle *non-stationarities*, the utilization of this MCMC approach is proposed within a hybrid method. Here, the objective is to separate *spatially dependent* Gaussian sources from their mixtures, where the mixing system is *time-varying* unlike the scenario given in the previous section. That is, the elements of the mixing matrix in (6.5) change over time. So, (6.5) can be written as follows:

$$\mathbf{y}_t = \mathbf{m} + \mathbf{A}_t \mathbf{s}_t + \mathbf{n}_t \quad (6.15)$$

where $\mathbf{A}_t = \left(\mathbf{a}_1^T(t), \mathbf{a}_2^T(t), ..., \mathbf{a}_{d_1}^T(t)\right)^T$. Here, the use of particle filtering is proposed to estimate *time-varying* elements of the mixing matrix. Also, it is assumed that there is no constant mean in the mixture, i.e. $\mathbf{m} = \mathbf{0}$, and there is *a priori* information about the statistics of the sources and the noise. Even if all the statistics of the sources and the noise are known *a priori*, separating the *time evolution* of the mixing matrix and the sources needs two sets of states in the particle filtering model, i.e. $\mathbf{x}_t^{(i)} = \left\{\mathbf{s}_t^{(i)}, \tilde{\mathbf{A}}_t^{(i)}\right\}$ for *i = 1,2, …, N*, at any time instant *t* and $\tilde{\mathbf{A}}_t = vec\left(\mathbf{A}_t\right)$. Since particle filter estimates the *joint* pdf given by $p\left(\mathbf{x}_{0:t} \mid \mathbf{y}_{1:t}\right)$; sources, must be integrated out from this *joint* pdf in order to obtain the *marginal* pdf estimate of the mixing matrix $\tilde{\mathbf{A}}_t^{(i)}$. From (6.8), it is seen that the importance weights of each particle is expressed in terms of the likelihood function (since Bootstrap particle filtering is used). To obtain the marginal importance weight of $\tilde{\mathbf{A}}_t^{(i)}$, the likelihood function is integrated with respect to the sources:

$$p\left(\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}\right) \propto \int p\left(\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}, \mathbf{s}_t^{(i)}\right) p\left(\mathbf{s}_t^{(i)}\right) d\mathbf{s}_t^{(i)} \tag{6.16}$$

From (6.15), it is seen that $p\left(\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}, \mathbf{s}_t^{(i)}\right)$ conditional density has a Gaussian distribution. Since the sources are also Gaussian distributed, the integration (6.16) has also a Gaussian distribution. Thus, instead of estimating (6.16) for each particle, only mean and variance estimations can be found respectively as follows, using Monte Carlo integration:

$$\hat{\mathbf{m}}_y = E\left[\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}\right] = \int \mathbf{y}_t \, p\left(\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}\right) d\mathbf{y}_t \approx \frac{1}{N} \sum_{l=1}^{N} \left(\mathbf{y}_t^{(l)}\right) \tag{6.17.a}$$

$$\begin{aligned} \hat{\mathbf{\Sigma}}_y &= E\left[\left(\mathbf{y}_t - E\left[\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}\right]\right)^2\right] = \int \left(\mathbf{y}_t - E\left[\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}\right]\right)^2 p\left(\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}\right) d\mathbf{y}_t \\ &\approx \frac{1}{N} \sum_{l=1}^{N} \left(\mathbf{y}_t^{(l)} - E\left[\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}\right]\right)^2 \end{aligned} \tag{6.17.b}$$

where samples $\mathbf{y}_t^{(l)}$, can be easily drawn from the Gaussian pdf, given by (6.16), i.e. $\mathbf{y}_t^{(l)} \sim p\left(\mathbf{y}_t \middle| \tilde{\mathbf{A}}_t^{(i)}\right) = \mathcal{N}\left(\hat{\mathbf{m}}_y, \hat{\mathbf{\Sigma}}_y\right)$. Here, $N$ denotes the total number of particles. By these operations, importance weight calculation, (6.8), turns into the following form:

$$w_t^{(i)} = p\left(\mathbf{y}_t \mid \mathbf{x}_t^{(i)}\right) \approx p\left(\mathbf{y}_t \mid \tilde{\mathbf{A}}_t^{(i)}\right) \tag{6.18}$$

where $p(\mathbf{y}_t \mid \tilde{\mathbf{A}}_t^{(i)})$ is Gaussian, whose mean and covariance matrices can be found by (6.17). In addition, *a priori* state transition, which is discussed above, is modeled by an artificial random walk, as shown below:

$$\tilde{\mathbf{A}}_t = \tilde{\mathbf{A}}_{t-1} + \mathbf{v}_{a,t} \tag{6.19}$$

where $\mathbf{v}_{a,t} \sim N(\mathbf{0}, \mathbf{Q})$ and $\mathbf{Q}$ is diagonal. (6.15) and (6.19) fits into the general state-space formulation of the particle filtering, as shown below for the sake of completeness:

$$\begin{aligned}\tilde{\mathbf{A}}_t &= \tilde{\mathbf{A}}_{t-1} + \mathbf{v}_{a,t} \\ \mathbf{y}_t &= \mathbf{A}_t \mathbf{s}_t + \mathbf{n}_t\end{aligned} \tag{6.20}$$

Having found the time evolution of the mixing matrix elements by integrating the sources as outlined above, sources also need to be extracted from the mixtures. Had the mixing matrix coefficients been constant over time, the MCMC method given Section 6.2.1 could have been used for estimating the sources. However, this is not the case in this scenario. Thus, it is assumed that the elements of the mixing matrix do not change considerably over small blocks of data. In this case, one may use MCMC in these blocks, however the choice of the appropriate prior distributions of the mixing matrix elements arises as a problem. This is due to the requirement of using *informative prior distributions* for the parameters given in (6.13) (see (Rowe, 2003), pp. 56).

Since estimates of these mixing matrix elements can be obtained by using the particle filtering scheme, which is explained above, it is proposed to use these estimates in order to form some *approximate* informative priors for the mixing matrix elements, which are denoted by $\mathbf{C_0}$ in (6.13) and (6.14). A pseudo-code of the proposed method is outlined in Table 6.2.

Table 6.2. Pseudo-code of MCS technique

**PARTICLE FILTER PART**

I. Initialize the particles $\tilde{\mathbf{A}}_0^{(i)} \sim \mathcal{N}(\boldsymbol{\mu}_A, \mathbf{P}_A)$ , $w_0^{(i)} = 1$ for $i = 1, 2, ..., N$ where $\mathbf{P}_A$ is a diagonal covariance matrix.

II. Draw new samples for *i = 1, 2, …, N*

$$\mathbf{x}_t^{(i)} = \left\{ \mathbf{s}_t^{(i)}, \tilde{\mathbf{A}}_t^{(i)} \right\} \; : \; \tilde{\mathbf{A}}_t^{(i)} \sim \mathcal{N}\left( \tilde{\mathbf{A}}_{t-1}^{(i)}, \mathbf{Q} \right) \quad \text{and} \quad \mathbf{s}_t \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{s}, \boldsymbol{\Sigma}_\mathbf{s})$$

Hyperparameters: $\mathbf{v}_a \sim N(\mathbf{0}, \mathbf{Q})$, $\boldsymbol{\mu}_\mathbf{s} \sim N(\boldsymbol{\mu}_\mathbf{1}, \mathbf{P}_\mathbf{1})$, $\boldsymbol{\Sigma}_\mathbf{s} \sim \mathcal{IW}(\boldsymbol{\mathcal{V}}, d_2, \bar{\eta})$

III. Calculate the importance weights by integrating out the sources:

Table 6.2. (continued)

$$w_t^{(i)} = p(\mathbf{y}_t \mid \tilde{\mathbf{A}}_t^{(i)})$$

where $p(\mathbf{y}_t \mid \tilde{\mathbf{A}}_t^{(i)}) = \mathcal{N}(\mathbf{y}_t; \hat{\mathbf{m}}_y, \hat{\boldsymbol{\Sigma}}_y)$ and $\mathcal{N}(\mathbf{y}_t; \hat{\mathbf{m}}_y, \hat{\boldsymbol{\Sigma}}_y)$ denotes the evaluation of the observation data $\mathbf{y_t}$ at the Gaussian pdf, whose mean and covariance matrices are given by (6.17)

IV. Normalize the importance weights:

$$\tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j=1}^{N} w_t^{(i)}}, \; i = 1, ..., N$$

V. Resample at each iteration from $\left\{\mathbf{A}_t^{(i)}, \tilde{w}_t^{(i)}\right\}_{i=1}^{N}$ and make the unnormalized importance weights equal to each other.

VI. Go to Step (II) and repeat.

**MCMC PART**

I. Estimate the Minimum Mean Square Error (MMSE) estimate of $\hat{\mathbf{A}}$ as follows:

$$\hat{\mathbf{A}}_t = \int \mathbf{A}_{0:t} p\left(\mathbf{A}_{0:t} \mid \mathbf{y}_{1:t}\right) d\mathbf{A}_{0:t} \approx \sum_{i=1}^{N} \mathbf{A}_{0:t}^{(i)} \tilde{w}_t^{(i)}$$

II. For data blocks of size $\tau'$, calculate the mean of each mixing matrix element and use that as the mixing matrix prior in the first equation of (6.14):

$$\mathbf{A}_0 = \frac{1}{\tau'} \sum_{i=1}^{\tau'} \hat{\mathbf{A}}_i \rightarrow \mathbf{C}_0 = \left(\mathbf{m}_0, \mathbf{A}_0\right)$$

For other parameters, use the priors given in (6.13).

III. Then cycle through the Gibbs iterations by using the posterior conditionals

i) Start with the initial $\mathbf{S}_0, \boldsymbol{\Sigma}_{\mathbf{n,0}}$,

ii) Cycle through the following posteriors:

$$\boldsymbol{\mathcal{C}}_{(r+1)} \sim p\left(\boldsymbol{\mathcal{C}} \mid \boldsymbol{\mathcal{S}}_{(r)}, \boldsymbol{\Sigma}_{\mathbf{s}(r)}, \boldsymbol{\Sigma}_{\mathbf{n}(l)}, \mathbf{Y}\right),$$

$$\boldsymbol{\Sigma}_{\mathbf{n},(r+1)} \sim p\left(\boldsymbol{\Sigma}_{\mathbf{n}} \mid \boldsymbol{\mathcal{S}}_{(r)}, \boldsymbol{\Sigma}_{\mathbf{s}(r)}, \boldsymbol{\mathcal{C}}_{(r+1)}, \mathbf{Y}\right)$$

$$\boldsymbol{\Sigma}_{\mathbf{s}(r+1)} \sim p\left(\boldsymbol{\Sigma}_{\mathbf{s}} \mid \boldsymbol{\mathcal{S}}_{(r)}, \boldsymbol{\mathcal{C}}_{(r+1)}, \boldsymbol{\Sigma}_{\mathbf{n},(r+1)}, \mathbf{Y}\right)$$

$$\boldsymbol{\mathcal{S}}_{(r+1)} \sim p\left(\boldsymbol{\mathcal{S}} \mid \boldsymbol{\Sigma}_{\mathbf{s}(r+1)}, \boldsymbol{\mathcal{C}}_{(r+1)}, \boldsymbol{\Sigma}_{\mathbf{n},(r+1)}, \mathbf{Y}\right)$$

iii) Discard the variates of burn in period and estimate the parameters as follows:

$$\boldsymbol{\mathcal{S}} = \frac{1}{L}\sum_{r=1}^{L} \boldsymbol{\mathcal{S}}_r, \boldsymbol{\Sigma}_{\mathbf{s}} = \frac{1}{L}\sum_{r=1}^{L} \boldsymbol{\Sigma}_{\mathbf{s}(r)}, \boldsymbol{\mathcal{C}} = \frac{1}{L}\sum_{r=1}^{L} \boldsymbol{\mathcal{C}}_r, \boldsymbol{\Sigma}_{\mathbf{n}} = \frac{1}{L}\sum_{r=1}^{L} \boldsymbol{\Sigma}_{\mathbf{n}(r)}$$

### 6.3. Modeling Non-stationary Mixtures of Cross-Correlated AR Sources (MCARS)

Previously, it was stated that an ambiguity trouble arises between the unknown mixing matrix and the source vector during the modeling of the DCA problem. The reason for this difficulty has been explained by the inability to construct an appropriate approximation of the optimal importance function in case of unknown state transitions. Thus, to overcome this difficulty, it was assumed that we know the statistics of the sources. As a result of this assumption, we got rid of modeling the transitions of the source variables. Instead, these state variables are integrated out to estimate the mixing matrix by means of a particle filter. However, waveforms of the sources could not be obtained by this particle filter and a MCMC scheme was needed to extract these source waveforms, given the estimates of the mixing matrix elements found by the particle filter. On the other hand, one cannot have *informative a priori* information about the sources, always. Thus, a more general method needs to be developed for such cases. Next, a new particle filtering technique is proposed, where the need for *informative a priori* information about the sources is avoided. Here, a *hybrid importance function* is proposed to sample particles of both the mixing matrix elements and the sources, unlike the previous case. As a result of this method, a general solution is provided to estimate non-stationary mixtures of cross correlated AR processes. The general state-space formulation for this problem is given below:

$$\mathbf{s}_t = \mathbf{\Phi}_{1,t}\mathbf{s}_{t-1} + \mathbf{\Phi}_{2,t}\mathbf{s}_{t-2} + \cdots + \mathbf{\Phi}_{K,t}\mathbf{s}_{t-K} + \boldsymbol{\upsilon}_t \tag{6.21.a}$$

$$\mathbf{y}_t = \mathbf{A}_t\mathbf{s}_t + \mathbf{n}_t \tag{6.21.b}$$

where $\mathbf{y}_t = \left[ y_{1,t}, y_{2,t}, \ldots, y_{d_1,t} \right]^T$ , $\mathbf{s}_t = \left( s_1(t), ..., s_{d_2}(t) \right)^T$ , $\mathbf{A}_t = \left( \mathbf{a}_1^T(t), \mathbf{a}_2^T(t), ..., \mathbf{a}_{d_1}^T(t) \right)^T$ , $\mathbf{n}_t = \left( n_1(t), ..., n_{d_1}(t) \right)^T$ and $\boldsymbol{\upsilon}_t = \left( \upsilon_1(t), ..., \upsilon_{d_1}(t) \right)^T$ denote the mixture, source, mixing matrix, observation and process noise matrices, respectively. $\mathbf{\Phi}_{j,t}$ denote $\left( d_2 \text{x} d_2 \right)$ matrix of AR coefficients for $j = 1, 2, \ldots, K$ and the driving noise process is assumed to have a mean and covariance matrix as shown below:

$$E\left[\boldsymbol{\upsilon}_t\right]=\mathbf{0},\quad E\left[\boldsymbol{\upsilon}_{t_1}\boldsymbol{\upsilon}_{t_2}^T\right]=\begin{cases}\boldsymbol{\Sigma}_{\boldsymbol{\upsilon}}, & t_1=t_2\\ \mathbf{0}, & t_1\neq t_2\end{cases} \tag{6.22}$$

where matrix $\boldsymbol{\Sigma}_{\boldsymbol{\upsilon}}$ is a non-diagonal, positive definite matrix, whereas the covariance matrix of the observation noise has a diagonal form, i.e. $E\left[\mathbf{n}_{t_1}\mathbf{n}_{t_2}^T\right]=\boldsymbol{\Sigma}_{\mathbf{n}}$ for $\forall t_1,t_2$.

Here, the use of a particle filtering scheme is proposed with the following definition of the state vector:

$$\mathbf{x}_t=\left[\tilde{\boldsymbol{\Phi}}_t^T,\mathbf{A}_t^T,\mathbf{s}_t^T,\mathbf{n}_t^T,\boldsymbol{\upsilon}_t^T,\boldsymbol{\chi}_t^T,\boldsymbol{\varsigma}_t^T\right]^T \tag{6.23}$$

where $\tilde{\boldsymbol{\Phi}}_t^T,\mathbf{A}_t^T,\mathbf{s}_t^T,\mathbf{n}_t^T,\boldsymbol{\upsilon}_t^T,\boldsymbol{\chi}_t^T$, and $\boldsymbol{\varsigma}_t^T$ correspond to the state variables representing the unknown AR coefficients, mixing matrix elements, sources, measurement and process noise processes, covariance matrix elements of the process and observation noise processes, respectively. They are given as follows:

$$\begin{aligned}
\tilde{\boldsymbol{\Phi}}_t&=\left[\tilde{\boldsymbol{\Phi}}_{1,t}^T,\tilde{\boldsymbol{\Phi}}_{2,t}^T,\ldots,\tilde{\boldsymbol{\Phi}}_{K,t}^T\right]^T,\\
\tilde{\mathbf{A}}_t&=\left[a_1,a_2,\ldots,a_d\right]^T,\\
\mathbf{s}_t&=\left[s_{1,t},\ldots,s_{d_2,t}\right]^T,\\
\boldsymbol{\upsilon}_t&=\left[\upsilon_{1,t},\ldots,\upsilon_{d_2,t}\right]^T,\\
\mathbf{n}_t&=\left[n_{1,t},\ldots,n_{d_2,t}\right]^T,\\
\boldsymbol{\chi}_t&=\left[\chi_{1,t},\chi_{2,t},\ldots,\chi_{d_2(d_2+1)/2}\right]^T,\\
\boldsymbol{\varsigma}_t&=\left[\varsigma_{1,t},\varsigma_{2,t},\ldots,\varsigma_{d_1,t}\right]^T
\end{aligned} \tag{6.24}$$

where $d=d_1$ x $d_2$ and $\varsigma_{l,t}=\log\sigma_{n_{l,t}}^2$, for $l=1,...,d_1$ is used to model the diagonal covariance matrix of the observation noise process as shown below:

$$\boldsymbol{\Sigma}_{\mathbf{n}}=diag\left(\sigma_{n_1}^2,\ldots,\sigma_{n_{d_1}}^2\right) \tag{6.25}$$

In (6.24), logarithms of these variance components are modeled to be as particles. By this approach, each variance component is guaranteed to be kept positive during the state-transitions. This will be clearer in the sequel. Similarly, unknown covariance matrix of the process noise is decomposed into the product of a lower and an upper triangular matrix, i.e. Cholesky decomposition is applied to the covariance matrix $\mathbf{\Sigma}_{\mathbf{v}}$ to keep it as a positive definite matrix during the state transitions. This decomposition is shown below:

$$\mathbf{\Sigma}_{\mathbf{v}_t} = \begin{bmatrix} \chi_{1,t} & 0 & \cdots & 0 \\ \chi_{2,t} & \chi_{d_2+1,t} & 0 & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \chi_{d_2,t} & \cdots & \cdots & \chi_{d_2(d_2+1)/2,t} \end{bmatrix} \begin{bmatrix} \chi_{1,t} & \chi_{2,t} & \cdots & \chi_{d_2,t} \\ 0 & \chi_{d_2+1,t} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \chi_{d_2(d_2+1)/2,t} \end{bmatrix} \tag{6.26}$$

In addition to this, in order to avoid multiple solutions, estimate of matrix $\mathbf{A}_t$ is taken to be constrained in the following form, since same observation can be obtained by dividing $\mathbf{A}_t$ and multiplying $\mathbf{s}_t$ with the same non-zero constant *c*.

$$\begin{array}{ccc} d_1 < d_2 & d_1 = d_2 & d_1 > d_2 \\ \hat{\mathbf{A}}_t = \begin{bmatrix} 1 & \hat{a}_{12} & 1 \\ \hat{a}_{21} & 1 & \hat{a}_{32} \end{bmatrix} & \hat{\mathbf{A}}_t = \begin{bmatrix} 1 & \hat{a}_{12} & \hat{a}_{13} \\ \hat{a}_{21} & 1 & \hat{a}_{23} \\ \hat{a}_{31} & \hat{a}_{32} & 1 \end{bmatrix} & \hat{\mathbf{A}}_t = \begin{bmatrix} 1 & \hat{a}_{12} \\ \hat{a}_{21} & 1 \\ \hat{a}_{31} & \hat{a}_{32} \end{bmatrix} \end{array} \tag{6.27}$$

For the sake of simplicity, this constraint is illustrated for three possible conditions as shown above. It is seen that, one element in each column should be made equal to a constant.

After defining the state variables given in (6.24), the following artificial state transition model is used to model the process equation of the particle filter represented by (3.20.a) in general. In the following equation, any VAR(*K*) process is expressed in terms of a VAR(1) which is illustrated in its last line by using the definitions in (2.12-2.15). For the sake of completeness, these definitions are recapitulated after the presentation of the artificial state transition model of (6.28).

$$
\begin{aligned}
\mathbf{x}_t &= \mathbf{x}_{t-1} + \mathbf{V}_t \\
&\textit{where} \\
\boldsymbol{\chi}_t &= \boldsymbol{\chi}_{t-1} + \mathbf{V}_{\chi,t} \\
\boldsymbol{\varsigma}_t &= \boldsymbol{\varsigma}_{t-1} + \mathbf{V}_{\varsigma,t} \\
\boldsymbol{\upsilon}_t &= \boldsymbol{\upsilon}_{t-1} + \mathbf{V}_{\boldsymbol{\upsilon},t} \\
\mathbf{n}_t &= \mathbf{n}_{t-1} + \mathbf{V}_{\mathbf{n},t} \\
\boldsymbol{\Phi}_t &= \boldsymbol{\Phi}_{t-1} + \mathbf{V}_{\boldsymbol{\Phi},t} \\
\mathbf{A}_t &= \mathbf{A}_{t-1} + \mathbf{V}_{\mathbf{A},t} \\
\mathbf{s}_t &= \mathbf{G}\left(\boldsymbol{\Phi}_t\right)\mathbf{s}_{t-1} + \mathbf{V}_{\mathbf{s},t}\left(\boldsymbol{\upsilon}_t\right)
\end{aligned}
\tag{6.28}
$$

where

$$
\mathbf{G}\left(\boldsymbol{\Phi}_t\right) = \begin{bmatrix}
\boldsymbol{\Phi}_{1,t} & \boldsymbol{\Phi}_{2,t} & \boldsymbol{\Phi}_{3,t} & \cdots & \boldsymbol{\Phi}_{K-1,t} & \boldsymbol{\Phi}_{K,t} \\
\mathbf{I}_{d_1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{I}_{d_1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\
\vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{d_1} & \mathbf{0}
\end{bmatrix} \text{ and } \tilde{\boldsymbol{\upsilon}}_t = \begin{bmatrix} \boldsymbol{\upsilon}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}
\tag{6.29}
$$

$$
E\left[\tilde{\boldsymbol{\upsilon}}_{t_1}\tilde{\boldsymbol{\upsilon}}_{t_2}^T\right] = \begin{cases} \boldsymbol{\Omega}, & t_1 = t_2 \\ \mathbf{0}, & \textit{otherwise} \end{cases} \text{ and } \boldsymbol{\Omega} = \begin{bmatrix}
\boldsymbol{\Sigma}_{\boldsymbol{\upsilon}}(\boldsymbol{\chi}) & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\vdots & \vdots & \cdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \mathbf{0}
\end{bmatrix}
\tag{6.30}
$$

However, it should be noted that samples from the artificial state transition equation of (6.28) must be drawn hierarchically, as shown below:

$$
\begin{aligned}
p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}\right) = {} & p\left(\mathbf{s}_t \middle| \mathbf{s}_{t-1}, \boldsymbol{\Phi}_t, \boldsymbol{\upsilon}_t\right) p\left(\boldsymbol{\Phi}_t \middle| \boldsymbol{\Phi}_{t-1}, \boldsymbol{\upsilon}_t\right) p\left(\mathbf{A}_t \middle| \mathbf{A}_{t-1}, \mathbf{n}_t\right) \\
& \text{x } p\left(\boldsymbol{\upsilon}_t \middle| \boldsymbol{\upsilon}_{t-1}, \boldsymbol{\chi}_t\right) p\left(\mathbf{n}_t \middle| \mathbf{n}_{t-1}, \boldsymbol{\varsigma}_t\right) \\
& \text{x } p\left(\boldsymbol{\varsigma}_t \middle| \boldsymbol{\varsigma}_{t-1}\right) p\left(\boldsymbol{\chi}_t \middle| \boldsymbol{\chi}_{t-1}\right)
\end{aligned}
\tag{6.31}
$$

Here, modeling of (6.28) is used, since the original time variations of the states are unknown. Each component of the noise process $\mathbf{V}_t$ is taken to be a zero mean Gaussian vector of appropriate dimension. Their covariance matrices are taken to be diagonal with constant variance elements known as the drift parameters. As a result of these, new particles are drawn from the following importance functions in the following order:

$$\begin{aligned}
\boldsymbol{\varsigma}_t^{(i)} &\sim \mathcal{N}\left(\boldsymbol{\varsigma}_{t-1}^{(i)}, \sigma_{\mathbf{V}_\varsigma}^2 \mathbf{I}\right) \rightarrow \text{ Form } \boldsymbol{\Sigma}_{\mathbf{n}}^{(i)}\left(\boldsymbol{\varsigma}_t^{(i)}\right) \\
\boldsymbol{\chi}_t^{(i)} &\sim \mathcal{N}\left(\boldsymbol{\chi}_{t-1}^{(i)}, \sigma_{\mathbf{V}_\chi}^2 \mathbf{I}\right) \rightarrow \text{ Form } \boldsymbol{\Sigma}_{\boldsymbol{\upsilon}}^{(i)}\left(\boldsymbol{\chi}_t^{(i)}\right) \\
\boldsymbol{\upsilon}_t^{(i)} &\sim \mathcal{N}\left(\boldsymbol{\upsilon}_{t-1}^{(i)}, \boldsymbol{\Sigma}_{\boldsymbol{\upsilon}}^{(i)}\left(\boldsymbol{\chi}_t^{(i)}\right)\right) \\
\mathbf{n}_t^{(i)} &\sim \mathcal{N}\left(\mathbf{n}_{t-1}^{(i)}, \boldsymbol{\Sigma}_{\mathbf{n}}^{(i)}\left(\boldsymbol{\varsigma}_t^{(i)}\right)\right) \\
\boldsymbol{\Phi}_t^{(i)} &\sim \mathcal{N}\left(\boldsymbol{\Phi}_{t-1}^{(i)}, \sigma_{\mathbf{V}_\Phi}^2 \mathbf{I}\right) \mathbb{I}_{\bar{\boldsymbol{\Phi}}} \\
\mathbf{A}_t^{(i)} &\sim \mathcal{N}\left(\mathbf{A}_{t-1}^{(i)}, \sigma_{\mathbf{V}_\mathbf{A}}^2 \mathbf{I}\right) \mathbb{I}_{\bar{\mathbf{A}}} \\
\overline{\mathbf{s}}_t^{(i)} &\sim \mathcal{N}\left(\mathbf{G}\left(\boldsymbol{\Phi}_t^{(i)}\right)\mathbf{s}_{t-1}^{(i)}, \mathbf{V}_{\mathbf{s},t}\left(\boldsymbol{\upsilon}_t^{(i)}\right)\right) \\
\mathbf{s}_t^{(i)} &\sim \overline{p}\,\mathcal{N}\left(\overline{\mathbf{s}}_t^{(i)}, \sigma_\mathbf{s}^2 \mathbf{I}\right) + \left(1-\overline{p}\right)\mathcal{N}\left(\hat{\mathbf{s}}_{t-1}, \sigma_\mathbf{s}^2 \mathbf{I}\right)
\end{aligned} \tag{6.32}$$

where $0 < \overline{p} < 1$ denotes the probability of the components of a Gaussian mixture process, from which the new particles of sources are sampled. Above, $\mathbb{I}_{\bar{\boldsymbol{\Phi}}}$ and $\mathbb{I}_{\bar{\mathbf{A}}}$ are used to exploit *a priori* information about the corresponding states as shown below:

$$\mathbb{I}_{\bar{\boldsymbol{\Phi}}} = \begin{cases} 1, & \boldsymbol{\Phi} \in \bar{\boldsymbol{\Phi}} \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \mathbb{I}_{\bar{\mathbf{A}}} = \begin{cases} 1, & \mathbf{A} \in \bar{\mathbf{A}} \\ 0, & \text{otherwise} \end{cases} \tag{6.33}$$

where $\bar{\boldsymbol{\Phi}}$ denotes the set of stable AR matrices. Similarly, $\bar{\mathbf{A}}$ denotes the set of mixing matrices where the non-diagonal elements are less than 1. This condition represents *a priori* information that the $l^{th}$ source is dominant at the $l^{th}$ sensor. In the artificial state-transition model, $\hat{\mathbf{s}}_{t-1}$ denotes the MAP estimate of source vector $\mathbf{s}_{t-1}$, which is calculated during the preceding time instant *t-1*, by the following relationship (Van Trees, 1968):

$$\hat{\mathbf{s}}_{t-1} = \left(\hat{\mathbf{A}}_{t-1}^T \hat{\mathbf{A}}_{t-1}\right)^{-1} \hat{\mathbf{A}}_{t-1}^T \mathbf{y}_{t-1} \tag{6.34}$$

In (6.34), $\hat{\mathbf{A}}_{t-1}$ denotes the MMSE estimate of the mixing matrix at time *t-1*.

Above importance function is taken to be the same function which is obtained from the artificial state-transition equations defined in (6.28). Again, the use of Bootstrap particle filtering with the artificial state-transition pdf is proposed here, i.e. $q\left(\mathbf{x}_t \left| \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t} \right.\right) = p\left(\mathbf{x}_t \left| \mathbf{x}_{t-1} \right.\right)$ which is defined as in (6.31) and then elaborated as in (6.32). Due to the utilization of the Bootstrap particle filtering, importance weight of each particle is calculated by means of the likelihood function as shown below (Table 3.3):

$$\begin{aligned} w_t^{(i)} &= p\left(\mathbf{y}_t \left| \mathbf{x}_t^{(i)} \right.\right) \\ p\left(\mathbf{y}_t \left| \mathbf{x}_t^{(i)} \right.\right) &= \mathcal{N}\left(\mathbf{y}_t; \mathbf{A}_t^{(i)} \mathbf{s}_t^{(i)}, \boldsymbol{\Sigma}_{\mathbf{n}}^{(i)}\right) \end{aligned} \tag{6.33}$$

After calculating the importance weight of each particle, these weights are normalized and resampling is performed by the replacement of *N* particles. A pseudo-code of the proposed method is shown below:

Table 6.3. Pseudo-code of MCARS method

For *i =1 to N*,

1. INITIATION:

Draw samples from the initial distributions of the state variables: $\mathbf{x}_t^{(i)} \sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$,

where **m** and **P** denote the mean and covariance matrices of the Gaussian distribution. Note that, $\mathbf{P}_0$ is a diagonal matrix.

For *t = 1 to* $\tau$ , ( $\tau$ denotes the data length)

2. STATE TRANSITIONS:

First, determine constant values of the drift parameters:

$$\sigma^2_{\mathbf{V}_\varsigma},\ \sigma^2_{\mathbf{V}_\chi},\ \sigma^2_{\mathbf{V}_\Phi},\ \sigma^2_{\mathbf{V}_\mathbf{A}},\ \sigma^2_{\mathbf{s}}$$

Secondly, draw new particles for each state variable by using the proposed hierarchical importance function as shown below:

Table 6.3. (continued)

$$p\left(\mathbf{x}_t\middle|\mathbf{x}_{t-1}\right)=p\left(\mathbf{s}_t\middle|\mathbf{s}_{t-1},\boldsymbol{\Phi}_t,\boldsymbol{\upsilon}_t\right)p\left(\boldsymbol{\Phi}_t\middle|\boldsymbol{\Phi}_{t-1},\boldsymbol{\upsilon}_t\right)p\left(\mathbf{A}_t\middle|\mathbf{A}_{t-1},\mathbf{n}_t\right)$$
$$\text{x}\ p\left(\boldsymbol{\upsilon}_t\middle|\boldsymbol{\upsilon}_{t-1},\boldsymbol{\chi}_t\right)p\left(\mathbf{n}_t\middle|\mathbf{n}_{t-1},\boldsymbol{\varsigma}_t\right)$$
$$\text{x}\ p\left(\boldsymbol{\varsigma}_t\middle|\boldsymbol{\varsigma}_{t-1}\right)p\left(\boldsymbol{\chi}_t\middle|\boldsymbol{\chi}_{t-1}\right)$$

where sample drawing from these conditional distributions are given in the hierarchical order as demonstrated below:

$$\boldsymbol{\varsigma}_t^{(i)}\sim\mathcal{N}\left(\boldsymbol{\varsigma}_{t-1}^{(i)},\sigma_{\mathbf{V}_\varsigma}^2\mathbf{I}\right)\rightarrow\ \text{Form}\ \boldsymbol{\Sigma}_{\mathbf{n}}^{(i)}\left(\boldsymbol{\varsigma}_t^{(i)}\right)$$
$$\boldsymbol{\chi}_t^{(i)}\sim\mathcal{N}\left(\boldsymbol{\chi}_{t-1}^{(i)},\sigma_{\mathbf{V}_\chi}^2\mathbf{I}\right)\rightarrow\ \text{Form}\ \boldsymbol{\Sigma}_{\boldsymbol{\upsilon}}^{(i)}\left(\boldsymbol{\chi}_t^{(i)}\right)$$
$$\boldsymbol{\upsilon}_t^{(i)}\sim\mathcal{N}\left(\boldsymbol{\upsilon}_{t-1}^{(i)},\boldsymbol{\Sigma}_{\boldsymbol{\upsilon}}^{(i)}\left(\boldsymbol{\chi}_t^{(i)}\right)\right)$$
$$\mathbf{n}_t^{(i)}\sim\mathcal{N}\left(\mathbf{n}_{t-1}^{(i)},\boldsymbol{\Sigma}_{\mathbf{n}}^{(i)}\left(\boldsymbol{\varsigma}_t^{(i)}\right)\right)$$
$$\boldsymbol{\Phi}_t^{(i)}\sim\mathcal{N}\left(\boldsymbol{\Phi}_{t-1}^{(i)},\sigma_{\mathbf{V}_\Phi}^2\mathbf{I}\right)\mathbb{I}_{\bar{\boldsymbol{\Phi}}}$$
$$\mathbf{A}_t^{(i)}\sim\mathcal{N}\left(\mathbf{A}_{t-1}^{(i)},\sigma_{\mathbf{V}_\mathbf{A}}^2\mathbf{I}\right)\mathbb{I}_{\bar{\mathbf{A}}}$$
$$\overline{\mathbf{s}}_t^{(i)}\sim\mathcal{N}\left(\mathbf{G}\left(\boldsymbol{\Phi}_t^{(i)}\right)\mathbf{s}_{t-1}^{(i)},\mathbf{V}_{\mathbf{s},t}\left(\boldsymbol{\upsilon}_t^{(i)}\right)\right)$$
$$\mathbf{s}_t^{(i)}\sim\overline{p}\,\mathcal{N}\left(\overline{\mathbf{s}}_t^{(i)},\sigma_\mathbf{s}^2\mathbf{I}\right)+\left(1-\overline{p}\right)\mathcal{N}\left(\hat{\mathbf{s}}_{t-1},\sigma_\mathbf{s}^2\mathbf{I}\right)$$

where the MAP estimate of the source vector at time $(t-1)$ is utilized:

$$\hat{\mathbf{s}}_{t-1}=\left(\hat{\mathbf{A}}_{t-1}^T\hat{\mathbf{A}}_{t-1}\right)^{-1}\hat{\mathbf{A}}_{t-1}^T\mathbf{y}_{t-1}$$

3. CALCULATE THE IMPORTANCE WEIGHT OF EACH PARTICLE:

$$w_t^{(i)}=p\left(\mathbf{y}_t\middle|\mathbf{x}_t^{(i)}\right)$$
$$p\left(\mathbf{y}_t\middle|\mathbf{x}_t^{(i)}\right)=\mathcal{N}\left(\mathbf{y}_t;\mathbf{A}_t^{(i)}\mathbf{s}_t^{(i)},\boldsymbol{\Sigma}_\mathbf{n}^{(i)}\right)$$

4. NORMALIZE THE WEIGHTS:

$$\tilde{w}_t^i=\frac{w_t^i}{\sum_{i=1}^N w_t^i}$$

5. RESAMPLE

6. GO TO STEP 2.

## 6.4. Experiments

In this chapter, successful performances of the proposed methods are illustrated by computer simulations. First, results of MCS method are demonstrated which are followed by those of MCARS technique.

### 6.4.1. Performance Analysis of MCS method:

Here, modeling of a non-stationary mixture of two cross-correlated, temporally independent Gaussian processes is performed by using the MCS technique. Two sources are distributed as follows:

$$\mathbf{s}_t = \begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix} \right) \tag{6.35}$$

where $\boldsymbol{\mu}_{\mathbf{s}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and $\boldsymbol{\Sigma}_{\mathbf{s}} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$, corresponding to (6.6.b). It is seen that these two sources are cross-correlated, however they do not possess any time structure. In order to use MCS, we assume that these statistics are known *a priori*. In this problem, the mixture of these sources is simulated as follows:

$$\mathbf{y}_t = \mathbf{A}_t \mathbf{s}_t + \mathbf{n}_t \tag{6.36}$$

where the mixing matrix and the observation noise statistics are shown below:

$$\mathbf{A}_t = \begin{bmatrix} 1 & cos\left(\dfrac{2\pi t}{T}\right) \\ sin\left(\dfrac{2\pi t}{T}\right) & 1 \end{bmatrix} \text{ and } \mathbf{n}_t \sim \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix} \right) \tag{6.37}$$

where $T = 128$ is chosen and $\tau = 400$ data (400 measurements of vector $\mathbf{y}_t$) are generated. Assuming that we know the source and observation noise statistics, MCS is applied to estimate the mixing matrix and the sources. First, mixing matrix is estimated by the

particle filter by integrating out the sources. Here, $\tilde{\mathbf{A}}_0^{(i)} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}2 & 0\\0 & 2\end{bmatrix}\right)$ and $\tilde{\mathbf{A}}_t \sim \mathcal{N}\left(\tilde{\mathbf{A}}_{t-1}^{(i)}, \begin{bmatrix}0.05 & 0\\0 & 0.05\end{bmatrix}\right)$ values are used in the particle filtering part of MCS (Table 6.2). 100 particles are used. For the sake of simplicity, hyperparameters $(\boldsymbol{\mu}_1, \mathbf{P}_1)$ and $(\boldsymbol{\mathcal{V}}, d_1, \bar{\eta})$ are not used here, since we assume that we know the values of $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$ (Since they are assumed to be known here, there is no need to assign a probabilistic model for them). Then, the estimate of the mixing matrix, found by the particle filter, is substituted into matrix $\mathcal{C}_0$, constituting the prior of the mixing matrix needed in the MCMC part for each data block assumed to be stationary. For the sake of simplicity, prior selections in (6.13.c) and (6.13.d) are omitted, since we assume to know the values of these matrices. Therefore, during the MCMC part of the algorithm, (6.14.c) and (6.14.e) are also omitted due to the same reason. $L = 800$ Gibbs sampling iterations are used to calculate the point estimates of $\boldsymbol{\mathcal{S}}, \boldsymbol{\Sigma}_s, \boldsymbol{\mathcal{C}}$ and $\boldsymbol{\Sigma}_n$ after omitting their values within the first 200 iterations. This transient is taken to be the burn-in period of the Gibbs Sampler. 40 sub-blocks are chosen. $\mathbf{H} = \begin{bmatrix}0.005 & 0\\0 & 0.005\end{bmatrix}$ is chosen, since the sources are i.i.d. in time (Small values of $\mathbf{H}$ implies small correlation in time (Rowe, 2003)). $\mathbf{S_0} = \mathbf{I}_{(2x10)}$ is chosen for each data block for an arbitrary initial point. 100 realizations have been performed and the following Normalized Mean Square Error criterion is used to evaluate the estimation performances:

$$NMSE(t) = \frac{\sum_{i=1}^{100}\left(\hat{x}_{j,i}(t) - x_{j,i}(t)\right)^2}{\sum_{i=1}^{100}\sum_{t=1}^{\tau} x_{j,i}^2(t)}, \qquad j = 1,2 \tag{6.38}$$

In the following figures, observation signals, source waveforms, and their MMSE estimates are illustrated for one realization in the ensemble in Figures 6.3, 6.4 and 6.6, respectively. By observing 100 realizations of this ensemble, the mean of the estimated mixing matrix elements, their NMSE curves and the NMSE curves of the estimated mixing matrix elements are illustrated in Figures 6.5, 6.7 and 6.8, respectively.

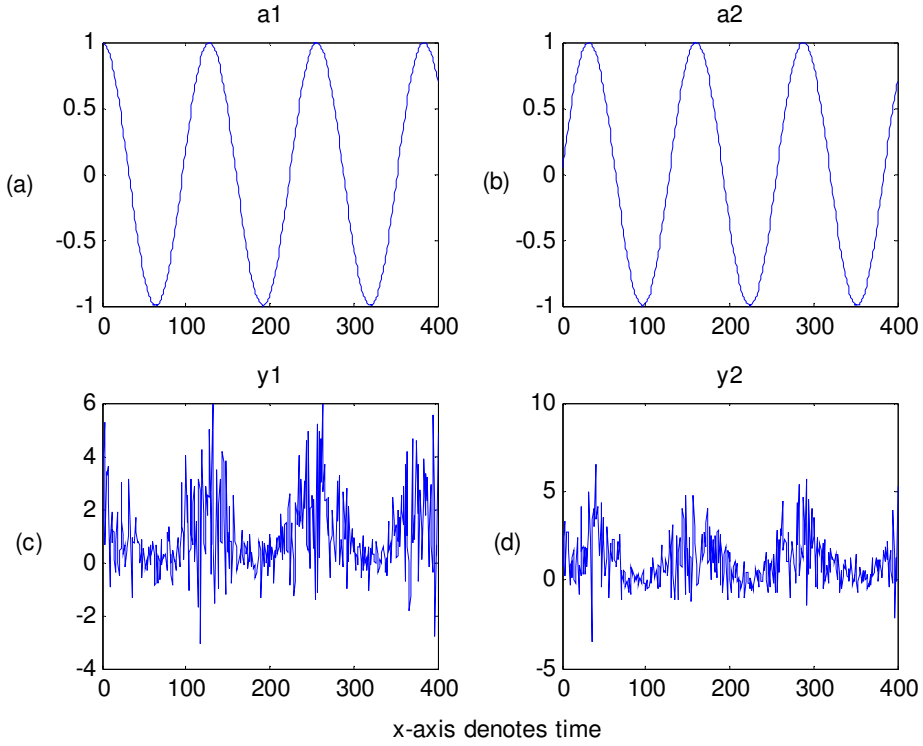


Figure 6.3. MCS Experiment: a) Waveform of $a_{12}(t)$, a) Waveform of $a_{21}(t)$, c) Observation 1, d) Observation 2

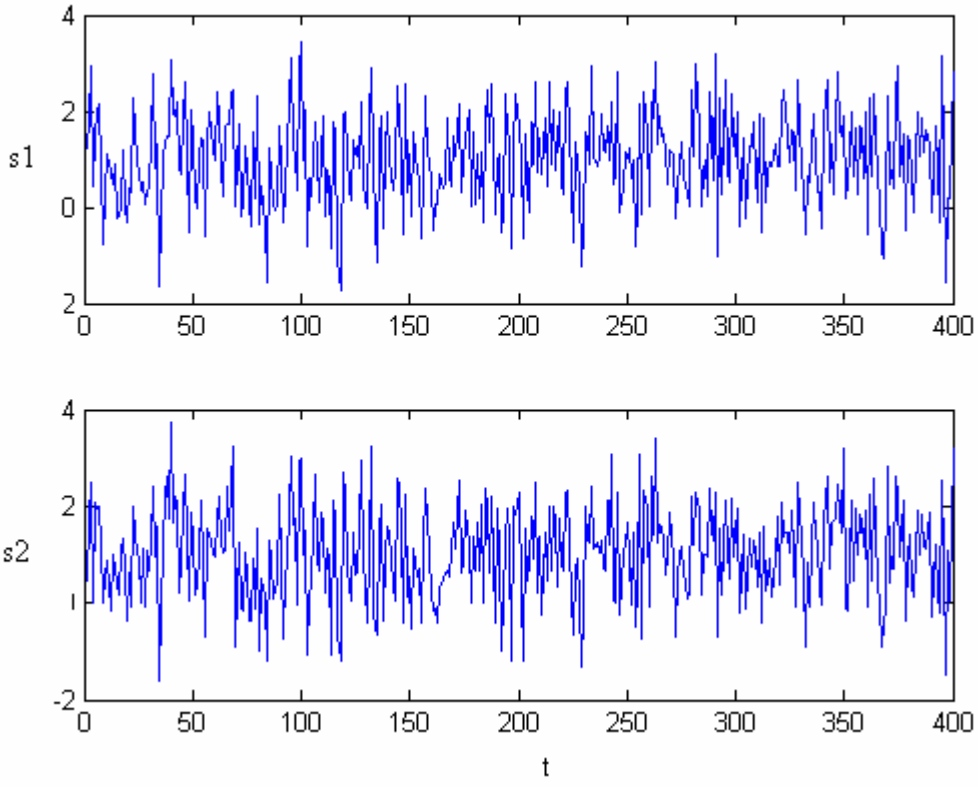


Figure 6.4. MCS Experiment: Source waveforms

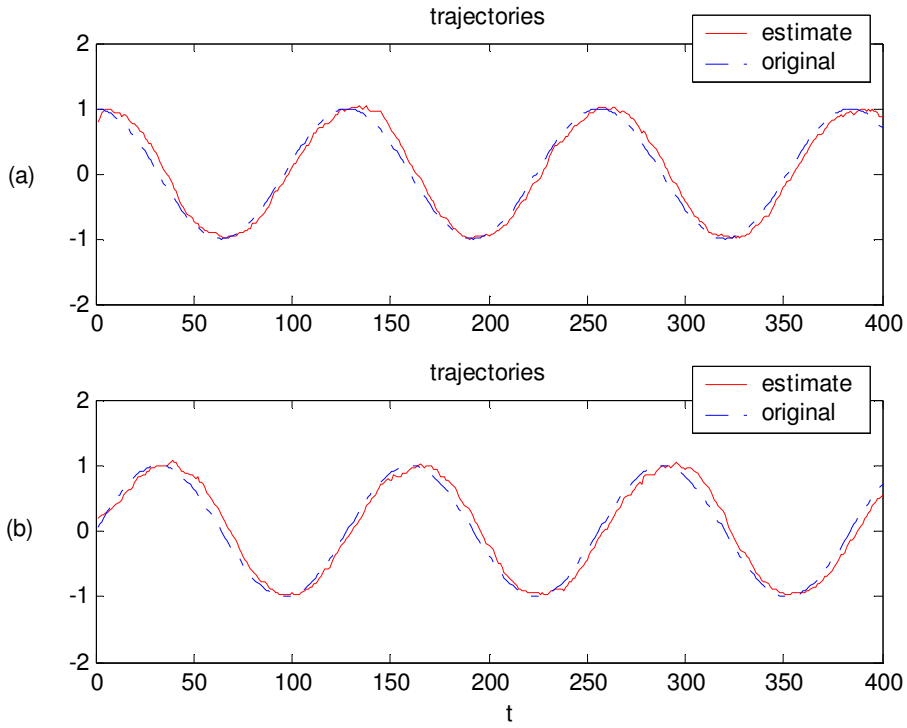


Figure 6.5. MCS Experiment: Mixing matrix elements and their MMSE estimates, a) $a_{12}(t)$, b) $a_{21}(t)$

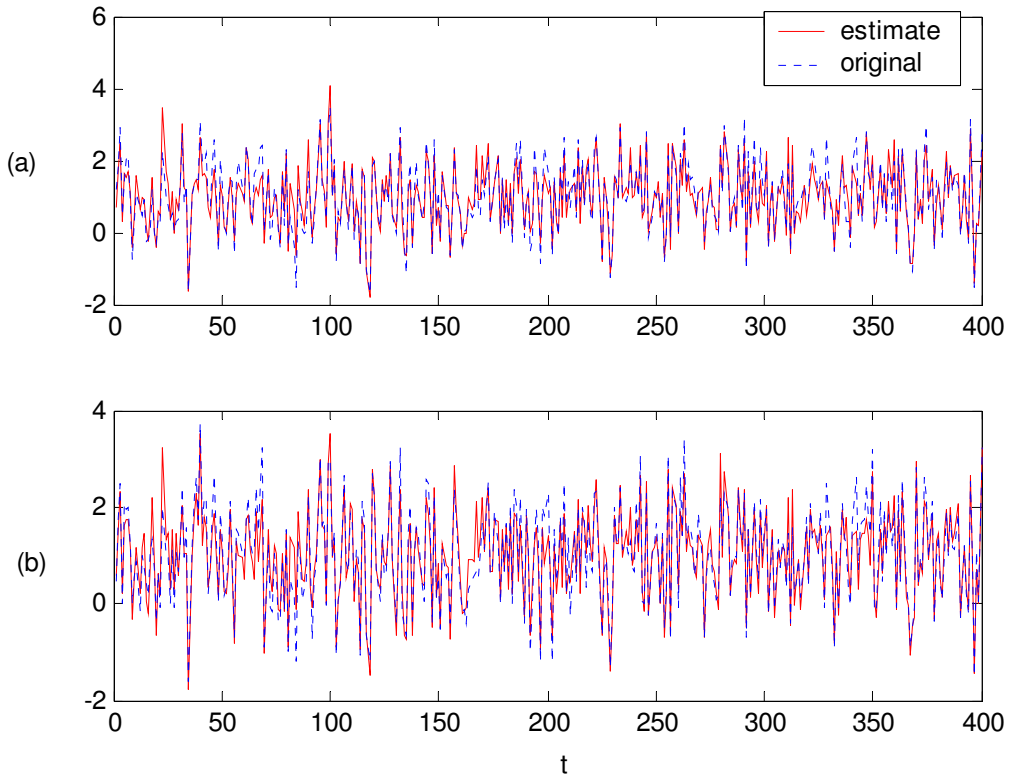


Figure 6.6. MCS Experiment: Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

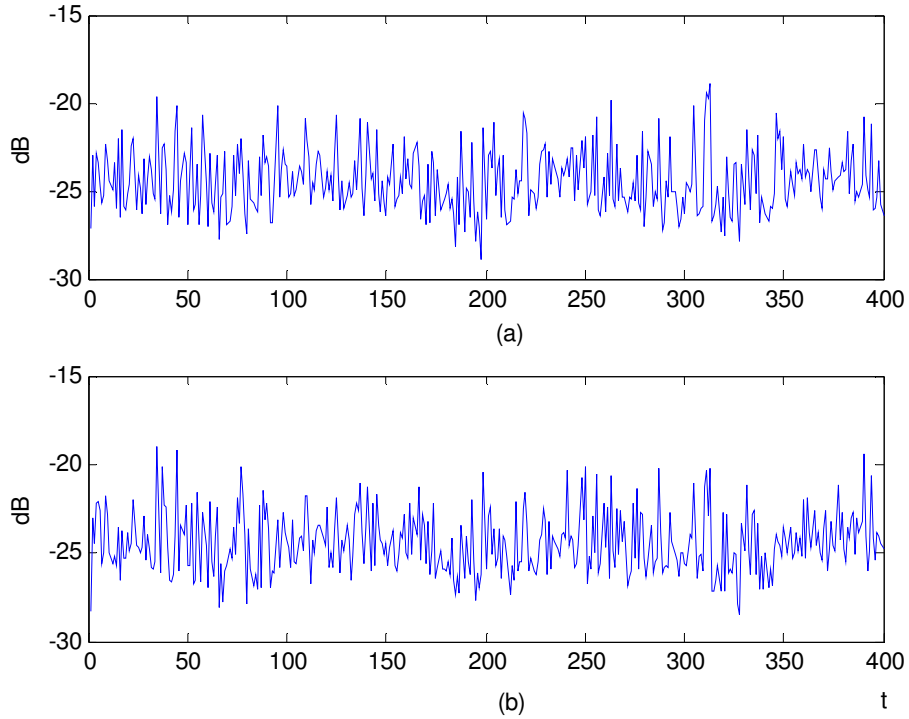


Figure 6.7. MCS Experiment: NMSE curves of sources estimates, a) $s_1(t)$, b) $s_2(t)$

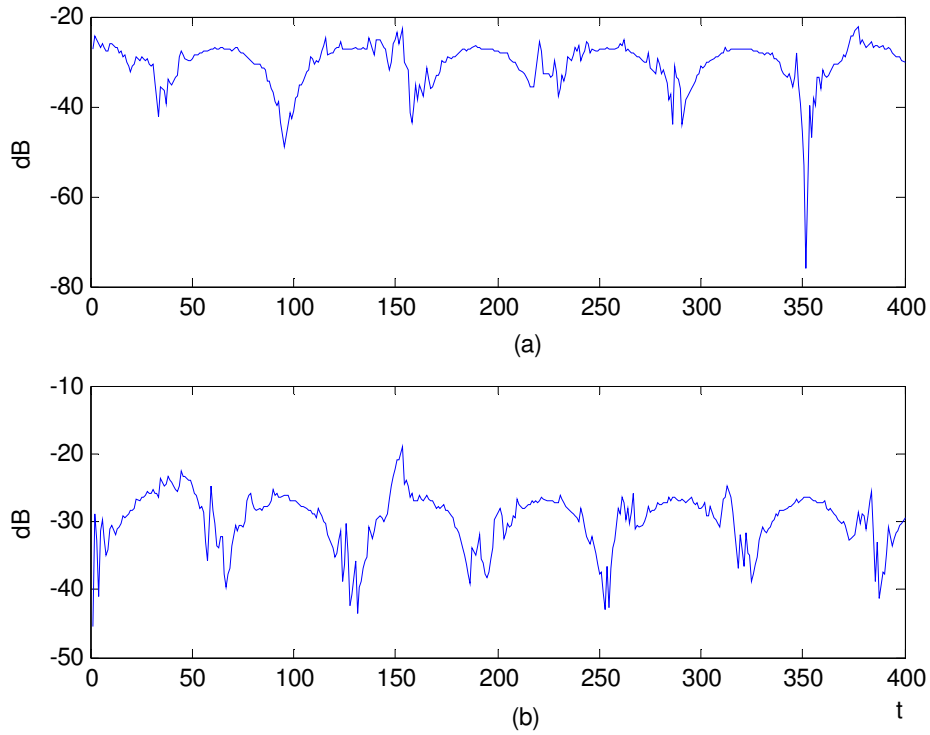


Figure 6.8. MCS Experiment: NMSE curves of mixing matrix elements,

a) $a_{12}(t)$, b) $a_{21}(t)$

The above results verify the successful performance of our method. Both sources and the mixing matrix elements have been estimated successfully for cross-correlated sources. Mixtures obtained by the multiplication of the sources with sinusoidal mixing matrices can be thought as basic modulation operations used in telecommunications.

### 6.4.2. Performance Analysis of MCARS method:

Here, modeling of a non-stationary mixture of two cross-correlated AR Gaussian processes is performed by using the MCARS technique. The experimental scenario is illustrated as follows:

$$\begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{bmatrix} \begin{bmatrix} s_{1,t-1} \\ s_{2,t-1} \end{bmatrix} + \begin{bmatrix} \upsilon_{1,t} \\ \upsilon_{2,t} \end{bmatrix} \tag{6.40}$$

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix} \begin{bmatrix} s_{1,t} \\ s_{2,t} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix} \tag{6.41}$$

where the AR matrix and the mixing matrix are represented as follows:

$$\mathbf{\Phi}_{1,t} = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{bmatrix} \text{ and } \mathbf{A}_{t} = \begin{bmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{bmatrix} \tag{6.42}$$

As mentioned earlier, a successful modeling cannot be provided if both the AR coefficient matrix and the mixing matrix are time-varying. A similar conclusion has also been drawn by (Everson and Roberts, 2000) for ICA. Therefore, in order to show the performance of our method, we will consider two main groups of simulations that can be encountered in the physical world:

- Time-varying AR matrix and time-invariant mixing matrix
- Time-invariant AR matrix and time-varying mixing matrix

Moreover, cross-correlation between the sources can arise either due to the non-diagonal entries of the AR matrix, or due to the correlated driving process. Therefore, our simulations will also consider these different cases as demonstrated below:

**Experiment 1 (E1):** In this experiment, the scenario given by (6.40) and (6.41) is considered. In this experiment, non-stationarity arises due to the time-varying nature of the mixing matrix as shown below:

$$\mathbf{A}_t = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 2 \end{bmatrix} \text{ for } t < 1000, \ \mathbf{A}_t = \begin{bmatrix} 2 & -1.5 \\ -1.5 & 2 \end{bmatrix} \text{ for } t > 1000 \tag{6.43}$$

while the cross-correlation between the sources is incorporated through the non-diagonal covariance matrix of the zero-mean driving process:

$$\boldsymbol{\upsilon}_t = \begin{bmatrix} \upsilon_{1,t} \\ \upsilon_{2,t} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.01 & 0.01/2 \\ 0.01/2 & 0.01 \end{bmatrix} \right) \tag{6.44}$$

In this experiment, AR matrix is taken to be diagonal in the following form:

$$\boldsymbol{\Phi}_{1,t} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.9 \end{bmatrix} \tag{6.45}$$

In all experiments from now on, the observation noise process is taken to be distributed as follows:

$$\mathbf{n}_t = \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix} = \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 10^{-5} & 0 \\ 0 & 10^{-5} \end{bmatrix} \right) \tag{6.46}$$

This modeling can simulate problems where separation of cross-correlated signals is needed when the mixing system changes abruptly in time. Similar situations may arise in geophysics (earth layer modeling), teleconference applications with changing channels due to the movement or wireless communications with similar varying channel characteristics.

The results of this experiment are shown next: First, the observation and source signals are illustrated:

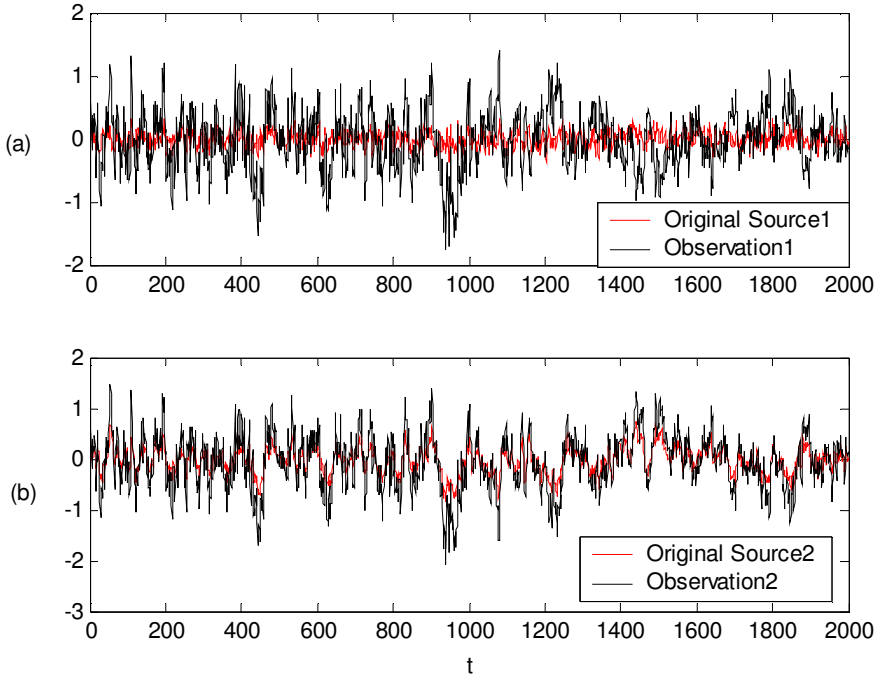


Figure 6.9. E1: Observation and source signals of 1 realization

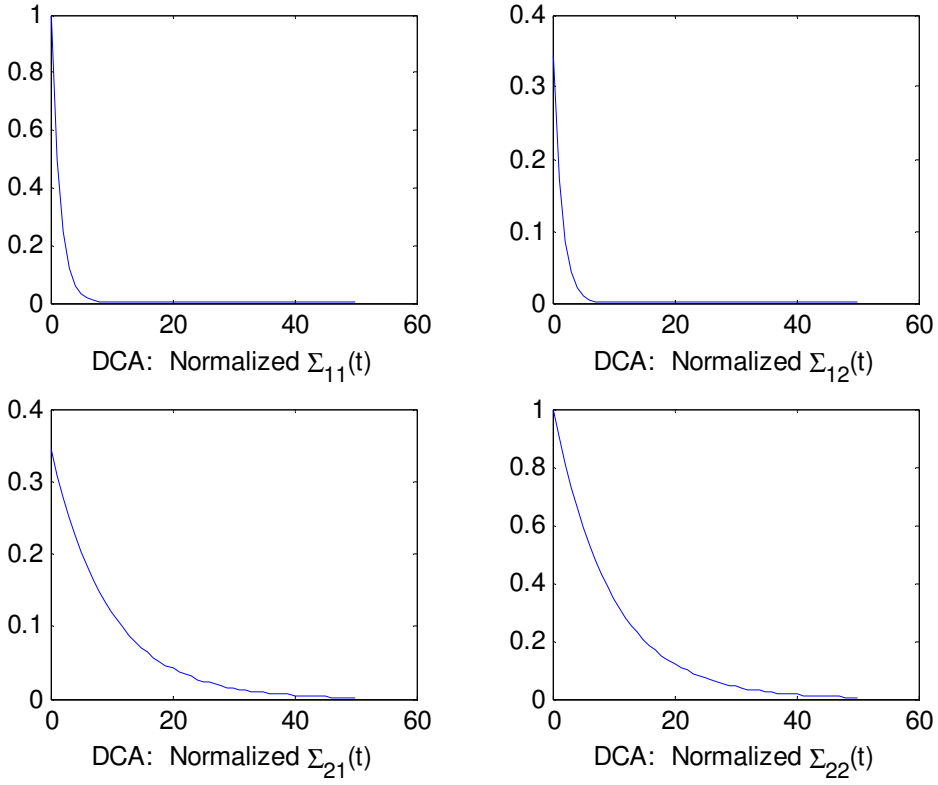


Figure 6.10. E1: Normalized covariance matrix of VAR(1) source vector

Next, the mean of the mixing matrix estimate is shown over 100 realizations. Then, the inverse of this estimate is used to extract the sources. We may also use the source estimate in an arbitrary realization in the ensemble without waiting for the smoothed estimate (ensemble average) of the mixing matrix. This situation is illustrated afterwards.

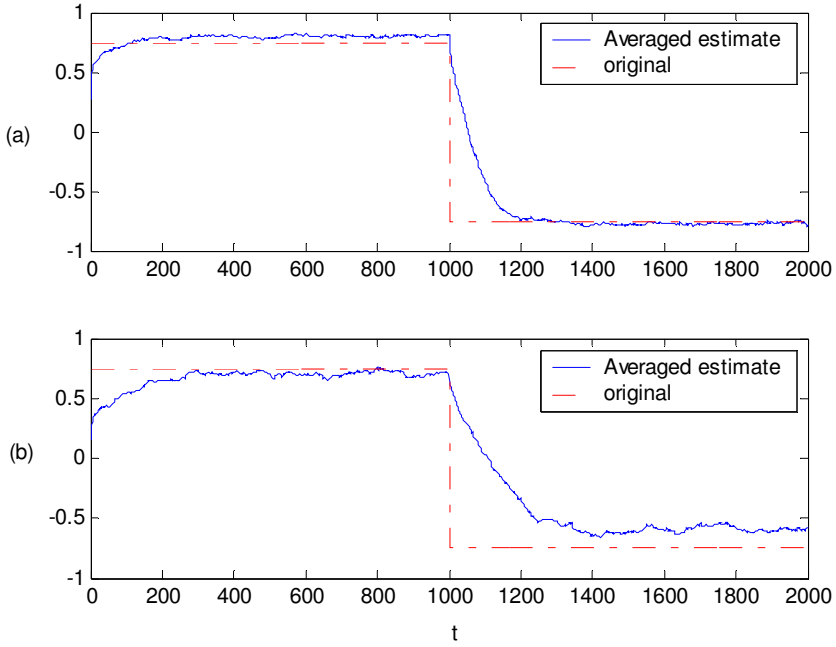


Figure 6.11. E1: Non-diagonal mixing matrix elements and their MMSE estimates

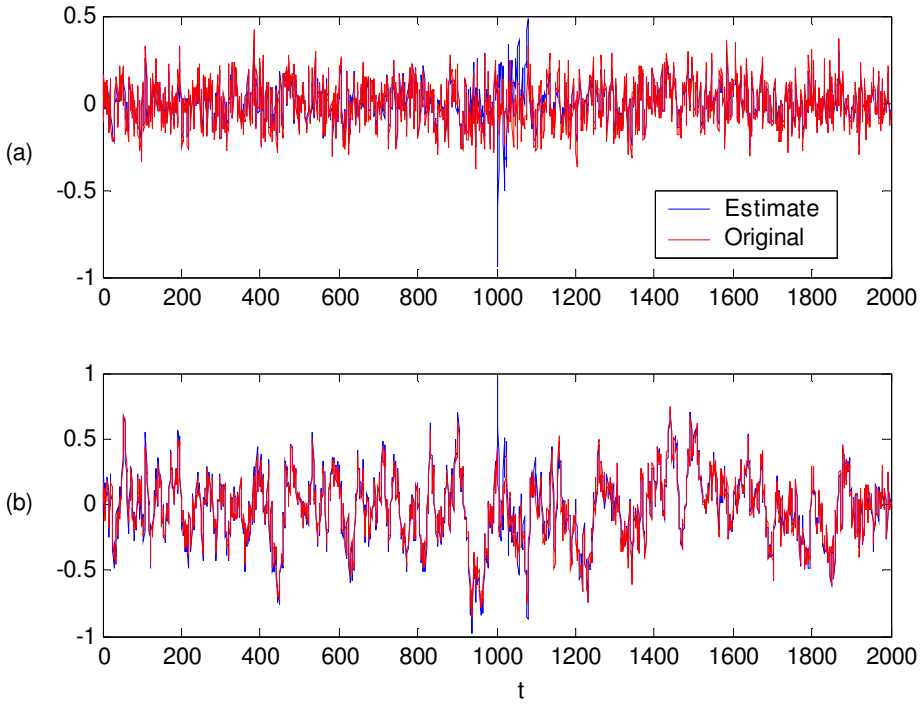


Figure 6.12. E1: Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

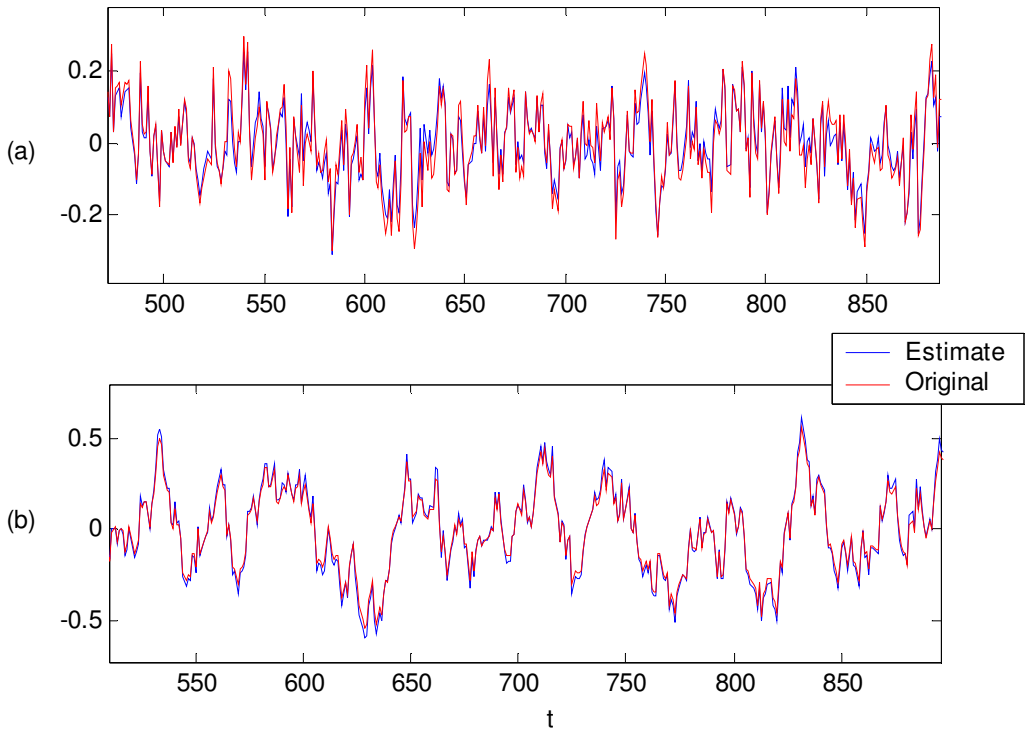


Figure 6.13. E1: Zoomed Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

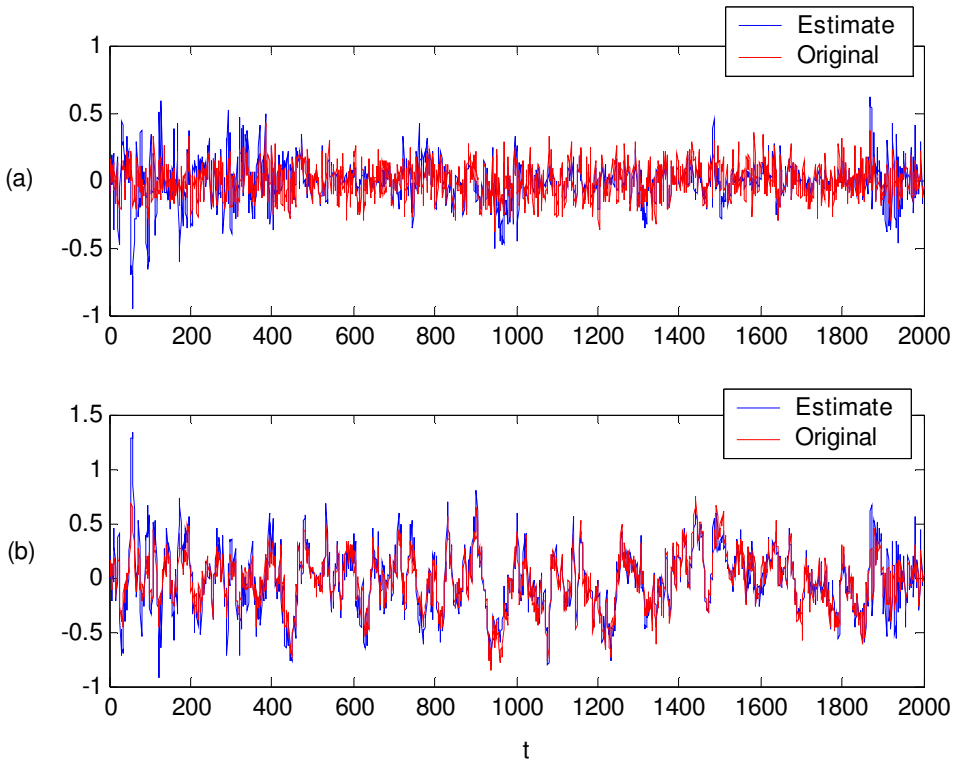


Figure 6.14. E1: Arbitrary realization: Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

The MMSE estimates of the sources given in Figures 6.12 and 6.13 have been found by multiplying the observation data by the inverse of the smoothed estimated mixing matrix $\hat{\mathbf{A}}_t$ (shown in Figure 6.11) at each time instant as shown below:

$$\hat{\mathbf{s}}_t = \hat{\mathbf{A}}_t^{-1}\mathbf{y}_t \tag{6.47}$$

On the other hand, a MMSE estimate of the source vector estimated from an arbitrary realization is also shown afterwards (smoothed mixing matrix is not used here):

When Figures 6.12 and 6.14 are compared, it is observed that the previous one is more accurate within the steady-state regions, since inverse of the smoothed mixing is utilized here. However, during a single realization among the ensemble, we can obtain source estimates with better tracking capabilities. This is observed during the abrupt change of the mixing matrix at $t = 1000$. This change has a larger artifact in Figure 6.12.

Next the NMSE curves of the mixing matrix and source estimates will be presented:

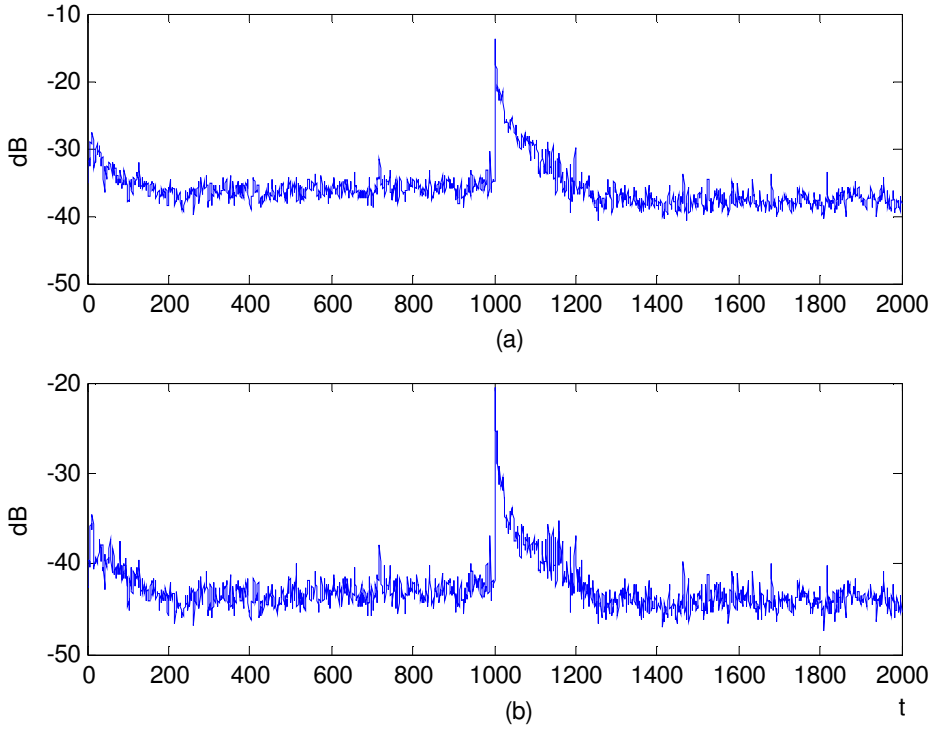


Figure 6.15. E1: NMSE curves of sources, a) First source, b) Second source

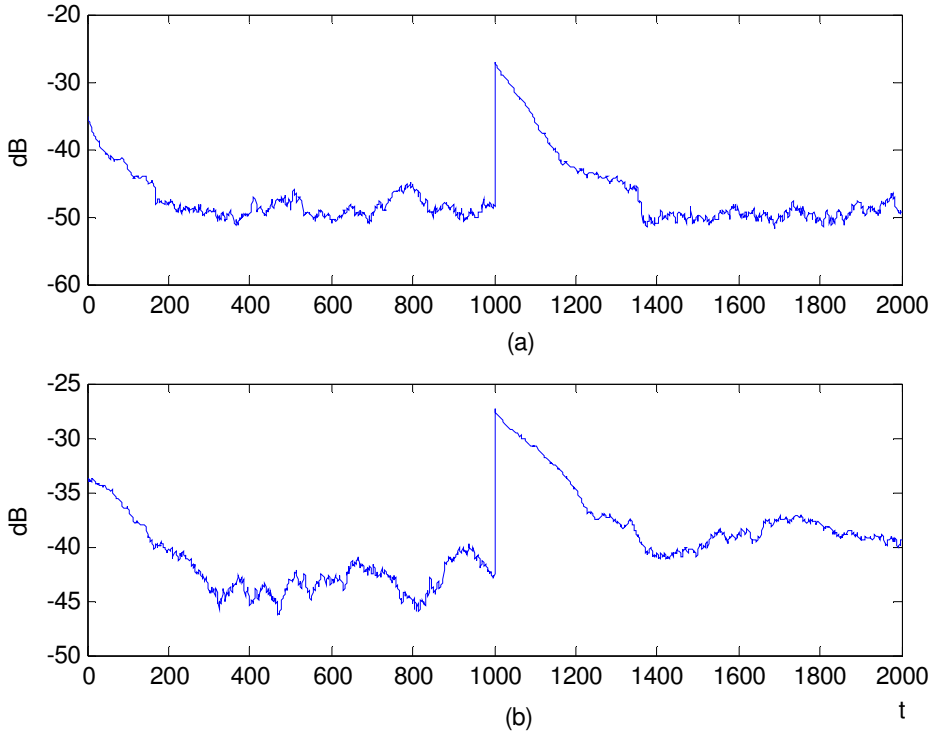


Figure 6.16. E1: NMSE curves of mixing matrix estimates, a) $a_{12}(t)$, b) $a_{21}(t)$

From these results, it is observed that both the estimation and the tracking performances of our method are successful.

**Experiment 2 (E2):** In the preceding experiment, cross-correlation between the sources has been simulated by using the covariance of the driving process, as shown in (6.44). Here, to bring solution for more general cases, correlations that may arise due to the AR coefficient matrix in (6.41) and (6.45) is considered, too. Thus, in addition to (6.44), cross-correlations for such cases are modeled here, by using the following matrix:

$$\mathbf{\Phi}_{1,t} = \begin{bmatrix} 0.5 & -0.1 \\ -0.1 & 0.9 \end{bmatrix} \tag{6.48}$$

which satisfies the stationarity condition of a VAR(1) process as described in Chapter 2. Again, non-stationarity of the observed mixture, $\mathbf{y}_t$, is due to the time-varying structure of the mixing matrix. Here, the same mixing matrix in (6.43) is used. Below, observation and source processes are illustrated, which are followed by the covariance matrix of sources.

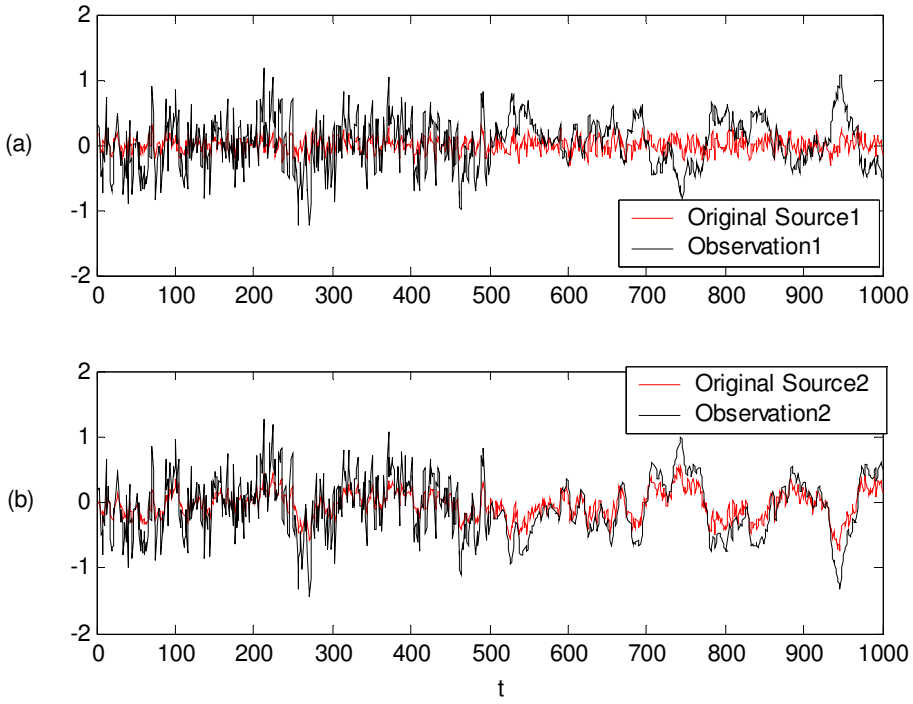


Figure 6.17. E2: Observation and source signals of 1 realization

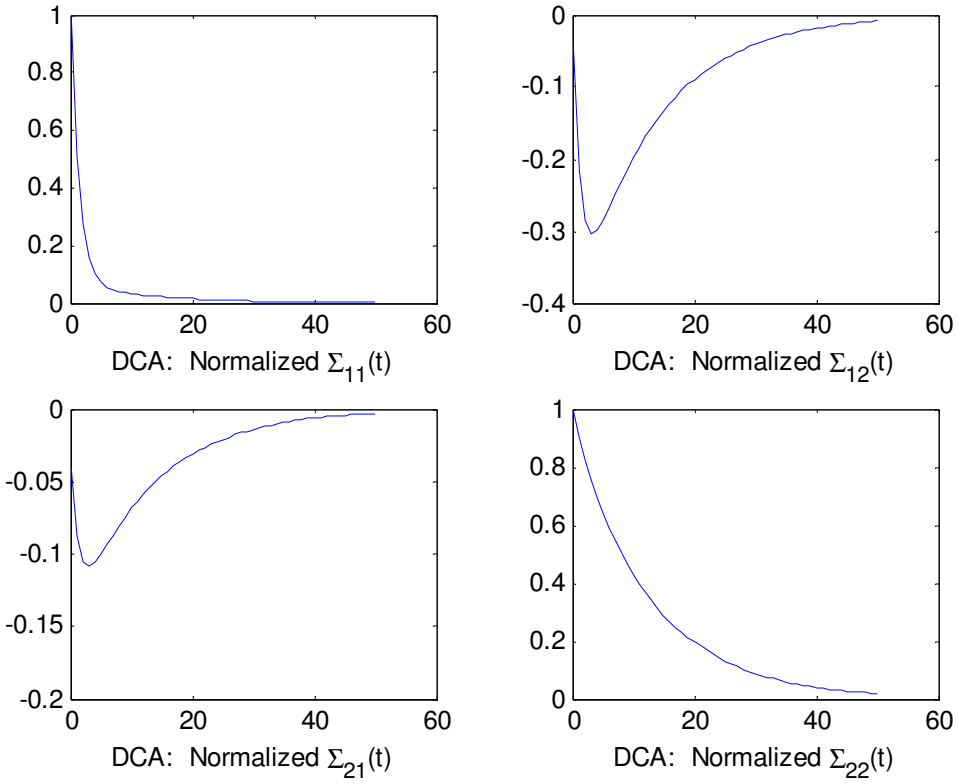


Figure 6.18. E2: Normalized covariance matrix of VAR(1) source vector

Next, the ensemble mean of the mixing matrix estimate is illustrated:

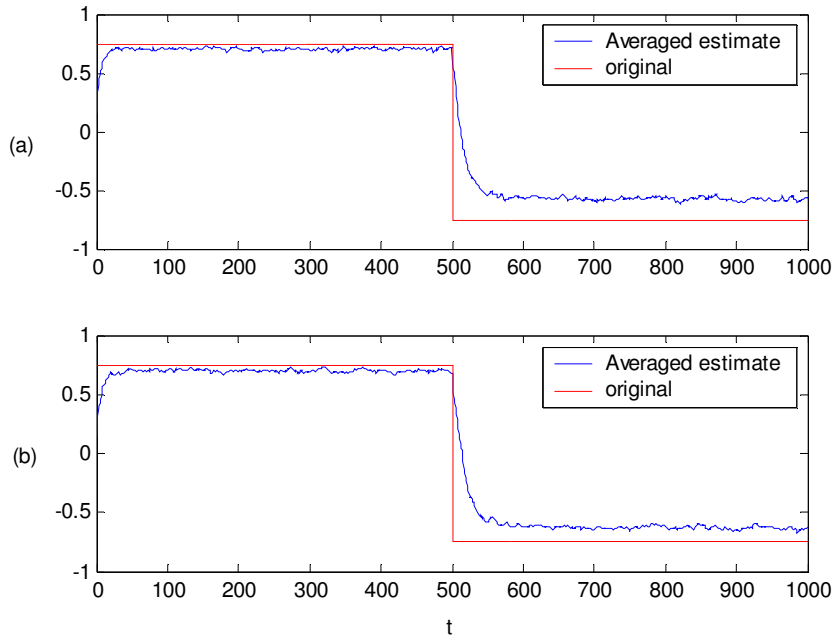


Figure 6.19. E2: Mixing matrix elements and their MMSE estimates, a) $a_{12}(t)$, b) $a_{21}(t)$

Next, sources are illustrated which are estimated by using the mean estimate of the mixing matrix elements found above by using (6.47). This is followed by their zoomed waveforms.

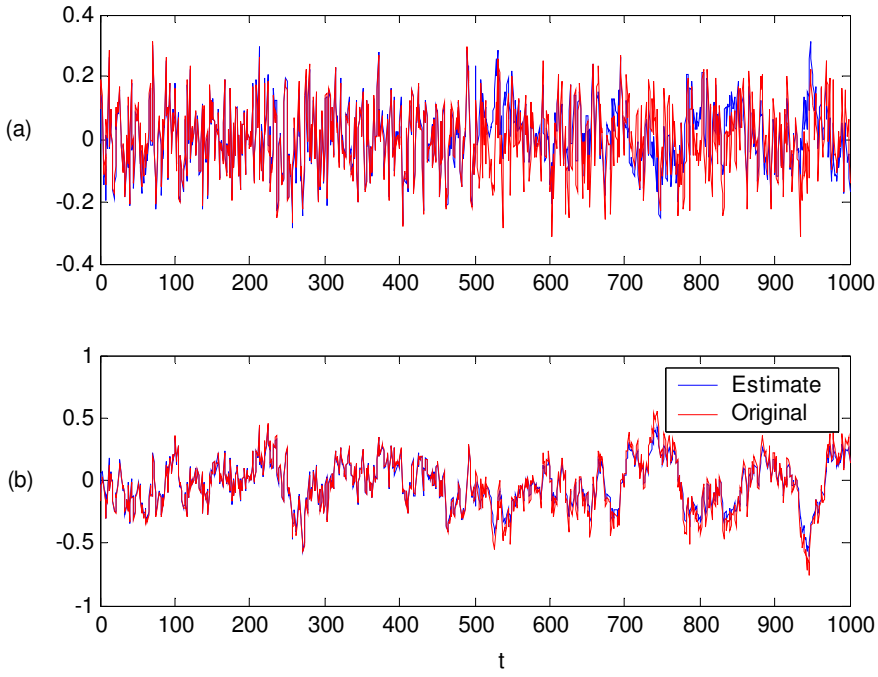


Figure 6.20. E2: Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

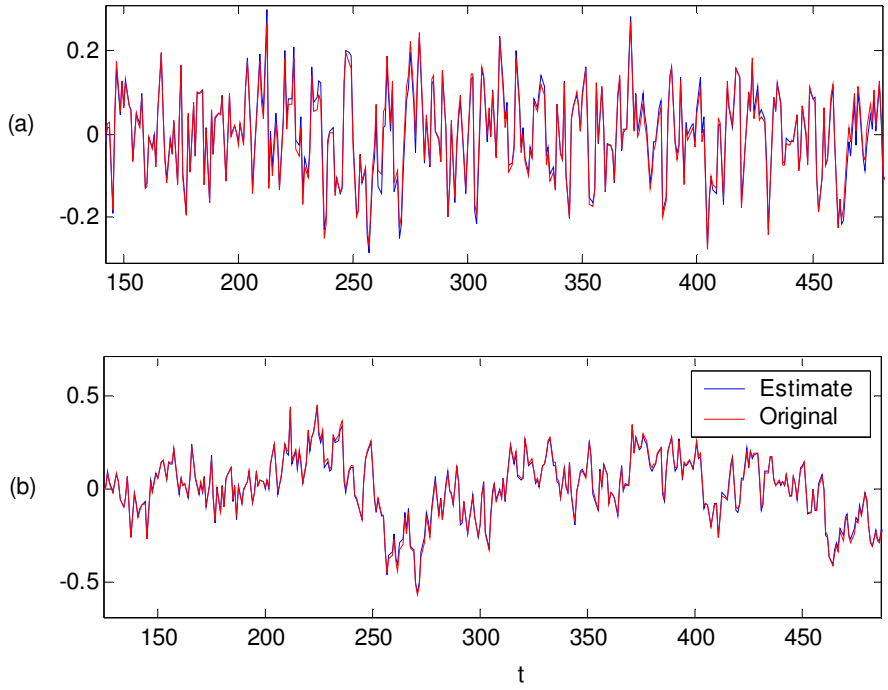


Figure 6.21. E2: Zoomed version of Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$ (before jump)

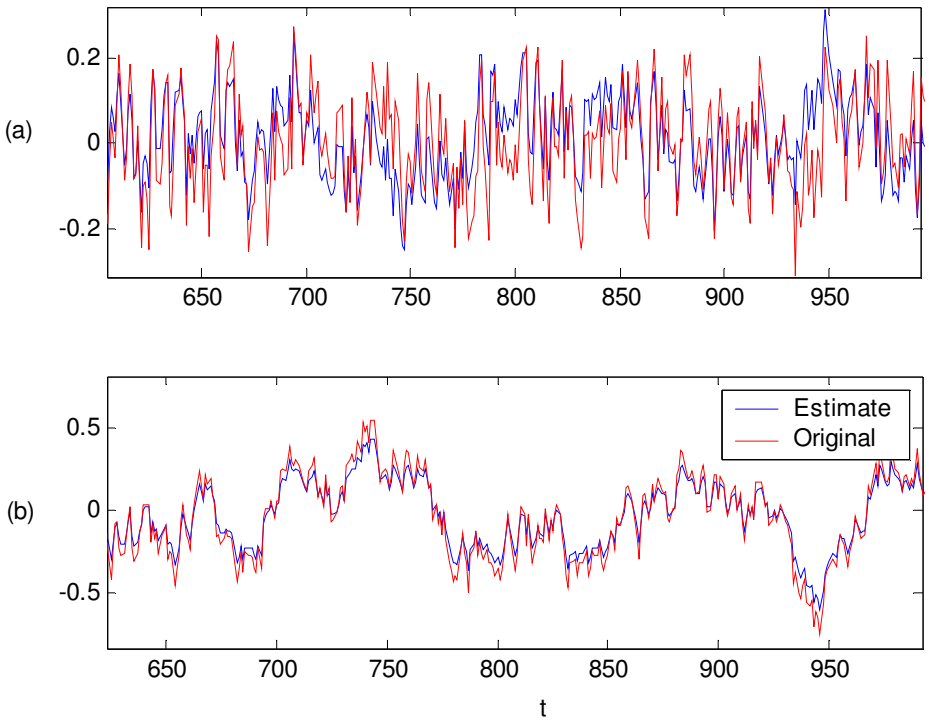


Figure 6.22. E2: Zoomed version of Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$ (after jump)

Sources can be obtained as follows for an arbitrary realization, instead of using the smoothed estimate of the mixing matrix:

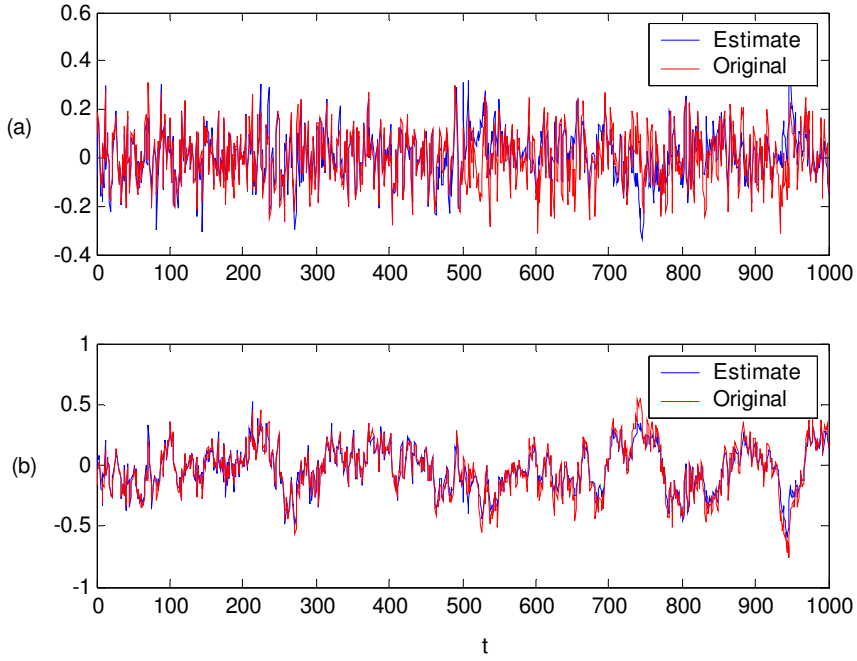


Figure 6.23. E2: Arbitrary realization: Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

Next, the NMSE curves of each parameter are presented for performance analysis.

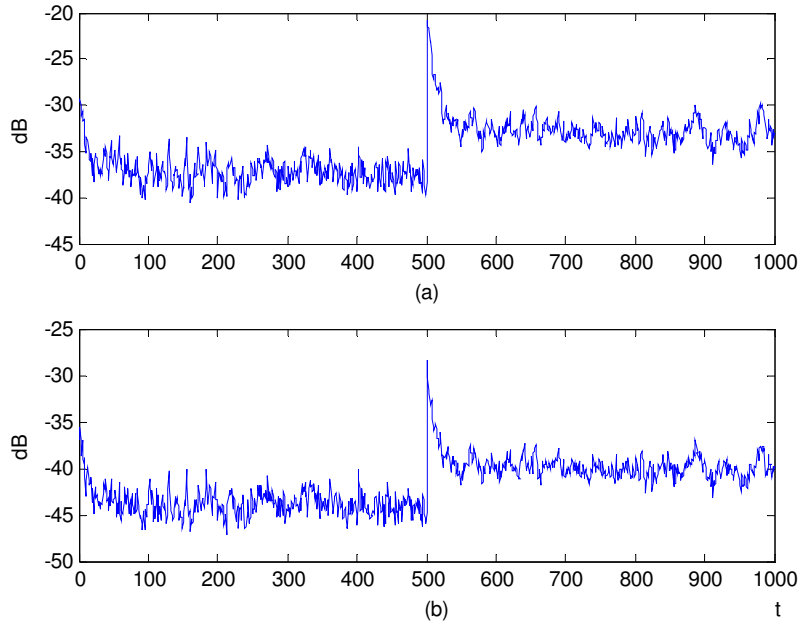


Figure 6.24. E2: NMSE curves of sources, a) First source, b) Second source

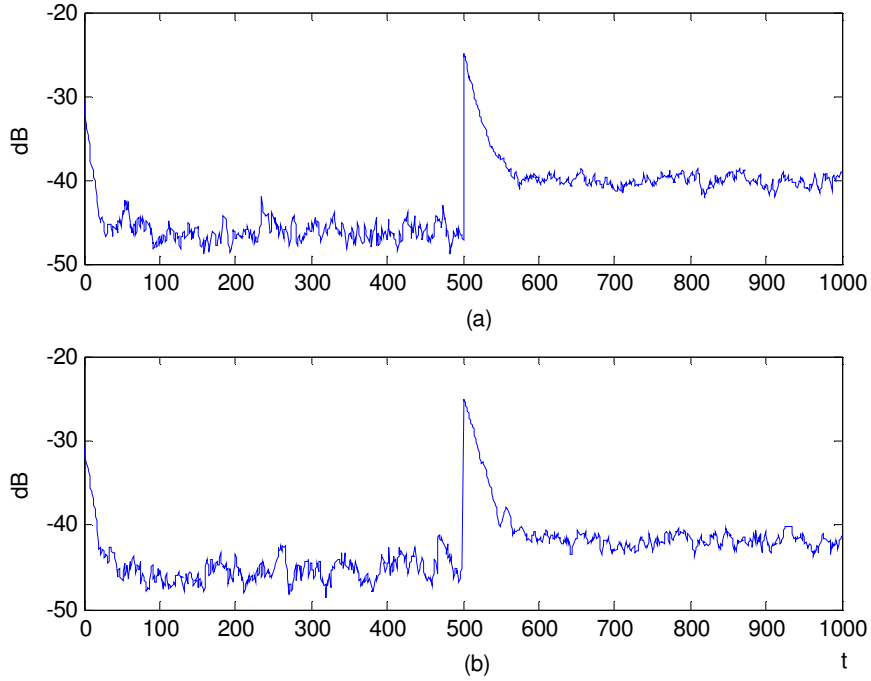


Figure 6.25. E2: NMSE curves of mixing matrix estimates, a) $a_{12}(t)$, b) $a_{21}(t)$

It is observed that there are no significant performance differences between Figures 6.20 and 6.23. This is due to the small error in the second half of the mixing matrix estimate shown in Figure 6.19. Otherwise, provided that the mean estimate (smoothed) of the mixing matrix can be obtained with a negligible estimation error, then using the smoothed estimate gives superior results when compared to those of a single realization of the particle filtering algorithm (Similar to better results of Figure 6.12 when compared with Figure 6.14). The small error in the second half of Figure 6.19 arises due to the long dependence of the covariance matrix elements of the sources in time. Since we use information from the past observation via MAP estimate of the sources, these may lead to small errors when the mixing matrix changes abruptly at $t = 500$.

**Experiment 3 (E3).** In this section, we will simulate problems where the AR coefficient matrix is time-varying, whereas the mixing matrix is constant. Therefore, the AR matrix changes in time as shown below:

$$\mathbf{\Phi}_{1,t} = \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.5 \end{bmatrix} \text{ for } t < 500, \ \mathbf{\Phi}_{1,t} = \begin{bmatrix} -0.5 & -0.1 \\ -0.4 & -0.5 \end{bmatrix} \text{ for } t > 500 \qquad (6.49)$$

where $\tau = 1000$ data are used. Meanwhile, the mixing matrix is kept constant as follows:

$$\mathbf{A}_t = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 2 \end{bmatrix} \tag{6.50}$$

On the other hand, the cross-correlation between the sources is provided through the AR coefficient matrix, here. The components of the driving noise process are chosen to be uncorrelated with the following covariance matrix:

$$\mathbf{v}_t = \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \right) \tag{6.51}$$

In the meantime, the observation noise is distributed as in (6.46), again. Below, the observation and source processes are illustrated which is followed by their zoomed waveforms in order to show the difference between the source and observation processes:

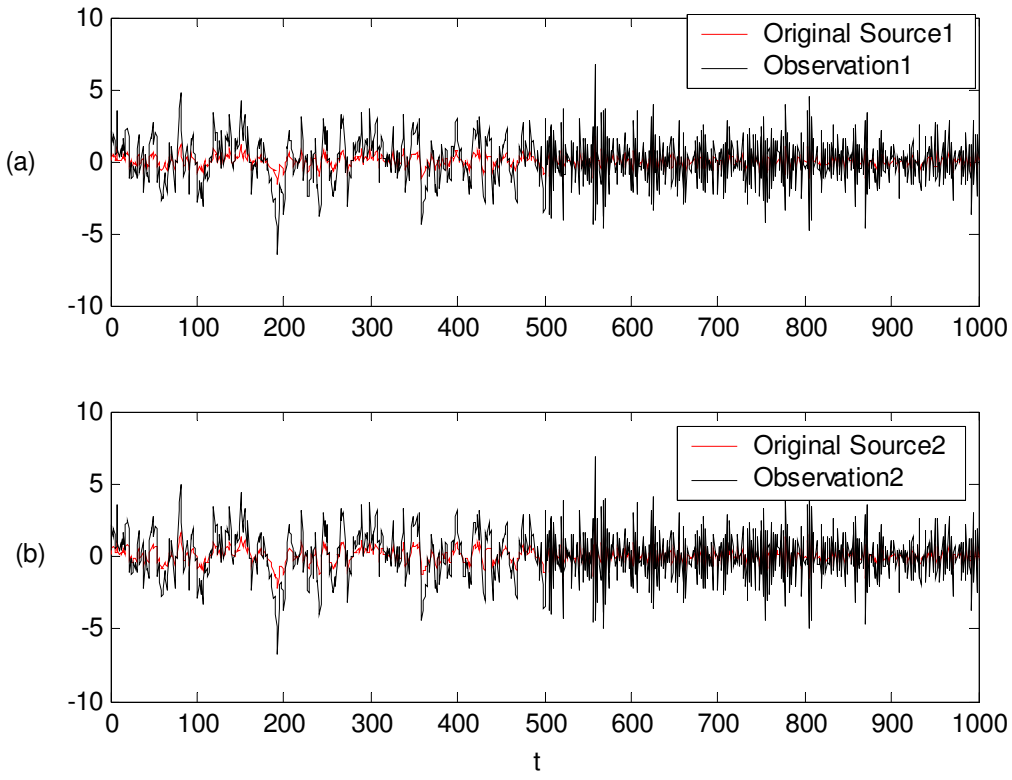


Figure 6.26. E3: Observation and source waveforms of 1 realization

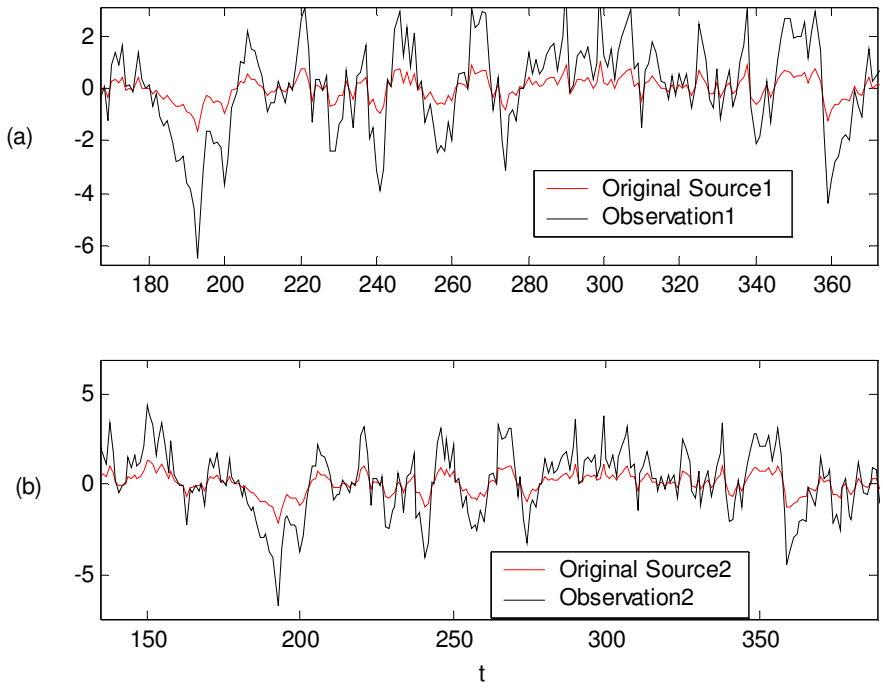


Figure 6.27. E3: Zoomed observation and source waveforms (before change)

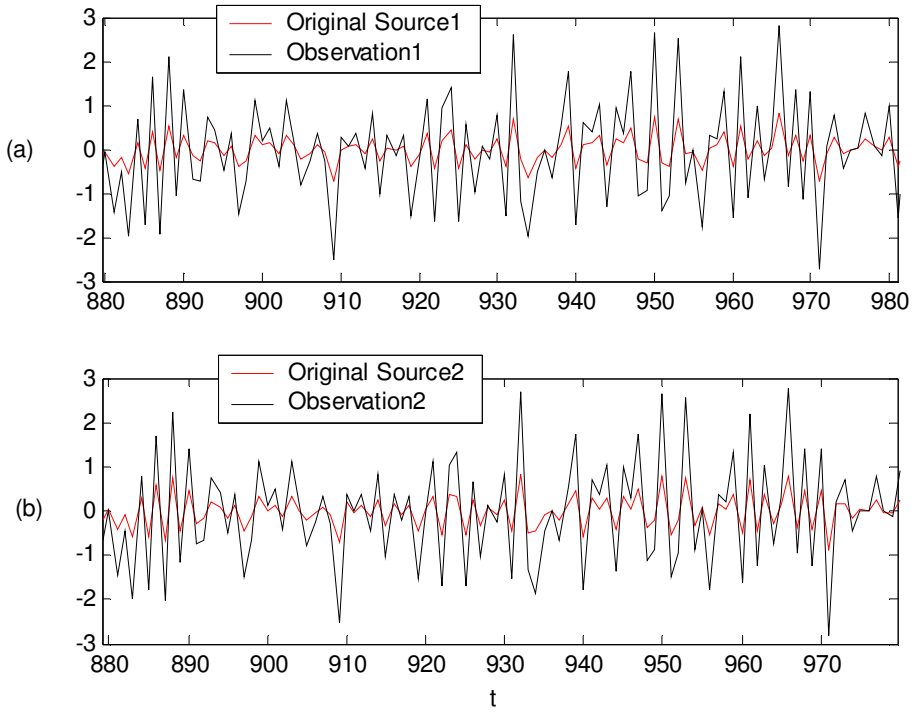


Figure 6.28. E3: Zoomed observation and source waveforms (after change)

Next, the covariance matrix of the VAR(1) process is illustrated for its two different regions, respectively:

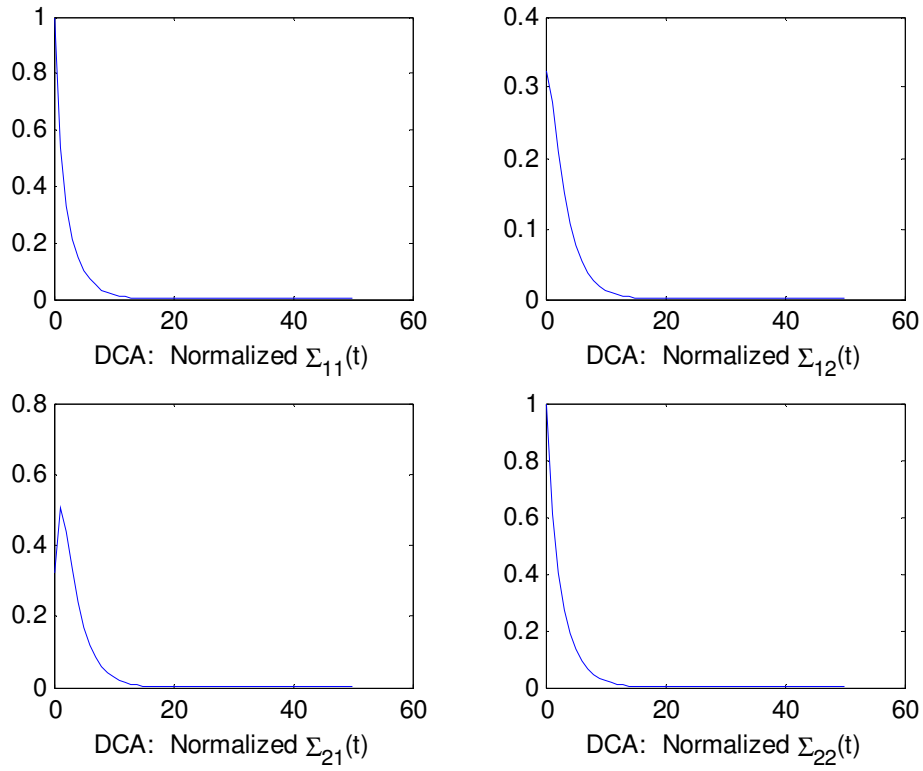


Figure 6.29. a) E3: Normalized covariance matrix of VAR(1) sources before change in time

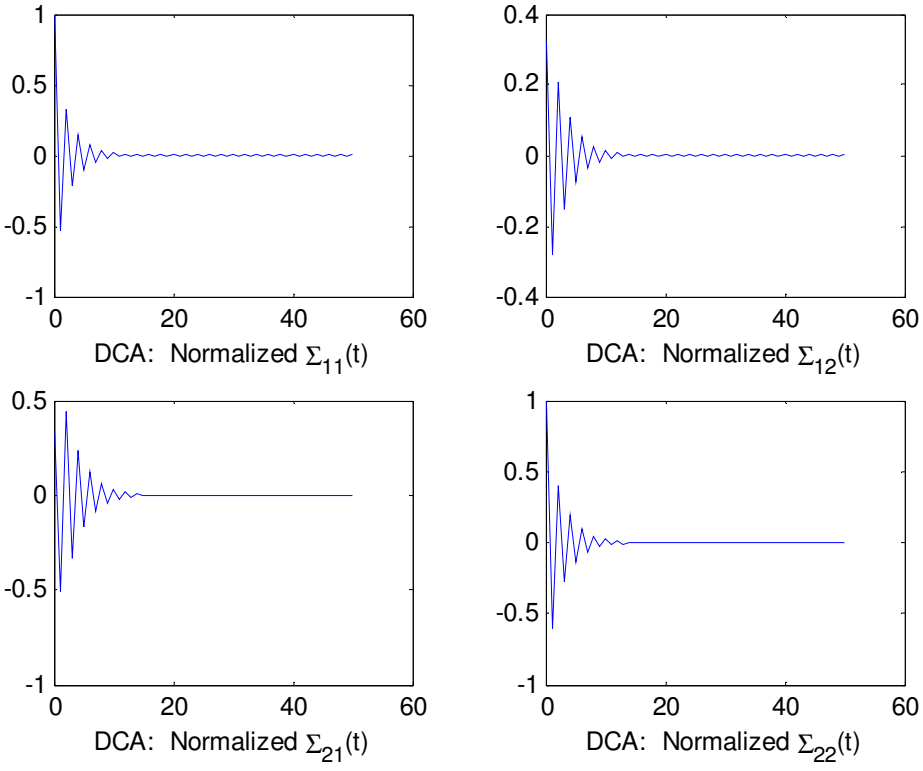


Figure 6.29. b) E3: Normalized covariance matrix of VAR(1) sources after change in time

Below, the mean of the estimated mixing matrix elements are shown:

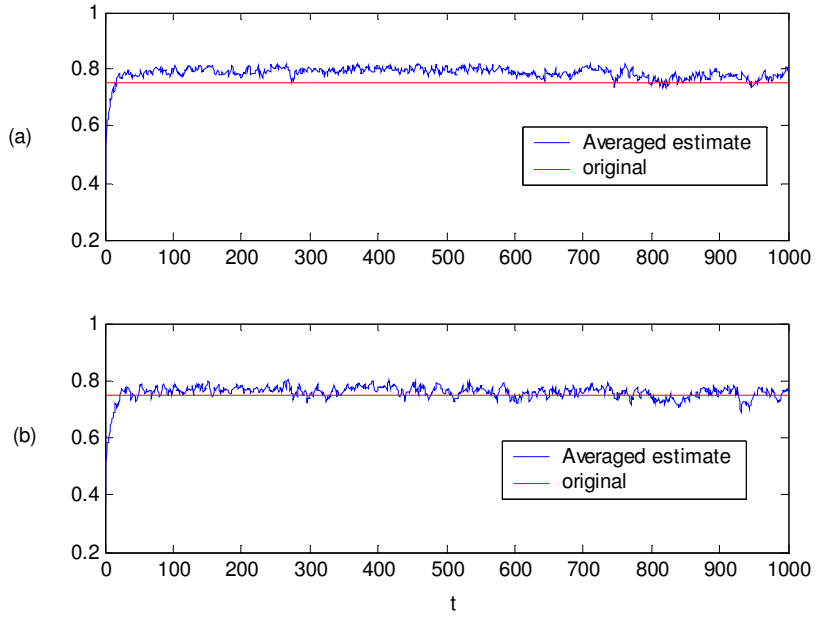


Figure 6.30. E3: Mixing matrix elements and their MMSE estimates, a) $a_{12}(t)$, b) $a_{21}(t)$

Finally, the source estimates and their zoomed waveforms are given, respectively as follows. Here, estimates of an arbitrary realization are not shown, since they are almost the same as those obtained by using the smoothed mixing matrix estimate.

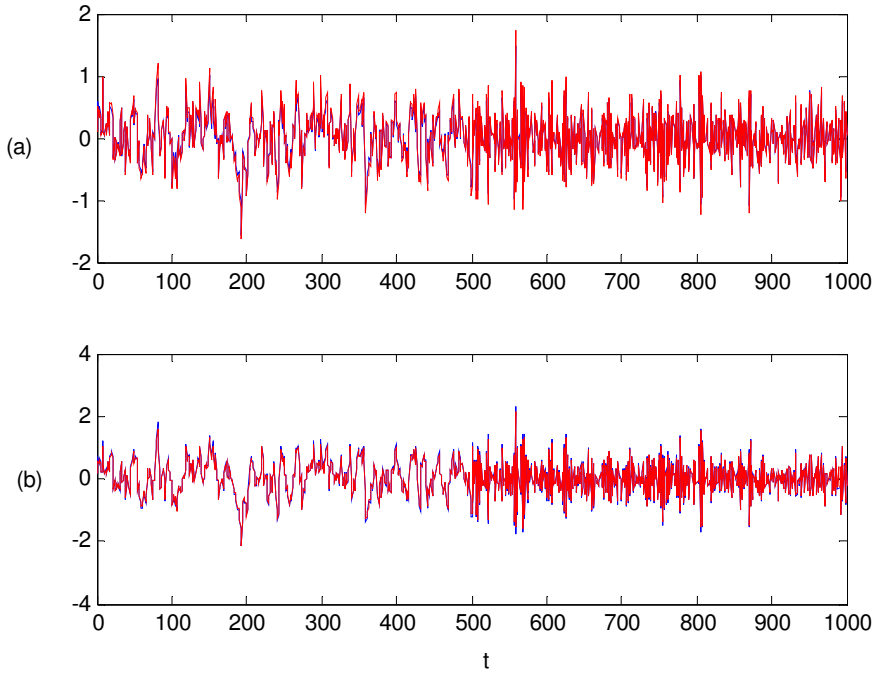


Figure 6.31. E3: Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

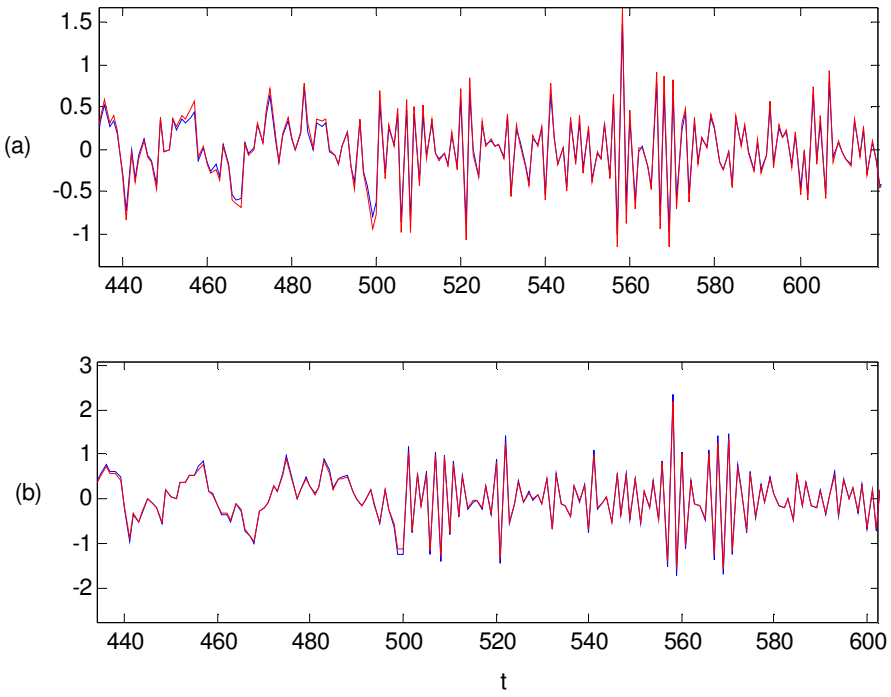


Figure 6.32. E3: Zoomed version of Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

Next, the NMSE curves of the estimated parameters are given.

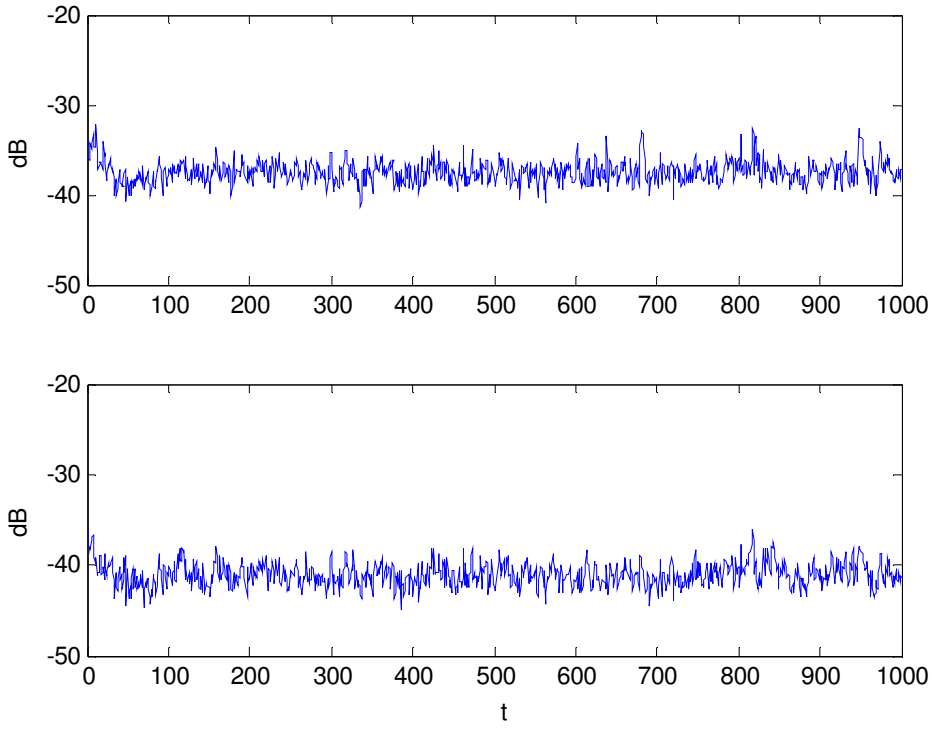


Figure 6.33. E3: NMSE curves of sources, a) First source, b) Second source

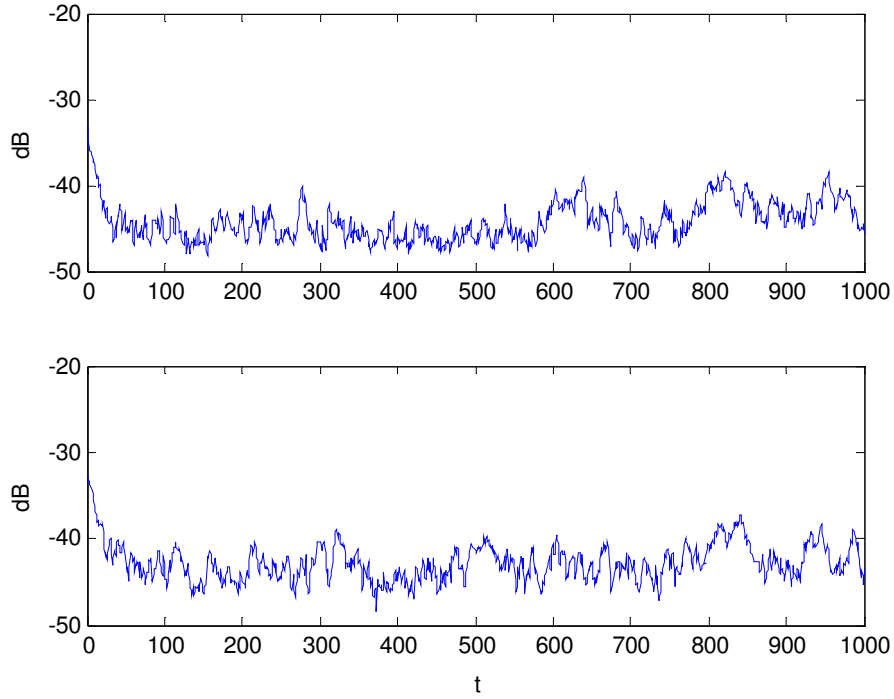


Figure 6.34. E3: NMSE curves of mixing matrix estimates, a) $a_{12}(t)$, b) $a_{21}(t)$

From the above figures, it is seen that the source estimates are almost the same as their original waveforms. It is seen that even at $t = 500$, which is the change point of the source characteristics, the estimates are not affected. Such a perfect estimation is due to the constant mixing matrix. This result justifies the need to take information as much as possible from the observations during the construction of an importance function. Here, the mixing matrix is learned by the algorithm and then it is used to utilize the information coming from the observations.

**Experiment 4 (E4).** A final experiment is presented here to justify that our method is a unifying approach and it can also be used in ICA problems. Therefore, sources are taken to be statistically independent. However, they possess temporal correlations, i.e. AR processes. Following matrices are used here:

$$\begin{aligned} \mathbf{A}_t &= \begin{bmatrix} 2 & 1.5 \\ 1.5 & 2 \end{bmatrix} \text{ for } t < 1000,\ \mathbf{A}_t = \begin{bmatrix} 2 & -1.5 \\ -1.5 & 2 \end{bmatrix} \text{ for } t > 1000 \\ \boldsymbol{\upsilon}_t &= \begin{bmatrix} \upsilon_{1,t} \\ \upsilon_{2,t} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix} \right), \boldsymbol{\Phi}_{1,t} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.9 \end{bmatrix} \end{aligned} \tag{6.52}$$

First, the observation and the source processes are shown as follows:

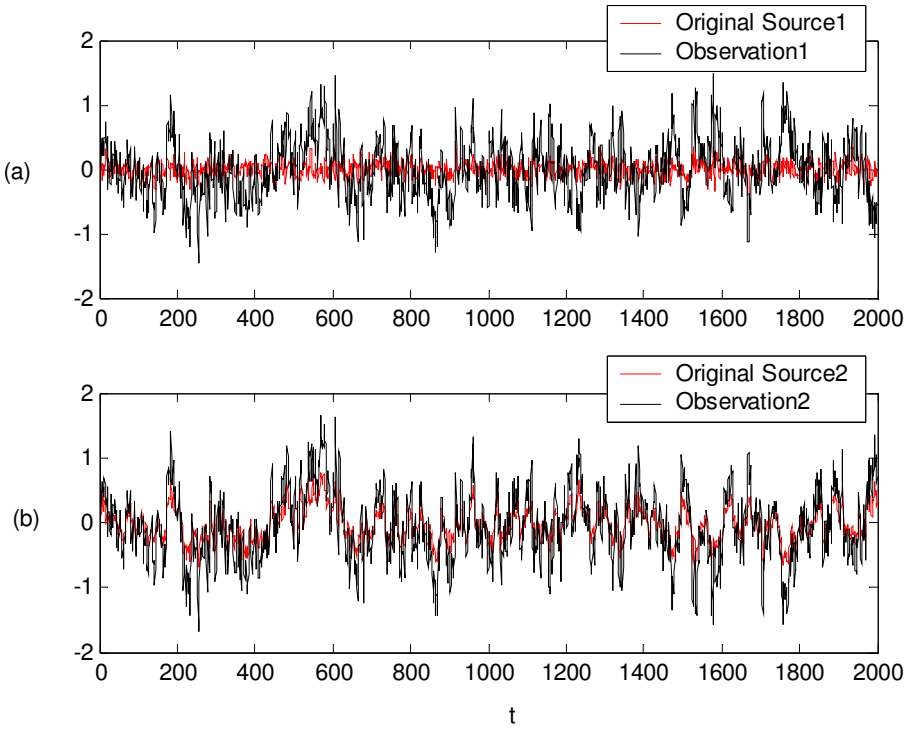


Figure 6.35. E4: Observation and source waveforms of 1 realization

Next, the mean estimate of the mixing matrix elements are shown followed by the source estimates and their zoomed waveforms:

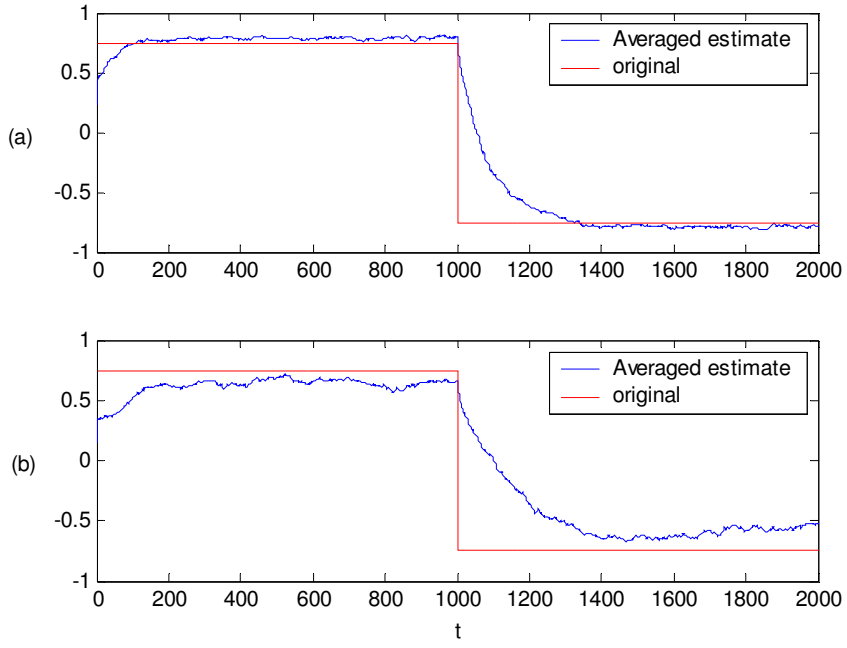


Figure 6.36. E4: Mixing matrix elements and their MMSE estimates, a) $a_{12}(t)$, b) $a_{21}(t)$

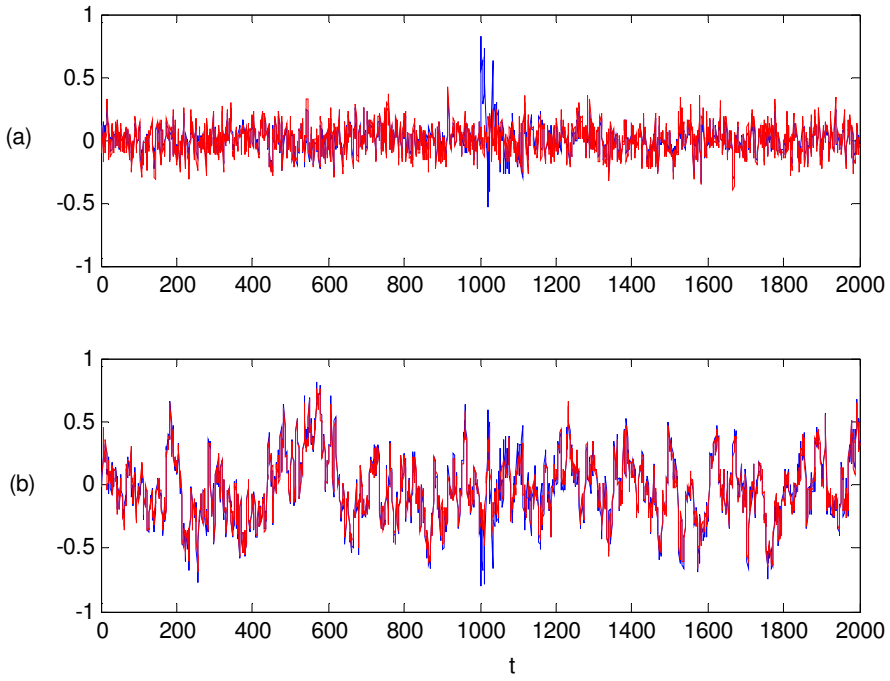


Figure 6.37. E4: Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

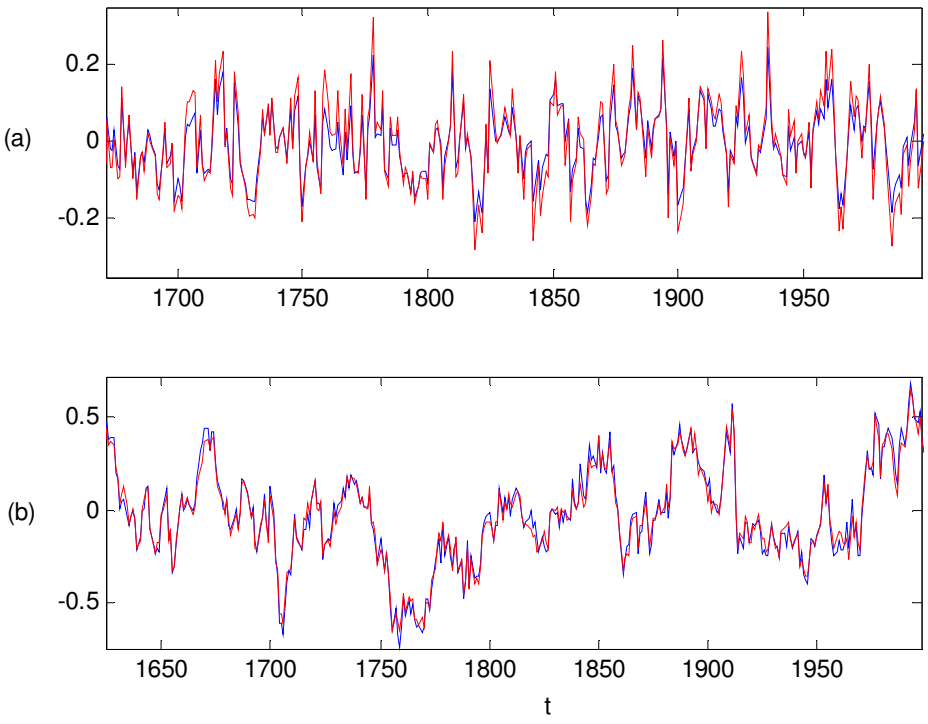


Figure 6.38. E4: Zoomed version of Sources and their MMSE estimates, a) $s_1(t)$, b) $s_2(t)$

Following two figures are provided to show the NMSE curves of the mixing matrix and source estimates.

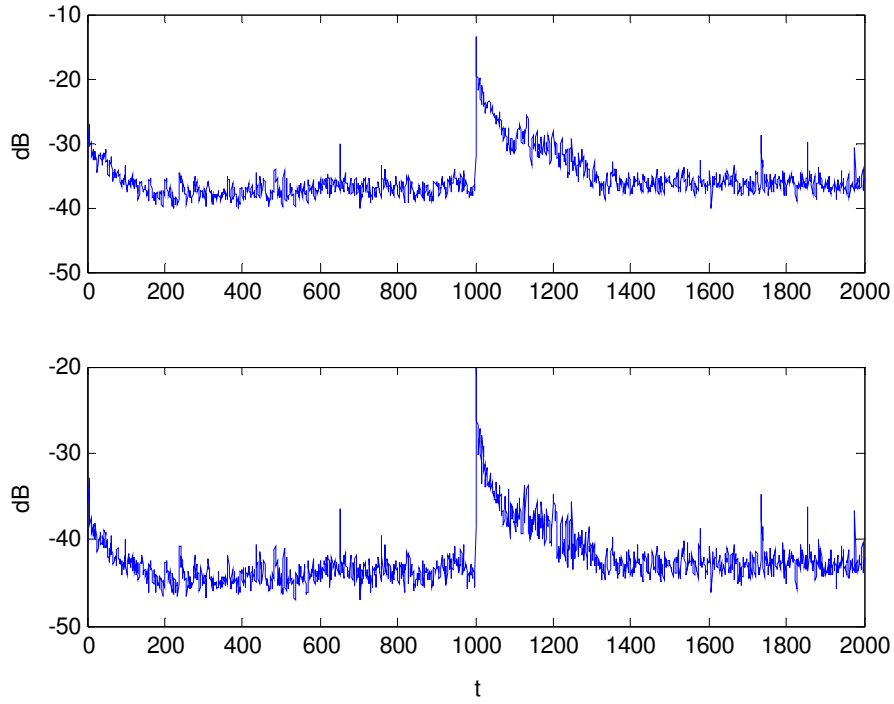


Figure 6.39. E4: NMSE curves of sources, a) First source, b) Second source

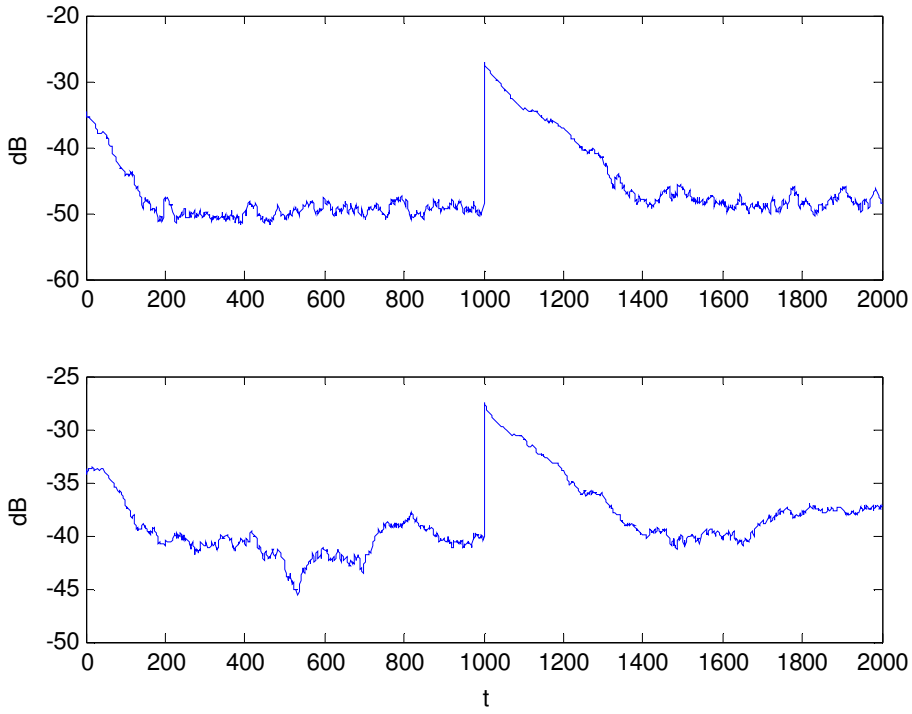


Figure 6.40. E4: NMSE curves of mixing matrix estimates, a) $a_{12}(t)$, b) $a_{21}(t)$

It is observed from these results that our MCARS method can also model ICA problems, successfully. This verifies that our technique presents a unifying approach that can be used in the sequential modeling of non-stationary mixtures of signals.

For the sake of simplicity, the experiments are summarized in the following table.

Table 6.4. Experimental Scenarios of MCARS method

| Experiment No. | Source Correlation | Time-varying $\mathbf{A}_t$ | Time-varying $\mathbf{\Phi}_{1,t}$ |
|---|---|---|---|
| Exp. 1 | by $\boldsymbol{\upsilon}_t$ | Yes | No |
| Exp. 2 | by $\mathbf{\Phi}_{1,t}$ and $\boldsymbol{\upsilon}_t$ | Yes | No |
| Exp. 3 | by $\mathbf{\Phi}_{1,t}$ | No | Yes |
| Exp. 4 | No | Yes | No |

## 6.5. Discussions

In this chapter, we propose two novel techniques to model mixtures of cross-correlated processes. First approach (MCS) can be successfully used when we have strong beliefs about the source distributions. This is a hybrid technique where the mixing matrix is found by particle filtering, whereas the sources are extracted by a MCMC scheme. However, this method can be used for cross-correlated processes with temporal independence. This method was successfully tested in the modeling of non-stationary mixtures of cross-correlated Gaussian sources. Here, sinusoidal mixing matrices have been used to simulate modulation problems that are widely encountered in telecommunications.

In the second part, we generalize our particle filtering methodology to model non-stationary mixtures of cross-correlated AR sources. This technique (MCARS) is capable of modeling such mixtures even though we have vague priors regarding the parameters of interest. Moreover, the need for locally stationary data in the MCS method is relaxed here, allowing the modeling of more general non-stationary data. This has been accomplished by the proposition of a hierarchical particle filtering methodology where importance functions are selected as mixtures using the current values of the sources and their MAP estimates

from the preceding iteration. Various simulation scenarios have been used to test the method, where mixing matrix and the AR coefficient matrix are changed in time. These changes have been performed in terms of abrupt jumps in time. In physical world, such changes can occur during wireless communications or cocktail party problems due to the movement of the sources. In this section, time-variation of the mixing matrices and the AR coefficient elements has been taken as abrupt jumps rather than sinusoidal changes. This is a deliberate choice, which allows the algorithm to learn the reason of time-variation: Since AR processes possess dynamic structures in time, the algorithm would fail to understand the cause of time-variation if the mixing matrix elements were also changed continuously in time. Having constant mixing matrix elements for a period of time allows the algorithm to extract the dynamic behavior of the AR sources. This ambiguity problem is also addressed in (Everson and Roberts, 2000). Had the time-variations of the state variables been known as in (Andrieu and Godsill, 2000; Vermaak *et al*. 2002; Dally and Reilly, 2005), then the problem would have been easier to handle. Therefore, experimental scenarios have been chosen for cases where we do not know the time-variation of the parameters. Successful simulation results show that our methodology is highly flexible and provides a unifying solution for different problems such as ICA and DCA.

# 7. CONCLUSIONS

This thesis brings a unifying methodology to model non-stationary non-Gaussian processes. In the literature, although successful approaches have been developed to model non-stationary processes, they cannot fulfill the increasing need of such modeling schemes. Most of these techniques include variants of Kalman filtering which provide the optimal solution only in case of linearly expressed Gaussian distributed systems. Although its variants, such as the EKF and UKF, can be used in case of nonlinear situations, the need for a unifying and flexible approach has arisen.

With the recent increase in computational capabilities, Bayesian sequential methods, known as the particle filters, have been utilized to process non-stationary non-Gaussian processes in diverse fields, such as biomedicine, telecommunications, astrophysics, geophysics, financial time series analysis and especially target tracking. Particle filters can be thought of as the generalization of Kalman filtering to nonlinear and/or non-Gaussian systems which can be expressed in terms of state-space equations. Although different forms of particle filtering techniques have been developed in the literature, their common property is the need to express a problem in terms of state-space equations whose *functional forms need to be known*. Even though there has been several approaches for artificial modeling of these state-space equations (Djuric *et al*., 2001; Djuric *et al*., 2002), they can only be used in case of processes with *closed form distributions* whose statistical properties are known.

In this thesis, a unifying modeling scheme is proposed where the need of closed form distributions with known statistical properties is relaxed and particle filtering methodology is utilized to model non-stationary non-Gaussian processes. To bring a solution for such a general situation, we model non-stationary *α-stable* processes. The motivation for using these processes in modeling non-Gaussian processes is that they share many common properties with Gaussian distributions, such as the CLT and the stability properties. Therefore, *α-stable* processes constitute a direct generalization to non-Gaussian processes whose pdf's cannot be expressed in closed form except for some limited cases. These

processes are widely used to model impulsive data encountered in computer communications, radar and sonar applications, astrophysics and mobile communications. We propose three novel techniques to model non-stationary *α-stable* processes. In these modeling schemes, we utilized linear autoregressions to express time-structures of the processes, which are widely encountered in the literature (Hamilton, 1994). Our first method is used to model TVAR coefficients of *α-stable* processes, given the *true* values of the distribution parameters. This time-varying nature of the AR coefficients causes data to be non-stationary and it is believed in the literature that teletraffic data possess such a time-varying dependence in time. For our first method, we also provide the PCRLB that bounds the MSE of the estimations from below. By this way, we provide a lower bound for our modeling methodologies to compare with our empirical simulation results. In our second method (HSMC), we propose a hybrid technique to model TVAR *α-stable* processes with *unknown* distribution parameters. Here, both the unknown TVAR coefficients and *constant* distribution parameters of the *α-stable* processes are modeled. Later, we extend our modeling scheme so that in addition to the TVAR coefficients, unknown and *time-varying* distribution parameters of *α-stable* processes can also be modeled (SSMC). To the best of our knowledge, this is the first technique that can be used in the modeling of TVAR *α-stable* processes with time-varying distribution parameters. By our unifying approach, non-stationary *α-stable* processes can be modeled in their most general case. According to the computer simulations, it is demonstrated that both the HSMC and SSMC methods approach to the PCRLB values. Moreover, further empirical studies have shown that estimates with lower variances can be obtained for the distribution parameters in HSMC method when compared with those of SSMC. On the other hand, estimations of TVAR coefficients show similar performances in both methods. That is why, HSMC method should be preferred if TVAR *α-stable* processes with constant unknown distribution parameters are to be modeled. However, if the distribution parameters are also time-varying, SSMC should be used, since HSMC has been developed for modeling constant unknown distribution parameters. The successful performance evaluation of these techniques possess very promising results for the future application areas in this research field where switching between different AR *α-stable* processes need to be modeled in time.

Secondly, we extend our particle filtering methodology so that relationships between different non-stationary non-Gaussian processes can also be modeled; i.e univariate

process modeling of the preceding section is extended to multivariate cases. In order to propose such a technique, we made use of VAR processes where cross-correlations between different processes are modeled through a VAR structure here. Although relationships between different cross-correlated AR processes have been examined in the literature for chemical processes (Hsu, 1997), mobile communications (Jachan and Martz, 2005), and biomedical applications (Möller *et al*., 2001; Sato *et al*., 2006); only Gaussian processes have been modeled in these studies. To the best of our knowledge, our particle filtering methodology is the first approach where cross-correlated VAR non-Gaussian processes can also be modeled. To justify our theory, we have performed computer simulations where both Gaussian and non-Gaussian *time-varying* VAR processes are modeled. Successful performance of our method serves as a major contribution to study cross-correlations between different non-Gaussian processes in future applications. Moreover, by the extension of our particle filtering methodology used for examining cross-correlated processes, we also form a background material to be used in the modeling of *mixtures* of cross-correlated VAR processes, where mixtures of cross-correlated VAR processes are observed instead of their own samples.

Finally, we extend our particle filtering methodology to the most general case, where mixtures of cross-correlated processes are modeled. This problem can also be treated as a DCA application, since latent, cross-correlated processes (sources) are modeled in addition to the mixing matrix. Thus, sources are separated from their observed mixtures. Although source separation problem has been widely studied in the literature, most of the addressed techniques assume statistical independence between the sources, hence they could only provide solutions to ICA (Hyvarinen, 2001). Moreover, despite the diversity of these techniques, they are not flexible in general: Most of the basic ICA methods consider square mixing matrices and do not model noise. Although more flexible approaches have been brought by the utilization of the particle filters (Ahmed *et al*., 2000; Everson and Roberts, 2000), i.i.d. processes have been examined in these studies and time structured sources could only be inserted into this flexible approach later by (Andrieu and Godsill, 2000; Dally and Reilly, 2005). On the other hand, despite the success of the latter techniques, they are also incapable of providing a solution for the DCA problem, since statistical independence between the sources is assumed. A very limited number of studies have been performed to solve the DCA problem (Nuzillard and Nuzillard, 1999; Bach and Jordan,

2000; Barros, 2000). Similar to classical ICA approaches, these methods are not flexible and require specific assumptions about the processes (such as the presence of an orthogonal region in the mixture spectrum) for a possible solution. Among these solutions, the most flexible approach has been brought by (Rowe, 2003) using MCMC strategies. However, its inability to be extended for non-stationary cases, except for a very special sub-case, avoids this method to be used in more general problems. In this thesis, we bring a unifying solution to these problems. By extending our particle filtering methodology, we provide a novel solution to model mixtures of cross-correlated Gaussian processes which is named as the MCS method. Here, we propose a hybrid technique where the mixing matrix is estimated by a particle filter and the sources are extracted by a MCMC algorithm, provided that we have strong beliefs about the source distributions and the mixing matrix elements vary slowly. Successful simulation results have demonstrated that this method serves as a promising modeling scheme where sinusoidal mixing matrices are involved, such as the modulation operations encountered in telecommunications. Next, we drop our assumption that we have strong knowledge about the source distributions and study the most general case. Here, it is assumed that we can have vague priors about the parameters of interest and the source processes possess cross-correlated VAR structures. That is, we extend our particle filtering methodology so that the mixtures of cross-correlated VAR processes can be modeled. We name this technique as MCARS. This method utilizes a hierarchical importance function in the form of a mixture of the current sources and their preceding MAP estimates. To model the sources, VAR modeling of the preceding sections have been used. The successful performance of the proposed method has been demonstrated by computer simulations, where different possible scenarios are tested. As a result, it is concluded that we can successfully model non-stationary mixtures of cross-correlated AR processes which can provide a major contribution in possible future applications, such as wireless communications and teleconference problems where mixtures of cross-correlated sources are observed through time-varying channels due to the movement. Here, simulations have been performed to model time-varying mixing matrix and time-varying cross-correlation effects that can be encountered in these applications. It is concluded that successful modeling performances are obtained when both of the mixing matrix and the VAR coefficient matrix are not simultaneously time-varying. Finally, it is shown that our modeling technique can also model non-stationary mixtures of independent processes, justifying that it is a unifying solution. To the best of our knowledge, MCS and

MCARS are the first techniques that can be used for non-stationary DCA problem and therefore this contribution opens a new research direction for possible applications in this field.

## 7.1. Suggestions for Future Work

This thesis presents a unifying methodology to model non-stationary non-Gaussian processes. To provide such an approach, modeling of non-stationary *α-stable* processes has been performed since this family of distribution is a direct generalization of Gaussian distributions to non-Gaussian distributions, sharing common properties. Next, we extend our methodology to model non-stationary cross-correlated non-Gaussian processes. However, we have taken these non-Gaussian distributions as mixtures of Gaussians with *known* covariance matrices. A future work can be devoted to the modeling of the spatial dependencies between different non-Gaussian process samples where the use of covariance matrix is not enough as in case of *α-stable* distributions. Similarly, the last part of the thesis presents an approach to model non-stationary mixtures of cross-correlated processes. Here, cross-correlated processes have taken to be Gaussian distributed in order to exploit their spatial dependency through their covariance matrices. As a future work, new spatial dependence measures and their temporal relationships could be examined to express the relationships between different non-Gaussian distributions where the use of covariance matrices is not enough and cannot be put into a parametric form.

## 7.2. Publications

The main contributions of this thesis have been presented in the following publications:

- Gençağa, D., E.E. Kuruoğlu, A. Ertüzün, "Estimation of Time-Varying Autoregressive Symmetric Alpha Stable Processes Using a Hybrid Sequential Monte Carlo Method", *submitted to IEEE Transactions on Signal Processing.*

- Gençağa, D., E. E. Kuruoğlu, A. Ertüzün, "Estimation of Time-Varying Autoregressive Symmetric Alpha Stable Processes by Particle Filters", Technical Report, ISTI-CNR, 2006.

- Gençağa, D., E.E. Kuruoğlu and A. Ertüzün, "SAR Image Enhancement using Particle Filters", *ESA-EUSC 2005: Image Information Mining – Theory and Application to Earth Observation*, Frascati, Italy, October 2005.

- Gençağa, D., E.E. Kuruoğlu and A. Ertüzün, "Estimation of Time-Varying Autoregressive Symmetric Alpha Stable Processes by Particle Filters", *13th European Signal Processing Conference (EUSIPCO 2005)*, Antalya, Turkey, September 2005.

- Gençağa, D., E.E. Kuruoğlu and A. Ertüzün, "Bayesian Separation of Non-Stationary Mixtures of Dependent Gaussian Sources", *25th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2005)*, San Jose, CA, USA, August 2005.

- Gençağa, D., E.E. Kuruoğlu and A. Ertüzün, "Zamanla Değişen Özbağlanımlı Cauchy Süreçlerinin Parçacık Süzgeçleri ile Kestirimi", *13th Signal Processing and Applications Conference (IEEE-SİU 2005)*, Kayseri, Turkey, May 2005. (IEEE Best Student Paper Award)

- Gençağa, D. and A. Ertüzün, "Çevrimiçi Bağımsız Bileşen Ayrıştırılması", *11th Signal Processing and Applications Conference*, İstanbul, Turkey, June 2003.

# APPENDIX

In this section, characteristics of several important statistical distributions will be introduced. These will be followed by a brief presentation on ways of choosing conjugate prior distributions which have been utilized in the stationary source separation by MCMC, as outlined in Section 6.2.1. In this appendix, we will refer to (Rowe, 2003). First, some useful statistical distributions are defined as follows:

## A.1. Statistical Distributions

### A.1.1. Scalar Distributions

#### i. Gamma Distribution

A Gamma variate is the sum of the squares of $\nu_0$ centered independent Gaussian variates with common mean $m$ and variance $\upsilon^2$. Such a Gamma distributed random variable is denoted as follows:

$$y|\eta,\kappa \sim \mathcal{G}(\eta,\kappa)$$
$$p(y|\eta,\kappa) = \frac{y^{\eta-1}e^{-y/\kappa}}{\Gamma(\eta)\kappa^{\eta}} \tag{A.1}$$

where $\Gamma(.)$ denotes the gamma function and $y \in \mathbb{R}^+, \eta \in \mathbb{R}^+, \kappa \in \mathbb{R}^+$. The properties of this distribution are given below:

$$E[y] = \eta\kappa \tag{A.2}$$
$$Mode(y) = (\eta - 1)\kappa \tag{A.3}$$
$$\text{var}(y) = \eta\kappa^2 \tag{A.4}$$

### ii. Inverted Gamma Distribution

An Inverted Gamma variate is the reciprocal of a Gamma variate, $\sigma^2 = y^{-1}$. The distribution is illustrated below:

$$\sigma^2 \left| \eta, \kappa \sim \mathcal{IG}(\eta, \kappa) \right.$$
$$p\left(\sigma^2 \left| \eta, \kappa \right.\right) = \frac{\left(\sigma^2\right)^{-(\eta+1)} e^{\frac{-1}{\kappa\sigma^2}}}{\Gamma(\eta)\,\kappa^{\eta}} \tag{A.5}$$

where $\Gamma(.)$ denotes the gamma function and $\sigma^2 \in \mathbb{R}^+, \eta \in \mathbb{R}^+, \kappa \in \mathbb{R}^+$. Here, $\eta$ and $1/\kappa$ denote the shape and scale parameters, respectively. The properties of this distribution are given below:

$$E\left[\sigma^2\right] = \frac{1}{(\eta-1)\,\kappa} \tag{A.6}$$

$$Mode\left(\sigma^2\right) = \frac{1}{(\eta+1)\,\kappa} \tag{A.7}$$

$$\text{var}\left(\sigma^2\right) = \frac{1}{(\eta-1)^2\,(\eta-2)\,\kappa^2} \tag{A.8}$$

## A.1.2. Vector Distributions

### i. Multivariate Normal

If we collect *d* numbers of scalar observations from a Normal distribution and form a vector, then this $(d\text{x}1)$ dimensional vector has multivariate Normal distribution with mean **m** and covariance matrix $\mathbf{\Sigma}$ as shown below:

$$\mathbf{y} \sim \mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$$
$$p\left(\mathbf{y} \left| \mathbf{m}, \mathbf{\Sigma} \right.\right) = (2\pi)^{\frac{-d}{2}} \left|\mathbf{\Sigma}\right|^{\frac{-1}{2}} \exp\left( \frac{-1}{2} (\mathbf{y}-\mathbf{m})^T \mathbf{\Sigma}^{-1} (\mathbf{y}-\mathbf{m})^T \right) \tag{A.9}$$
$$\text{where } \mathbf{y} \in \mathbb{R}^d$$

### A.1.3. Matrix Distributions

#### i. Matrix Normal

If we collect $\tau$ numbers of $(d\text{x}1)$ dimensional Normal vectors, then the $(d\text{x}\tau)$ dimensional matrix is said to have a Matrix Normal distribution. This is a special case of a $d\tau$-variate Multivariate Normal distribution when the covariance matrix is separable. A $d\tau$-variate Multivariate Normal distribution with ($d\tau$ x1) dimensional mean vector **m** and ($d\tau$ x $d\tau$) dimensional covariance matrix $\mathbf{\Psi}$ is shown as follows:

$$p\left(\mathbf{y}\middle|\mathbf{m},\mathbf{\Psi}\right)=\left(2\pi\right)^{\frac{-d}{2}}\left|\mathbf{\Psi}\right|^{\frac{-1}{2}}\exp\left(\frac{-1}{2}\left(\mathbf{y}-\mathbf{m}\right)^{T}\mathbf{\Psi}^{-1}\left(\mathbf{y}-\mathbf{m}\right)\right) \tag{A.10}$$

A separable matrix is denoted by $\mathbf{\Psi}=\mathbf{P}\otimes\mathbf{\Sigma}$. This Kronecker product of $\tau$ and $d$ dimensional matrices is denoted as follows:

$$\mathbf{P}\otimes\mathbf{\Sigma}=\begin{pmatrix}\rho_{11}\mathbf{\Sigma} & \cdots & \rho_{1\tau}\mathbf{\Sigma}\\ \vdots & \vdots & \vdots\\ \rho_{\tau 1}\mathbf{\Sigma} & \cdots & \rho_{\tau\tau}\mathbf{\Sigma}\end{pmatrix} \tag{A.11}$$

where **P** and **Σ** denote temporal and spatial covariances, respectively. If (A.11) is substituted in (A.10), following distribution is obtained:

$$p\left(\mathbf{y}\middle|\mathbf{m},\mathbf{\Sigma},\mathbf{P}\right)=\left(2\pi\right)^{\frac{-d\tau}{2}}\left|\mathbf{P}\otimes\mathbf{\Sigma}\right|^{\frac{-1}{2}}\exp\left(\frac{-1}{2}\left(\mathbf{y}-\mathbf{m}\right)^{T}\left(\mathbf{P}\otimes\mathbf{\Sigma}\right)^{-1}\left(\mathbf{y}-\mathbf{m}\right)\right) \tag{A.12}$$

Here we will make use of the following matrix identities:

$$\left|\mathbf{P}\otimes\mathbf{\Sigma}\right|^{\frac{-1}{2}}=\left|\mathbf{P}\right|^{\frac{-d}{2}}\left|\mathbf{\Sigma}\right|^{\frac{-\tau}{2}} \tag{A.13}$$

and

$$(\mathbf{y}-\mathbf{m})^T(\mathbf{P}\otimes\boldsymbol{\Sigma})^{-1}(\mathbf{y}-\mathbf{m}) = tr\mathbf{P}^{-1}(\mathbf{Y}-\mathbf{M})\boldsymbol{\Sigma}^{1}(\mathbf{Y}-\mathbf{M})^T \quad \text{(A.14)}$$

where

$$\mathbf{y}=\left(\mathbf{y}_1^T,...,\mathbf{y}_\tau^T\right)^T, \mathbf{Y}^T=\left(\mathbf{y}_1,...,\mathbf{y}_\tau\right), \mathbf{m}=vec\left(\mathbf{M}^T\right)=\left(\mathbf{m}_1^T,...,\mathbf{m}_\tau^T\right)^T, \mathbf{M}^T=\left(\mathbf{m}_1,...,\mathbf{m}_\tau\right) \quad \text{(A.15)}$$

By using these definitions, Matrix Normal distribution is expressed in the following form:

$$p\left(\mathbf{Y}\middle|\mathbf{M},\boldsymbol{\Sigma},\mathbf{P}\right)=(2\pi)^{\frac{-d\tau}{2}}\left|\mathbf{P}\right|^{\frac{-d}{2}}\left|\boldsymbol{\Sigma}\right|^{\frac{-\tau}{2}}\exp\left(\frac{-1}{2}tr\mathbf{P}^{-1}(\mathbf{Y}-\mathbf{M})(\boldsymbol{\Sigma})^{-1}(\mathbf{Y}-\mathbf{M})^T\right) \quad \text{(A.16)}$$

In conclusion, a random variable having a $\tau$xd Matrix Normal distribution is denoted as follows:

$$\begin{gathered}\mathbf{Y}\sim\mathcal{N}(\mathbf{M},\mathbf{P}\otimes\boldsymbol{\Sigma})\\ where\ \ \mathbf{Y}\in\mathbb{R}^{\tau\ x\ d},\ \ \mathbf{M}\in\mathbb{R}^{\tau\ x\ d},\ \ \boldsymbol{\Sigma},\mathbf{P}>0\end{gathered} \quad \text{(A.17)}$$

where the properties are given as follows:

$$E[\mathbf{Y}]=\mathbf{M} \quad \text{(A.18)}$$

$$Mode[\mathbf{Y}]=\mathbf{M} \quad \text{(A.19)}$$

$$\text{var}\left(vec\left(\mathbf{Y}^T\right)\right)=\mathbf{P}\otimes\boldsymbol{\Sigma} \quad \text{(A.20)}$$

### ii. Wishart Distribution

A Wishart variate is equal to the transpose product $\mathbf{G}=(\mathbf{Y}-\mathbf{M})^T(\mathbf{Y}-\mathbf{M})$, where $\mathbf{Y}$ is a $\upsilon_0 xd$ Matrix Normal variate with mean $\mathbf{M}$ and covariance matrix $I_{\upsilon_0}\otimes\mathcal{B}$. A $dxd$ random symmetric matrix $\mathbf{G}$ having a Wishart distribution is shown as follows:

$\mathbf{G} \sim \mathcal{W}(\mathcal{B}, d, \upsilon_0)$ where the distribution is as follows:

$$p(\mathbf{G}|\mathcal{B}, d, \upsilon_0) = k_W |\mathcal{B}|^{\frac{-\upsilon_0}{2}} |\mathbf{G}|^{\frac{\upsilon_0 - d - 1}{2}} \exp\left(\frac{-1}{2} tr\mathcal{B}^{-1}\mathbf{G}\right)$$

where (A.21)

$$k_W^{-1} = 2^{\frac{\upsilon_0 d}{2}} \pi^{\frac{d(d-1)}{4}} \prod_{j=1}^{d} \Gamma\left(\frac{\upsilon_0 + 1 - j}{2}\right) \text{ with } \mathbf{G} > 0,\ \upsilon_0 \in \mathbb{R}^+, \mathcal{B} > 0$$

### iii. Inverted Wishart Distribution

An Inverted Wishart variate is the reciprocal of a Wishart variate, i.e. $\mathbf{\Sigma} = \mathbf{G}^{-1}$. It is represented by the following distribution:

$\mathbf{\Sigma} \sim \mathcal{IW}(\mathcal{V}, d, \upsilon)$ where the distribution is as follows:

$$p(\mathbf{\Sigma}|\mathcal{V}, \upsilon) = k_{IW} |\mathcal{V}|^{\frac{-\upsilon - d - 1}{2}} |\mathbf{\Sigma}|^{\frac{\upsilon}{2}} \exp\left(\frac{-1}{2} tr\mathbf{\Sigma}^{-1}\mathcal{V}\right)$$

where (A.22)

$$k_{IW}^{-1} = 2^{\frac{(\upsilon - d - 1)d}{2}} \pi^{\frac{d(d-1)}{4}} \prod_{j=1}^{d} \Gamma\left(\frac{\upsilon - d - j}{2}\right) \text{ with } \mathbf{\Sigma} > 0,\ \upsilon \in \mathbb{R}^+, \mathcal{V} > 0$$

Properties:

$$\begin{aligned}
&E[\mathbf{\Sigma} \mid \upsilon, \mathcal{V}] = \frac{\mathcal{V}}{\upsilon - 2d - 2} \\
&Mode[\mathbf{\Sigma} \mid \upsilon, \mathcal{V}] = \frac{\mathcal{V}}{\upsilon} \\
&\text{var}(\sigma_{ii} \mid \upsilon, \mathcal{V}) = \frac{2\nu_{ii}^2}{(\upsilon - 2d - 2)^2 (\upsilon - 2d - 4)} \\
&\text{var}(\sigma_{ii'} \mid \upsilon, \mathcal{V}) = \frac{\nu_{ii}\nu_{i'i'} + \frac{\upsilon - 2d}{(\upsilon - 2d - 2)}\nu_{ii'}^2}{(\upsilon - 2d - 2)^2 (\upsilon - 2d - 4)}
\end{aligned} \quad \text{(A.23)}$$

## A.2. Conjugate Prior Distributions

Conjugate prior distributions are defined as the distributions forms whose form conserves itself after the multiplication by the likelihood function. Thus, their posterior distributions also possess the same form. These priors are easily formed by using the likelihood function. Below, we explain how to choose some basic conjugate priors:

For univariate Normal distribution, first the likelihood is written as follows:

$$p\left(y \mid m,\sigma^2\right) \propto \left(\sigma^2\right)^{\frac{-1}{2}} \exp\left(\frac{-\left(y-m\right)^2}{2\sigma^2}\right) \tag{A.24}$$

Then, the roles of *y* and *m* are interchanged:

$$p\left(m\right) \propto \left(\sigma^2\right)^{\frac{-1}{2}} \exp\left(\frac{-\left(m-y\right)^2}{2\sigma^2}\right) \tag{A.25}$$

To get rid of the dependency on the data, observation data is replaced by a constant as follows:

$$p\left(m \mid \sigma^2\right) \propto \left(\sigma^2\right)^{\frac{-1}{2}} \exp\left(\frac{-\left(m-y_0\right)^2}{2\sigma^2}\right) \tag{A.26}$$

where $y_0$ is a hyperparameter.

If we wish to form a conjugate prior for the variance, this can be given by the Inverted Gamma distribution, using the above methodology. In the Normal likelihood function of (A.24), the roles of *y* and $\sigma^2$ are changed now:

$$p\left(\sigma^2\right) \propto \left(\sigma^2\right)^{\frac{-1}{2}} \exp\left(\frac{-\left(y-m\right)^2}{2\sigma^2}\right) \tag{A.27}$$

If the dependency on the observation data is removed, a distribution proportional to the Inverted Gamma is obtained as follows:

$$p\left(\sigma^2\right) \propto \left(\sigma^2\right)^{\frac{v}{2}} \exp\left(\frac{-(q)^2}{2\sigma^2}\right) \tag{A.28}$$

By using a similar methodology, conjugate priors for vector variates are obtained as follows for the mean vector and covariance matrix of a Normal distribution by writing the likelihood function first:

$$p\left(\mathbf{y} \mid \mathbf{m}, \boldsymbol{\Sigma}\right) \propto \left(\boldsymbol{\Sigma}\right)^{\frac{-1}{2}} \exp\left(\frac{-1}{2}\left(\mathbf{y}-\mathbf{m}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{y}-\mathbf{m}\right)\right) \tag{A.29}$$

Then, the roles of data and the mean vector are changed:

$$p\left(\mathbf{m} \mid \boldsymbol{\Sigma}\right) \propto \left(\boldsymbol{\Sigma}\right)^{\frac{-1}{2}} \exp\left(\frac{-1}{2}\left(\mathbf{m}-\mathbf{y}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{m}-\mathbf{y}\right)\right) \tag{A.29}$$

and afterwards, data dependency is removed which means that the conjugate prior of the mean vector is a Normal distribution:

$$p\left(\mathbf{m} \mid \boldsymbol{\Sigma}\right) \propto \left(\boldsymbol{\Sigma}\right)^{\frac{-1}{2}} \exp\left(\frac{-1}{2}\left(\mathbf{m}-\mathbf{m}_0\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{m}-\mathbf{m}_0\right)\right) \tag{A.30}$$

Similarly, if the roles of the observation and the covariance matrix are changed and the property of the trace operator is used, the following form can be written:

$$p\left(\boldsymbol{\Sigma}\right) \propto \left(\boldsymbol{\Sigma}\right)^{\frac{-1}{2}} \exp\left(\frac{-1}{2} tr\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}-\mathbf{m}\right)\left(\mathbf{y}-\mathbf{m}\right)^T\right) \tag{A.31}$$

If the data dependency is removed, the following Inverted Wishart distribution is obtained:

$$p(\boldsymbol{\Sigma}) \propto (\boldsymbol{\Sigma})^{\frac{-\upsilon}{2}} \exp\left(\frac{-1}{2} tr\boldsymbol{\Sigma}^{-1}\mathcal{V}\right) \tag{A.32}$$

For matrix variates, the same methodology is used again. For a Normal data in matrix form, the Matrix Normal likelihood can be written in the following form using (A.16):

$$p(\mathbf{Y}|\mathbf{M},\boldsymbol{\Sigma},\mathbf{P}) \propto |\mathbf{P}|^{\frac{-d}{2}} |\boldsymbol{\Sigma}|^{\frac{-\tau}{2}} \exp\left(\frac{-1}{2} tr\mathbf{P}^{-1}(\mathbf{Y}-\mathbf{M})(\boldsymbol{\Sigma})^{-1}(\mathbf{Y}-\mathbf{M})^{T}\right) \tag{A.33}$$

Again, the roles of the observation and the mean are changed:

$$p(\mathbf{Y}|\mathbf{M},\boldsymbol{\Sigma},\mathbf{P}) \propto |\mathbf{P}|^{\frac{-d}{2}} |\boldsymbol{\Sigma}|^{\frac{-\tau}{2}} \exp\left(\frac{-1}{2} tr\mathbf{P}^{-1}(\mathbf{M}-\mathbf{Y})(\boldsymbol{\Sigma})^{-1}(\mathbf{M}-\mathbf{Y})^{T}\right) \tag{A.34}$$

which shows that the prior of the mean should be a Matrix Normal as well. If data dependency is removed by putting a hyperparameter instead of the observation, we can write the following equation:

$$p(\mathbf{M}|\boldsymbol{\Sigma},\mathbf{P}) \propto |\mathbf{P}|^{\frac{-d}{2}} |\boldsymbol{\Sigma}|^{\frac{-\tau}{2}} \exp\left(\frac{-1}{2} tr\mathbf{P}^{-1}(\mathbf{M}-\mathbf{M}_0)(\boldsymbol{\Sigma})^{-1}(\mathbf{M}-\mathbf{M}_0)^{T}\right) \tag{A.35}$$

For the covariance matrix, we can write the following form by interchanging $\mathbf{Y}$ and $\boldsymbol{\Sigma}$, and then use the property of the trace operator:

$$p(\boldsymbol{\Sigma}) \propto |\mathbf{P}|^{\frac{-d}{2}} |\boldsymbol{\Sigma}|^{\frac{-\tau}{2}} \exp\left(\frac{-1}{2} tr(\boldsymbol{\Sigma})^{-1}(\mathbf{M}-\mathbf{M}_0)^{T}\mathbf{P}^{-1}(\mathbf{M}-\mathbf{M}_0)\right) \tag{A.36}$$

which fits into the following Inverted Wishart form:

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{\frac{-\upsilon}{2}} \exp\left(\frac{-1}{2} tr(\boldsymbol{\Sigma})^{-1}\mathcal{V}\right) \tag{A.37}$$

# REFERENCES

Anderson, B. D. and J. B. Moore, 1979, *Optimal Filtering*, Prentice-Hall, New Jersey.

Andrieu, C. and S. Godsill, 2000 , "A particle filter for model based audio source separation", *Proceedings of the International Workshop on Independent Component Analysis and Signal Separation: ICA'2000*, Helsinki, Finland.

Ahmed, C., A. Andrieu, A. Doucet, and P. J. W. Rayner, 2000, "On-line non-stationary ICA using mixture models", *Proceedings of International Conf. on Acoustics, Speech and Signal Processing*, pp. 3148-3151.

Amari, S., 1998, "Natural Gradient works efficiently in learning", *Neural Computation*, Vol. 10, No. 2, pp. 251-276.

Arulampalam, M. J., S. Maskell, N. Gordon and T. Clapp, 2002, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking", *IEEE Trans. on Signal Processing*, Vol. 50, No. 2, February.

Attias, H., 1999, "Independent Factor Analysis", *Neural Computation*, Vol. 14, No. 4, pp. 803-851.

Bach, F. R. and M. I. Jordan, 2002, "Tree-dependent component analysis", *Uncertainty in Artificial Intelligence (UAI): Proceedings of the Eighteenth Conference*.

Barros, A. K., 2000, "The Independence Assumption: Dependent Component Analysis," in M. Girolami (ed.), *Advances in Independent Component Analysis,* pp. 63-70, Springer, London.

Bates, S. and S. McLaughlin, 1997, "Testing the Gaussian assumption for self-similar teletraffic model", in *Proc. IEEE Higher Order Statistics Workshop*, Bannf, pp. 444-447.

Bedini L., D. Herranz, E. Salerno, C. Baccigalupi, E. E. Kuruoğlu and A. Tonazzini, 2005, "Separation of Correlated Astrophysical Sources Using Multiple-Lag Data Covariance Matrices", *EURASIP Journal on Applied Signal Processing*, Vol. 15, pp. 2400-2412.

Bell, A. and T. Sejnowski, 1995, "An information maximization approach to blind separation and blind deconvolution", *Neural Computation*, Vol. 7, No. 6, pp. 1004-1034.

Belouchrani, A. K., A. Meraim, J. F. Cardoso, and E. Moulines, 1997, "A blind source separation technique based on second order statistics", *IEEE Trans. on Signal Processing*, Vol. 45, No. 2, pp. 434-444.

Berzuini, C. and W. Gilks, 2001, "Resample-Move filtering with cross-model jumps", in Doucet, A., N. de Freitas and N. Gordon, (eds), *Sequential Monte Carlo Methods in Practice*, Springer, NY.

Bolic, M., P. M. Djuric and S. Hong, 2003, "New Resampling Algorithms for Particle Filters", *Proceedings of International Conf. on Acoustics, Speech and Signal Processing*, Hong Kong.

Box, G. E. P. and G. C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publishing Company.

Cardoso, J. F., 1998, "Blind Signal Separation: statistical principles", *Proceedings of the IEEE*, Vol. 9, No. 10, pp. 2009-2025, October.

Cardoso J. F. and B. H. Laheld, 1996, "Equivariant adaptive source separation", *IEEE Trans. on Signal Processing*, Vol. 44, No. 12, pp. 3017-3030.

Cardoso, J. F. and D. Pham., 2001, "Separation of non-stationary sources", in S. J. Roberts and R.M. Everson (eds.), *Independent Components Analysis: Principles and Practice*, Cambridge University Press.

Cardoso, J. F., 2001, "The three easy routes to independent component analysis; contrasts and geometry", in *Proc. of the ICA 2001 workshop*, San Diego, December.

Carpenter, J., P. Clifford, and P. Shephard, 1999, *Building robust simulation-based filters for evolving data sets*, unpublished, Department of Statistics, Oxford University.

Comon, P., 1994, "Independent Component Analysis, A new concept?", *Signal Processing*, Vol. 36, pp. 287-314.

Cover, T. and J. A. Thomas, 1991, *Elements of Information Theory*, John Wiley.

Dally, M. J. and J. P. Reilly, 2005, "Marginalized Particle Filtering for Blind System Identification", *Proceedings of European Signal Processing Conference*, Antalya 4-8, 2005.

Davis, R., A. K. Knight, and J. Liu, 1992, "M-estimation for autoregressions with infinite variance ," *Stochastic Proc. and their Applications,* Vol. 40, issue 1, pp. 145-180, February.

Djafari, A. M., 1999, "A Bayesian approach to source separation*", 19th International Workshop on Maximum Entropy and Bayesian Methods*, Idaho, USA, Aug. 4-8.

Djuric, P. M., J. H. Kotecha, J. Y. Tourneret, and S. Lesage, 2001, "Adaptive Signal Processing by Particle Filters and Discounting of Old Measurements," in *Proc. IEEE Int. Conf. on Acoustics., Speech, Signal Processing.* , Salt Lake City, USA, , pp. 3733-3736.

Djuric, P. M., J. H. Kotecha, F. Esteve, and E. Perret., 2002, "Sequential Parameter Estimation of Time-Varying Non-Gaussian Autoregressive Processes," *EURASIP Journal on Applied Signal Processing*, vol. 2002, issue 8, pp. 865-875.

Doucet, A. S. Godsill, and C. Andrieu, 2000, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, Vol. 10, No. 3, pp. 197-208, July.

Doucet, A., N. de Freitas and N. Gordon, (editors), 2001, *Sequential Monte Carlo Methods in Practice*, Springer, NY.

Everson, R. S. Roberts, 2000, "Particle Filters for Non-Stationary ICA" in M. Girolami (ed.), *Advances in Independent Component Analysis*, pp. 23-39, Springer, London.

Feller, W., 1966, *An Introduction to Probability Theory and Its Applications*, Vol. 2, John Wiley and Sons.

Gelman, A. B., J. S. Carlin, H. S. Stern, and D. B. Rubin, 1995, *Bayesian Data Analysis*, Chapman & Hall/ CRC.

Gençağa, D., E.E. Kuruoğlu and A. Ertüzün, 2005, "Zamanla Değişen Özbağlanımlı Cauchy Süreçlerinin Parçacık Süzgeçleri ile Kestirimi", *13th Signal Processing and Applications Conference (IEEE-SİU 2005)*, Kayseri, Turkey, May.

Gençağa, D., E.E. Kuruoğlu and A. Ertüzün, 2005, "Bayesian Separation of Non-Stationary Mixtures of Dependent Gaussian Sources", *25th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt 2005)*, San Jose, CA, USA, August.

Gençağa, D., E.E. Kuruoğlu and A. Ertüzün, 2005, "Estimation of Time-Varying Autoregressive Symmetric Alpha Stable Processes by Particle Filters", *13th European Signal Processing Conference (EUSIPCO 2005)*, Antalya, Turkey, September.

Gençağa, D., E.E. Kuruoğlu, A. Ertüzün, 2006, "Estimation of Time-Varying Autoregressive Symmetric Alpha Stable Processes Using a Hybrid Sequential Monte Carlo Method", *submitted to IEEE Transactions on Signal Processing.*

Gilks, W.R., S. Richardson, and D.J. Spiegelhalter, 1998, (eds.), *Markov Chain Monte Carlo in Practice*, Chapman and Hall.

Godsill, S. and E. E. Kuruoğlu, 1999, "Bayesian Inference for Time Series with Heavy-Tailed Symmetric α-Stable Noise Processes," in *Proc. Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, American University, Washington D.C., USA, June 3-5,.

Gordon, N. J., D. J. Salmond, A. F. M. Smith, 1993, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation", *IEE Proceedings-F*, Vol. 140, No. 2, April.

Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press, NJ.

Haykin, S., 1996, *Adaptive Filter Theory*, Prentice Hall,.

Haykin, S., 2000, *Unsupervised Adaptive Filtering, volume 1: Blind Source Separation*, Wiley Interscience.

Hsu, K. J., 1997, "Application of vector autoregressive time series analysis to aerosol studies", *Tellus*, 49B, pp. 327-342.

Hyvarinen, 1998, "Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood", *Neurocomputing*, Vol. 22, pp. 49-67.

Hyvarinen, 1999, "Gaussian moments for noisy independent component analysis", *IEEE Signal Processing Letters*, 6, pp. 145-147.

Hyvärinen, A. and E. Oja., 2000, "Independent Component Analysis: Algorithms and Applications", *Neural Networks*, 13(4-5):411-430.

Hyvarinen, A., J. Karhunen and E. Oja, 2001, *Independent Component Analysis*, John Wiley and Sons.

Jachan, M. and G. Matz, 2005, "Nonstationary vector AR modeling of wireless channels", *Proc. IEEE SPAWC-2005*, New York, June, Vol. 1, pp. 648-652.

Julier, S. J. and J. K. Uhlman, 1996, A General Method for Approximating Nonlinear Transformations of Probability Distributions, *Technical Report*, RRG, Dept. of Engineering Science, University of Oxford.

Jutten, C. and J. Herault, 1991, "Blind separation of sources, part 1: An adaptive algorithm based on neuromimetic architecture", *Signal Processing*, 24, pp. 1-10.

Kailath, T., A. H. Sayed, and B. Hassibi, 2000, 'Linear Estimation', Prentice Hall, NJ.

Kalman, R. E., 1960, "A new approach to linear filtering and prediction problems", *Trans. ASME J. Basic Eng.*, 82, pp. 34-45. [Also published as ASME Paper 59-IRD-11.]

Kanter, M. And W. L. Steiger, 1974, "Regression and Autoregression with Infinite Variance", *Adv. Appl. Prob.*, Vol. 6, pp. 768-783.

Knuth, K.H., 1999, "A Bayesian approach to source separation", *Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation: ICA'99*, Aussois, France, Jan., pp. 283-288.

Kuruoğlu, E. E., P. Rayner, and W. Fitzgerald, 1997, "Least $L_p$ Norm Estimation of Autoregressive Model Coefficients of Symmetric α-Stable Processes", *IEEE Signal Proc. Letters*, vol. 4, issue 7, pp. 201-203, July.

Kuruoğlu, E. E., C. Molina and W. J. Fitzgerald, 1998, "Approximation of Alpha-Stable Probability Densities Using Finite Mixtures of Gaussians", *in Proc. European Signal Processing Conference*, Rhodes, Greece, pp. 989-992.

Kuruoğlu, E. E., 2002, "Nonlinear Least $l_p$ –Norm Filters for Nonlinear Autoregressive α-Stable Processes", *Digital Signal Processing*, 12, pp. 119-142.

Kuruoğlu, E. E., L. Bedini, M. T. Paratore, E. Salerno and A. Tonazzini, 2003, "Source separation in astrophysical maps using independent factor analysis", *Neural Networks*, 16, pp. 479-491.

Liu, J. S. and R. Chen, 1998, "Sequential Monte Carlo methods for dynamic systems", Journal of the American Statistical Association, 93, pp. 1032-1044.

Lombardi, M. and S. Godsill, 2004, "Monte Carlo Bayesian Filtering and Smoothing for TVAR Signals in Symmetric α-stable noise," in *Proc. 12th European Signal Processing Conf.*, Vienna, Austria, , pp. 865-872.

Lütkepohl, 1993, *Introduction to Multiple Time Series Analysis*, Springer, Berlin.

MacKay, D., 2003, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press.

Maino D., A. Farusi, C. Baccigalupi, F. Perotta, A. J. Banday, L. Bedini, C. Burigana, G. De Zotti, K. M. Grski and E. Salerno, 2002, "All-sky astrophysical component separation with fast independent component analysis (fastica)", *Monthly Notices of the Royal Astronomical Society*, 334, pp. 53-68.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, 1953, "Equations of state calculations by fast computing machines", *Journal of Chemical Physics*, 21, pp. 1087-1091.

Moulines, E., J. F. Cardoso and E. Gassiar, 1997, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models", *Proceedings of ICASSP'97*, 5, pp. 3617-3620.

Möller, E., B. Schack, M. Arnold and H. Witte, 2001, "Instantaneous multivariate EEG coherence analysis by means of adaptive high-dimensional autoregressive models", *Journal of Neuroscience Methods*, 105, pp. 143-158.

Neumaier, A. D. and T. Schneider, 2001, "Estimation of parameters and eigenmodes of multivariate autoregressive models", *ACM Transactions on Mathematical Software*, Vol. 27, No. 1, pp. 27-57.

Nikias, C. L. and M. Shao, 1995, *Signal Processing with Alpha-Stable Distributions and Applications*, Prentice-Hall.

Nolan, J. P. and A. Swami, 1999, (Eds.), "Proceedings of workshop on applications of heavy-tailed distributions in economics, statistics and engineering," Washington, D.C., U.S.A.

Nuzillard, D. and J-M. Nuzillard, 1999, "Blind source separation applied to non-orthogonal signals", *Proc. ICA'99*, pp. 25-30.

Parga, N JP Nadal, 2000, "Blind Source Separation with time dependent mixtures" *Signal Processing*, Vol. 80, No. 10, pp. 2187-2194.

Pierce, R. D., 1997, "Application of the positive alpha stable distribution," in *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics*, pp. 420-424.

Pitt, M. K. and N. Shephard, 1999, "Filtering via Simulation: Auxiliary particle filters", Journal of the American Statistical Association, Vol. 94, No. 446, pp. 590-599.

Rappaport, T., 2001, *Wireless Communications: Principles and Practice*, Prentice Hall.

Resnick, S., 1997, "Heavy tail modeling and teletraffic data", *The Annals of Statistics*, vol. 25, no. 5, pp. 1805-1869, Oct.

Ristic, S. Arulampalam, and N. Gordon, 2004, *Beyond the Kalman Filter*, Artech House Publishers.

Robert, C. and G. Casella, 1999, *Monte Carlo Statistical Methods*, Springer.

Rowe, D. B., 2003, *Multivariate Bayesian Statistics*, *Models for Source Separation and Signal Unmixing*, Chapman & Hall/CRC.

Samorodnitsky, G. and M. S. Taqqu, 1994, *Stable Non-Gaussian Random Processes, Stochastic Models with Infinite Variance*, Chapman & Hall.

Sato, J. R., P. A. Morettin, P. R. Arantes and E. Amaro Jr., 2006, "Wavelet Based Time-Varying Vector Autoregressive Modelling", to be published in *Computational Statistics and Data Analysis*.

Shanmugan, K. S., and A. M. Breipohl, 1988, *Random Signals, Detection, Estimation and Data Analysis*, John Wiley & Sons.

Sivia, D. S., 1998, *Data Analysis: A Bayesian Tutorial*, Oxford University Press, NY.

Tanner, M., 1996, *Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions*, Springer, NY.

Thavaneswaran, A. and S. Peiris, 1999, "Estimation for Regression with Infinite Variance Errors," *Mathematical and Computer Modelling*, vol. 29, pp. 177-180.

Tichavsky, P., C. H. Muravchik, and A. Nehorai, 1998, "Posterior Cramer-Rao Bounds for Discrete-Time Nonlinear Filtering," *IEEE Transactions on Signal Processing,* Vol. 46, No. 5, May, pp.1386-1396.

Wan, E. and R. van der Merwe, 2000, "The Unscented Kalman Filter for Nonlinear Estimation", in *Proceedings of Symposium 2000 on Adaptive Systems for Signal Processing, Communication and Control (AS-SPCC)*, IEEE, Lake Louise, Alberta, Canada, October.

Van der Merwe, N. de Freitas, A. Doucet and E. Wan, 2000, "The Unscented Particle Filter", num. CUED/F-INFENG/TR 380, Cambridge University Engineering Department, Cambridge, England, August.

Van Trees, H. L., 1968, *Detection, Estimation and Modulation Theory, Part 1*, Wiley, NY.

Vermaak, J., C. Andrieu, A. Doucet and S. J. Godsill, 2002, “Particle Methods for Bayesian Modeling and Enhancement of Speech Signals”, *IEEE Trans. on Speech and Audio Processing*, Vol. 10, No. 3, pp.173-185.

Zolotarev, V. M., 1989, Ed., *One-Dimensional Stable Distributions*, Providence, R.I: Amer. Math. Soc.