Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

# Identification of Continuous-time Transfer Function Models from Non-uniformly Sampled Data in Presence of Colored Noise

F. Chen[1,2], H. Garnier[1,2], M. Gilson[1,2], J. C. Agüero[3], B. I. Godoy[3,4]

[1] *University of Lorraine, CRAN, UMR 7039, 2 rue Jean Lamour, F-54519 Vandœuvre-les-Nancy, France*
[2] *CNRS, CRAN, UMR 7039, France*
*fengwei.chen,hugues.garnier,marion.gilson@univ-lorraine.fr*
[3] *Centre of Complex Dynamics Systems and Control (CDSC)*
*The University of Newcastle, Australia.*
*{juan.aguero, boris.godoy}@newcastle.edu.au*
[4] *School of Electrical Eng. and Telecommunications*
*The University of New South Wales, Australia.*
*b.godoy@unsw.edu.au*

**Abstract:** This paper considers the problem of continuous-time model identification from non-uniformly sampled input-output data, having the measured output corrupted by colored noise. We concentrate on the continuous-time transfer function model identification. A Box-Jenkins model structure is used to describe the system, thus providing independent parameterizations for the plant and the noise. Monte Carlo simulation analysis is also used to illustrate the properties of the proposed estimation method.

Keywords: Continuous-time models, instrumental variable, non-uniform sampling, CARMA model estimation, system identification

## 1. INTRODUCTION

Non-uniform sampling is a common practice in scientific research, and the reasons for that can be manifold. One example is the Lebesgue sampling, where the sampler takes action only when a certain change in the signal has been detected. This leads to a non-uniformly distributed sampling time instants. Another reason to have irregular sampling is when, for example, a part of the data is lost. Typically, data is sampled equidistantly, but when we proceed to the recording stage, some can go missing unexpectedly. This phenomena also raises the issue of non-uniformly (irregularly) sampled data set.

In general, the discrete-time (DT) approach is not suited to handle non-uniformly sampled data. The reason is that the DT model parameter depends on the sampling period, see e.g Garnier and Wang (Eds.) [2008] and Chen et al. [2013]. It is, however, more natural to consider a continuous-time (CT) model when we encounter a sequence of non-uniformly sampled data.

The Box-Jenkins (BJ) model structure is a very well known model representation for dynamical systems. One of its main advantages is that the plant and the noise models (G and H, respectively) are independent (or equivalently, the denominators of the plant and the noise models have no common roots). In this paper, the system to be identified is assumed to have a BJ structure.

A vast literature is dedicated to the estimation of BJ models. One popular approach is the refined instrumental variable method for CT system identification (RIVC). In the RIVC algorithm [Young and Jakeman, 1980], which was first developed by Young and Jakeman considering that the additive noise is purely white in the open-loop situation. The extension of this initial algorithm to its hybrid form, where the process is modeled in CT while the noise is modeled in DT, was outlined in this earlier paper and was very recently developed and evaluated in Garnier et al. [2007], Young and Garnier [2006], Young et al. [2008]. However, the DT noise model assumption requires uniformly sampled data, thus the previous RIVC method cannot be extended to handle non-uniformly sampled data. In our previous work Chen et al. [2013], an CT approach was proposed to estimate a simplified CT BJ model, where the noise was regarded as a CT autoregressive (CAR) process. A shifted least-squares estimator [Larsson and Söderström, 2002] was used to estimate the noise model, it is computationally efficient and shows good convergence properties. Unfortunately, this method only considers a CAR noise model and cannot be extended to the CT autoregressive moving average (CARMA) noise model case.

As an extension to Chen et al. [2013], a CARMA noise model is considered in this paper. The difficulty to estimate CARMA models is that input (CT white noise) is missing. Thus, the widely used methods for deterministic system identification cannot be applied directly. However, if the second order property of the input is known, the missing data (e.g. state) can still be reconstructed by the Kalman filtering or smoothing, which can be used for parameter identification. Due to the fact that a CARMA model can be represented by a partially parameterized state-space model, we propose here to estimate the CARMA model in state-space form.

In Johansson [2009], identification of continuous-time state-space models, which are composed of an input-output model and a stochastic innovations model, from finite non-uniformly sampled input-output sequences was studied. Yuz et al. [2011] used the expectation-maximization (EM) algorithm to estimate a CT state space model from non-uniform fast sampled data, a so called incremental model was used to approximate a CT model. Dembo and Zeitouni [1986] also used the EM algorithm to estimate CT processes, but the state variables were estimated by a CT smoother. Our problem is a little bit different to Yuz et al. [2011], Dembo and Zeitouni [1986], because of the different noise assumptions. A maximum likelihood based method is used to solve the problem (see Section 4).

This paper is organized in the following way. We first define the parameter estimation problem in Section 2. In Section 3, the optimal IV estimator is recalled. Subsequently the CARMA noise estimation problem and the solution are described in Section 4. The optimal IV solution for identifying CT BJ models is summarized in Section 5. Finally, in Section 6, an example is presented to illustrate the properties of the proposed method.

## 2. PROBLEM STATEMENT

In this paper, the identification problem will be concentrated on the case of single-input, single-output system. So let us consider the following CT system

$$x(t) = G(p, \rho^o)u(t) = \frac{B(p, \rho^o)}{A(p, \rho^o)}u(t) \tag{1}$$

with

$$B(p, \rho^o) = b_0^o p^{n_b} + b_1^o p^{n_b - 1} + \cdots + b_{n_b}^o$$
$$A(p, \rho^o) = p^n + a_1^o p^{n_a - 1} + \cdots + a_{n_a}^o \quad (n_a > n_b)$$

where $u(t), x(t)$ are the excitation and the noise-free response. $p$ denotes the differentiation operator, i.e. $px = dx/dt$. It is further assumed that $B(p, \theta^o)$ and $A(p, \theta^o)$ are coprime.

In practical situations, the deterministic output $x(t)$ is inevitably corrupted by a colored noise $\xi(t)$, which is generated from the following process

$$\xi(t) = H(p, \eta^o) = \frac{C(p, \eta^o)}{D(p, \eta^o)}e(t) \tag{2}$$

with

$$C(p, \eta^o) = p^{n_c} + c_1^o p^{n_c - 1} + \cdots + c_{n_c}^o$$
$$D(p, \eta^o) = p^{n_d} + d_1^o p^{n_d - 1} + \cdots + d_{n_d}^o \quad (n_d > n_c)$$

where $e(t)$ is a CT Gaussian white noise has the covariance function of

$$\mathbb{E}\left\{e(t)e(\tau)\right\} = \sigma^2 \delta(t - \tau)$$

where $\delta$ is the Dirac's delta function. $H(p, \eta^o)$ is assumed to be stable and invertible stable.

The input and output signals $u(t)$, $y(t)$ are sampled instantaneously (see e.g. Wahlberg et al. [1993]) at irregular time instant $t_k$, for $k = 1, 2, \cdots, N$, and this gives rise to $u(t_k)$, $y(t_k)$. The time varying sampling interval is denoted as

$$h_k = t_{k+1} - t_k \tag{3}$$

Subsequently the data-generating system can be written in a more appropriate CT BJ form

$$\begin{cases} x(t) = G(p, \rho^o)u(t) \\ \xi(t) = H(p, \eta^o)e(t) \\ y(t_k) = x(t_k) + \xi(t_k) \end{cases} \tag{4}$$

It is assumed that there are no common factors in the plant $G(p, \rho^o)$ and the noise $H(p, \eta^o)$ components. Then, the parameter vector can be decomposed in the following way

$$\theta^o = [\rho^o;\ \eta^o] \tag{5}$$

where $\rho^o$ and $\eta^o$ include the dynamic plant model and the noise parameters stacked column-wise as

$$\rho^o = \begin{bmatrix} a_1^o & \cdots & a_{n_a}^o & b_0^o & \cdots & b_{n_b}^o \end{bmatrix}^T \in \mathbb{R}^{n_a + n_b + 1} \tag{6}$$
$$\eta^o = \begin{bmatrix} d_1^o & \cdots & d_{n_d}^o & c_1^o & \cdots & c_{n_c}^o \end{bmatrix}^T \in \mathbb{R}^{n_d + n_c} \tag{7}$$

The identification objective is then to estimate the parameters of the CT BJ model (4) from the non-uniformly sampled input and output data $Z^N = \{u(t_k); y(t_k)\}_{k=1}^N$.

## 3. OPTIMAL IV ESTIMATORS

In this section the main conditions for obtaining optimal (consistent and minimum variance) IV parameter estimate are recalled. Consider the general class of IV estimators

$$\hat{\rho} = \left\{ \text{sol}_\rho \frac{1}{N} \sum_{k=1}^{N} \zeta_f(t_k) \left[ y_f^{(n_a)}(t_k) - \varphi_f^T(t_k)\rho \right] = 0 \right\} \tag{8}$$

where the subscript $(\cdot)_f$ means the filtering operation, $(\cdot)_f = F(p)(\cdot)$. $\zeta(t_k)$ and $\varphi^T(t_k)$ are the instrument and the regressor.

$$\varphi^T(t_k) = \big[ -y^{(n_a-1)}(t_k) \ \cdots \ -y(t_k) \quad u^{(n_b)}(t_k) \ \cdots \ u(t_k) \big] \tag{9}$$

It has been shown that a minimum variance estimator is achieved under the following conditions [Young and Jakeman, 1980] (see also Söderström and Stoica [1983]):

$$\begin{cases} \zeta_f^{\text{opt}}(t_k) = F^{\text{opt}}(p)\mathring{\varphi}(t_k) \\ F^{\text{opt}}(p) = \dfrac{1}{H_o(p)A_o(p)} \end{cases} \tag{10}$$

where $\mathring{\varphi}(t_k)$ is the noise-free version of $\varphi(t_k)$

$$\mathring{\varphi}^T(t_k) = \big[ -x^{(n_a-1)}(t_k) \ \cdots \ -x(t_k) \quad u^{(n_b)}(t_k) \ \cdots \ u(t_k) \big] \tag{11}$$

## 4. NOISE MODEL ESTIMATION

In this section, identification of the CARMA noise model is considered. First we write (2) in the following alternative state space form (see Wahlberg et al. [1993], Goodwin et al. [2013])

$$\begin{cases} \dfrac{dz(t)}{dt} &= F_c z(t) + K_c e(t) \\ \xi(t) &= C z(t) \end{cases} \tag{12}$$

where $z(t)$ is a $n_d \times 1$-dimensional state variable and

$$F_c = \begin{bmatrix} -d_1 & \cdots & \cdots & -d_{n_d} \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & 1 & 0 \end{bmatrix}, K_c = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
$$C = [0 \ \cdots \ 0 \ 1 \ c_1 \ \cdots \ c_{n_c}]$$

Let $w(t) = K_c e(t)$ be process disturbance with covariance of

$$\mathbb{E}\left\{w(t)w^T(\tau)\right\} = Q_c \delta(t - \tau) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \delta(t - \tau) \tag{13}$$

Instantaneous sampling of (12) yields an equivalent DT state-space model (see e.g. Wahlberg et al. [1993])

$$\begin{cases} z(t_{k+1}) = F(h_k)z(t_k) + \tilde{w}(t_k) \\ \xi(t_k) = Cz(t_k) \end{cases} \tag{14}$$

with $F(h_k) = e^{F_c h_k}$, the covariance matrix of the DT noise term $\tilde{w}(t_k)$ is given by

$$\begin{aligned} \mathbb{E}\left\{\tilde{w}(t_k)\tilde{w}(t_i)^T\right\} &= Q(h_k)\delta_{k,i} \\ &= \int_0^{h_k} e^{F_c t} Q_c e^{F_c^T t} dt \; \delta_{k,i} \end{aligned} \tag{15}$$

where $\delta_{k,i}$ is the Kronecker's delta function. Note that even though $Q_c$ is singular, $Q(h_k)$ is still full rank [Goodwin et al., 2013]. For the sake of simplicity, let $F_k, Q_k, w_k, z_k$ and $\xi_k$ denote $F(h_k), Q(h_k), w(t_k), z(t_k)$ and $\xi(t_k)$, respectively.

In order to find out the connections between DT and CT parameters, expand $F_k$ to a Taylor series and omit the terms with the order higher than 1

$$F_k = I + F_k^{\delta} h_k \approx I + F_c h_k \tag{16}$$

then we have

$$\begin{cases} z_{k+1} - z_k = F_k^{\delta} h_k z_k + \tilde{w}_k \\ \xi_k = C z_k \end{cases} \tag{17}$$

It is necessary to point out that when $h_k$ goes to zero

$$\lim_{h_k \to 0} F_k^{\delta} = F_c, \; \lim_{h_k \to 0} \frac{Q_k}{h_k} = Q_c \tag{18}$$

### *4.1 Gradient-based parameter estimation scheme*

In this subsection, a maximum-likelihood method is introduced to estimate the CARMA noise model.

The log-likelihood function is defined as the logarithm of the conditional joint probability density function of $\xi_k$, which can be given as

$$l(\eta, Q_c) = \log p\left(\mathscr{Y}_N | \eta, Q_c\right) \tag{19}$$

where $\mathscr{Y}_k = \{\xi_1, \xi_2, \cdots, \xi_k\}$, $k = 1, \ldots, N$, denotes the observation. With this in mind, the maximum-likelihood (ML) estimate for $\eta, Q_c$ is defined as

$$(\hat{\eta}, \hat{Q}_c) = \arg \max_{\eta, Q_c} l(\eta, Q_c) \tag{20}$$

Numerical methods based on gradient search algorithms are the typical ways to solve (20).

By repeating the well-known Bayes' rule, $p(\mathscr{Y}_N|\eta, Q_c)$ can be decomposed as

$$\begin{aligned} p(\mathscr{Y}_N|\eta, Q_c) =& p(\xi_1|\eta, Q_c)p(\mathscr{Y}_{N-1}|\xi_1, \eta, Q_c) \\ =& p(\xi_1|\eta, Q_c)\prod_{k=2}^{N} p(\xi_k|\mathscr{Y}_{k-1}, \eta, Q_c) \end{aligned} \tag{21}$$

Consequently the log-likelihood function $l(\eta, Q_c)$ can be reformulated as

$$l(\eta, Q_c) = \sum_{k=2}^{N} \log p(\xi_k|\mathscr{Y}_{k-1}, \eta, Q_c) + \log p(\xi_1|\eta, Q_c) \tag{22}$$

where (22) contains conditional means. The Kalman filter (see e.g. Shumway and Stoffer [2011]) can be used to compute these terms.

It is well known, see e.g. Agüero et al. [2012], that the log-likelihood function (22) can be equivalently reformulated as

$$l(\eta, Q_c) = -\frac{1}{2}\sum_{k=1}^{N} \left(\varepsilon_k^T R_k^{-1} \varepsilon_k + \log \det R_k\right) + \text{constant} \tag{23}$$

where

$$\begin{aligned} \varepsilon_k &= \xi_k - \hat{\xi}_{k/k-1} \\ \hat{\xi}_{k/k-1} &= \mathbb{E}\{\xi_k|\mathscr{Y}_{k-1}, \eta, Q_c\} \\ R_k &= \mathbb{E}\{(\xi_k - \hat{\xi}_{k/k-1})(\xi_k - \hat{\xi}_{k/k-1})^T|\mathscr{Y}_{k-1}, \eta, Q_c\} \end{aligned}$$

can be obtained by using the Kalman Filter.

To maximize $l(\eta, Q_c)$, gradient methods can be applied. The partial derivative with respect to the $i$ th element of $\eta$ is given by

$$\begin{aligned} \frac{\partial l(\eta, Q_c)}{\partial \eta_i} =& -\frac{1}{2}\sum_{k=1}^{N}\left(2\frac{\partial \varepsilon_k^T}{\partial \eta_i} R_k^{-1}\varepsilon_k - \varepsilon_k^T R_k^{-1}\frac{\partial R_k}{\partial \eta_i} R_k^{-1}\varepsilon_k\right) \\ & -\frac{1}{2}\sum_{k=1}^{N} \text{Trace} R_k^{-1}\frac{\partial R_k}{\partial \eta_i} \end{aligned} \tag{24}$$

The derivative with respect to $Q_c$ can be similarly obtained. Thus, $\eta$ and $Q_c$ can be estimated in the following iterative way

$$(\hat{\eta}^{j+1}, \hat{Q}_c^{j+1}) = (\hat{\eta}^j, \hat{Q}_c^j) + \mu^j q^j \tag{25}$$

where $\mu^j$ is the step length that can be found by a line-search procedure. Notice that $\mu^j$ can be chosen such that $l(\hat{\eta}^{j+1}, \hat{Q}_c^{j+1}) \geq l(\hat{\eta}^j, \hat{Q}_c^j)$, and $q^j$ indicates the search direction that can be computed with an approximated Hessian.

### *4.2 Some modifications*

As can be seen from (24), the computation of the necessary derivatives are quite involved. Fortunately, there are alternatives ways to solve the optimization problem posed in (20). We split $\eta$ as

$$\eta = [\eta_D;\; \eta_C]$$

where $\eta_D$, $\eta_C$ denote the unknown parameters involved in $D(p, \eta)$ and $C(p, \eta)$, respectively. It turns out $\eta_C$ can easily be estimated when we froze the value for $\eta_D$ and $Q_c$. On the other hand, when we froze the value of $\eta_C$, the resulting estimation problem becomes more involved. To solve the latter problem, we will explore the use of the Expectation-Maximization (EM) algorithm as used in Yuz et al. [2011]. Notice that in Yuz et al. [2011], non-uniform fast sampled is data is used, but the system in (17) does not consider an output noise. Thus, the algorithm in Yuz et al. [2011] needs to be crafted to our particular problem.

One way to optimize the likelihood function (19) is by constructing a surrogate convex function. This is the principle of the EM algorithm, see e.g. Dempster et al. [1977], Dembo and Zeitouni [1986], Gibson and Ninness [2005], Shumway and Stoffer [2011], Agüero et al. [2012], where one optimizes the maximum likelihood function iteratively. Thus, the EM algorithm will generate a sequence of estimates $\left(\hat{\eta}_D^j, \hat{Q}_c^j\right)$, $j = 1, 2, \cdots$, of the parameters $(\eta_D, Q_c)$, which is guaranteed to converge to a local maximum of the log-likelihood function Dempster et al. [1977]. The basic idea is to use a *hidden* [1] variable, which in our case is taken to be $\mathscr{Z}_N := \{z_1, \ldots, z_N\}$. The complete data is given by $\mathscr{Z}_N$ since we have no measurement noise. Thus

$$\begin{aligned} \mathscr{Q}(\eta_D, Q_c, \hat{\eta}_D^j, \hat{Q}_c^j) &= \mathbb{E}\left\{\log p(\mathscr{Z}_N|\eta_D, Q_c)|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\} \\ &= \mathbb{E}\left\{l_c(\eta_D, Q_c)|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\} \end{aligned} \tag{26}$$

where $l_c(\eta_D, Q_c)$ can be expanded as follows using the Bayes' rule

[1] The terminology arises from the statistics literature.

$$l_c(\eta_D, Q_c) = \prod_{k=2}^{N} \log p(z_k|\mathscr{Z}_{k-1}, \eta_D, Q_c) + \log p(z_1|\eta_D, Q_c)$$

Because $z_k$ is a Markov process, the predictions for the future of the process are solely dependent on its present state, thus we have

$$l_c(\eta_D, Q_c) = \prod_{k=2}^{N} \log p(z_k|z_{k-1}, \eta_D, Q_c) + \log p(z_1|\eta_D, Q_c) \tag{27}$$

A new estimate of $\eta_D$ and $Q_c$, $\hat{\eta}_D^{j+1}$ and $\hat{Q}_c^{j+1}$, are then obtained by maximizing the function $\mathscr{Q}(\eta_D, Q_c, \hat{\eta}_D^j, \hat{Q}_c^j)$, that is,

$$(\hat{\eta}_D^{j+1}, \hat{Q}_c^{j+1}) = \arg \max_{\eta_D, Q_c} \mathscr{Q}(\eta_D, Q_c, \hat{\eta}_D^j, \hat{Q}_c^j) \tag{28}$$

Expressions in (26) and (28) are known as the E-step and M-step of the EM algorithm.

Considering the above, we suggest to estimate $\eta_D$ and $\eta_C$ separately, meaning we froze one of the parameters while we estimate the other one, and vice versa. This procedure is detailed below.

Provided that

$$z_{k+1} \sim \mathscr{N}(F_k z_k, Q_k) \tag{29}$$
$$z_1 \sim \mathscr{N}(\mu_1, P_1) \tag{30}$$

we have

$$\begin{aligned} &-2\mathscr{Q}(\eta_D, Q_c, \hat{\eta}_D^j, \hat{Q}_c^j) \\ =&\log \det P_1 + \sum_{k=2}^{N} \log \det Q_k \\ &+ \text{Trace}\, \mathbb{E}\left\{P_1^{-1}(z_1 - \mu_1)(z_1 - \mu_1)^T|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\} \\ &+ \sum_{k=2}^{N} \text{Trace}\, \mathbb{E}\Big\{{Q_k}^{-1}(z_{k+1} - F_k z_k) \\ &\times (z_{k+1} - F_k z_k)^T|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\Big\} + \text{constant} \end{aligned} \tag{31}$$

In (31), $x^T Ax$=Trace$\{A^T xx^T\}$ is used.

If the following terms are obtained by a Rauch–Tung–Striebel (RTS) smoother (see e.g. Shumway and Stoffer [2011])

$$z_{k/N} = \mathbb{E}\left\{z_k|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\}$$
$$P_{k/N} = \mathbb{E}\left\{\left(z_k - z_{k/N}\right)\left(z_k - z_{k/N}\right)^T|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\}$$
$$S_{k/N} = \mathbb{E}\left\{\left(z_k - z_{k/N}\right)\left(z_{k-1} - z_{k-1/N}\right)^T|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\}$$

Then

$$\mathbb{E}\left\{z_k z_k^T|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\} = z_{k/N} z_{k/N}^T + P_{k/N} \tag{32}$$
$$\mathbb{E}\left\{z_k z_{k-1}^T|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\} = z_{k/N} z_{k-1/N}^T + S_{k/N} \tag{33}$$

If $F_k = I + F_c h_k, Q_k = Q_c h_k$ is assumed, differentiate $\mathscr{Q}(\eta_D, Q_c, \hat{\eta}_D^j, \hat{Q}_c^j)$ with respect to $F_c, Q_c$ and set the derivative to zero, we have (see Yuz et al. [2011])

$$F_c = \Psi\Gamma^{-1} \tag{34}$$
$$Q_c = \left(\Phi - \Psi\Gamma^{-1}\Psi^T\right)/(N-1) \tag{35}$$

where

$$\Psi = \mathbb{E}\left\{\sum_{k=1}^{N-1}\left(z_{k+1}z_k^T - z_k z_k^T\right)|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\}$$
$$\Gamma = \mathbb{E}\left\{\sum_{k=1}^{N-1} z_k z_k^T h_k|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\}$$
$$\Phi = \mathbb{E}\left\{\sum_{k=1}^{N-1}\frac{(z_{k+1} - z_k)(z_{k+1} - z_k)^T}{h_k}\bigg|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{Q}_c^j\right\}$$

### 4.3 *The final algorithm for noise modeling*

The algorithm to estimate the CARMA noise model is composed of two stages, namely:

(i) We estimate $\eta_C$ by the traditional gradient algorithm with frozen $\eta_D$, and choosing the step size $\mu^j$ such that the log-likelihood function increases,
(ii) We estimate $\eta_D$ using the EM algorithm with frozen $\eta_C$.

---

*Algorithm 1: The combined Gradient and EM algorithms for noise modeling*

**Step 1. Initialization:** Set the initial guess $\hat{\eta}^0 = \left\{\hat{\eta}_D^0, \hat{\eta}_C^0\right\}, \hat{Q}_c^0$
**Step 2. Iteration:**
*for j=1:convergence*
(1) Estimate $\eta_C$ using numerical search method

$$\hat{\eta}_C^{j+1} = \hat{\eta}_C^j + \mu^j \bar{q}^j \tag{36}$$

where

$$\bar{q}^j = \left(\frac{\partial^2 l}{\partial\eta_C \partial\eta_C^T}\bigg|_{\eta_C^j, \hat{Q}_c^j}\right)^{-1} \frac{\partial l}{\partial\eta_C}\bigg|_{\eta_C^j, \hat{Q}_c^j} \tag{37}$$

(2) Estimate $\eta_D$ using EM algorithm
- E-step

$$\begin{aligned} &\mathscr{Q}(\eta_D, Q_c, \hat{\eta}_D^j, \hat{Q}_c^j) \\ =&\mathbb{E}\left\{l_c(\eta_D, Q_c)|\mathscr{Y}_N, \hat{\eta}_D^j, \hat{\eta}_C^{j+1}\right\} \end{aligned}$$

- M-step

$$\left(\hat{\eta}_D^{j+1}, \hat{Q}_c^{j+1}\right) = \arg \max_{\eta_D, Q_c} \mathscr{Q}(\eta_D, Q_c, \hat{\eta}_D^j, \hat{Q}_c^j)$$

(3) Form $\hat{\eta}^{j+1} = \left\{\hat{\eta}_D^{j+1}, \hat{\eta}_C^{j+1}\right\}$, then return to (1)
*end*

---

## 5. RIVC METHOD FOR CT BJ MODELS

The complete RIVC algorithm is summarized below.

---

*Algorithm 2: The RIVC algorithm for CT BJ model identification*

**Step 1. Initialization**: Specify $\hat{\eta}^0, \hat{Q}_c^0$. Apply the SRIVC algorithm [Chen et al., 2013, Young et al., 2008] to compute $\hat{\rho}^0$.
**Step 2. Iterative IV estimation with prefilters**:
$for\ j = 1 : convergence$
(1) If the estimated plant model of $(j-1)$th iteration is unstable, reflect the unstable zeros of the estimated $\hat{A}(p, \hat{\rho}^{j-1})$ polynomial into the stable region of the

| | $\hat{a}_1$ | $\hat{a}_2$ | $\hat{b}_0$ | $\hat{b}_1$ | $\hat{d}_1$ | $\hat{d}_2$ | $\hat{c}_1$ | $\hat{\sigma}^2$ | $N_{iter}$ |
|---|---|---|---|---|---|---|---|---|---|
| True value | 2.8 | 4 | 1 | 3 | 1 | 2 | 5 | $1.8 \times 10^{-3}$ | |
| SRIVC | 2.8006 ±0.1588 | 3.9849 ±0.3405 | 1.0017 ±0.0319 | 2.9961 ±0.2644 | —— | —— | —— | —— | 3.7 |
| RIVC | 2.8032 ±0.0918 | 4.0075 ±0.1472 | 1.0001 ±0.0083 | 3.0079 ±0.1277 | 1.0632 ±0.1319 | 1.9657 ±0.1061 | 5.0888 ±0.3253 | $2.1 \times 10^{-3}$ $\pm 7.5 \times 10^{-5}$ | 12.1 |

Table 1. Mean and standard deviation of the estimated parameters. $\hat{\sigma}^2$ - Estimated intensity of the CT white noise. $N_{iter}$ - Number of iterations for convergence.

complex plane. Generate the noise response from the system 'auxiliary model'

$$\hat{x}(t_k, \hat{\rho}^{j-1}) = \hat{G}(p, \hat{\rho}^{j-1})u(t_k)$$

(2) Obtain the latest estimate $\hat{\eta}^j$ of the CARMA noise model parameters based on the estimated noise sequence

$$\hat{\xi}(t_k) = y(t_k) - \hat{x}(t_k, \hat{\rho}^{j-1}) \tag{38}$$

using the combined Gradient and EM algorithms (see section 4).

(3) Prefilter $u(t_k)$, $y(t_k)$ and $\hat{x}(t_k, \hat{\rho}^{j-1})$ signals by the filter

$$F(p, \hat{\rho}^{j-1}, \hat{\eta}^j) = \frac{1}{\hat{A}(p, \hat{\rho}^{j-1})\hat{H}(p, \hat{\eta}^j)}$$

(4) Based on these prefiltered data, compute the estimate $\hat{\rho}^j$ of the plant model parameter vector from

$$\hat{\rho}^j = \left[\sum_{k=1}^{N} \zeta_f(t_k, \hat{\rho}^{j-1}, \hat{\eta}^j)\varphi_f^T(t_k, \hat{\rho}^{j-1}, \hat{\eta}^j)\right]^{-1} \sum_{k=1}^{N} \zeta_f(t_k, \hat{\rho}^{j-1}, \hat{\eta}^j) y_f^{(n_a)}(t_k, \hat{\rho}^{j-1}, \hat{\eta}^j) \tag{39}$$

where $\zeta_f(t_k, \hat{\rho}^{j-1}, \hat{\eta}^j)$, $\varphi_f(t_k, \hat{\rho}^{j-1}, \hat{\eta}^j)$ are given in (11) and (9) but where a dependency to the CARMA noise model parameter estimates $\hat{\eta}^j$ is made clear.

*end*

*Remark*: When computing $y_f^{(n_a)}(t_k) = p^{n_a}F(p)y(t_k)$, the degree of numerator is larger than the degree of denominator in $p^{n_a}L(p)$. This implies that pure time-derivatives of the signal are computed. In this case, we approximate the time-derivatives by difference quotient.

## 6. NUMERICAL EXAMPLES

Consider now the following continuous-time system

$$\begin{cases} x(t) &= \dfrac{p+3}{p^2 + 2.8p + 4}u(t) \\ \xi(t) &= \dfrac{p+5}{p^2 + p + 2}e(t) \\ y(t_k) &= x(t_k) + \xi(t_k) \end{cases} \tag{40}$$

$u(t)$ is chosen to be a pseudo-random binary sequence (PRBS) that is generated from a shift register with 9 stages. The shift clock of the shift register is assumed to be 1 sec.

The estimation procedure is carried out under the following conditions:

- The length of the sampled data is $N = 10,000$.
- The intensity of $e(t)$ is set to $\sigma^2 = 1.8 \times 10^{-3}$, thus the signal to noise ratio (SNR) is 15dB approximately.
- Monte Carlo simulations (MCS) of 100 runs are performed.
- The sampling period $h_k$ is uniformly distributed over the following interval.

$$h_k \sim U(0.01, 0.09) \tag{41}$$

- When using the SRIVC algorithm in step 1 (see Algorithm 2 above), we need to specify the cutoff frequency $\omega_c^{\rm SVF}$ of a state variable filter (see e.g. Young et al. [2008]). This value is set to $\omega_c^{\rm SVF} = 0.6\pi$.
- The initial values of $d_1, d_2, c_1$ are random numbers uniformly distributed in the following intervals, the initial guess of $Q_c$ is assumed to be a fixed value.

$$\hat{d}_1^0, \hat{d}_2^0 \sim U(0,3), \; \hat{c}_1^0 \sim U(0,10), \; \hat{Q}_c^0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{42}$$

- The stopping rule is chosen to be

$$\max\left\{\left|\frac{\hat{\rho}^{i+1} - \hat{\rho}^i}{\hat{\rho}^i}\right|\right\} \leq 10^{-6} \tag{43}$$

Figure 1 shows a portion of the measured input-output data where the time-varying sampling can be observed.

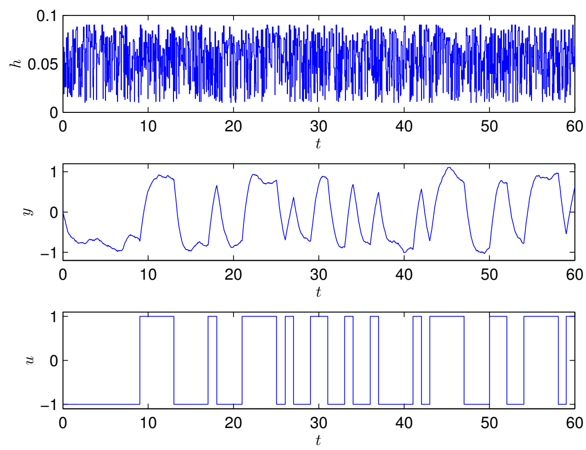


Fig. 1. A portion of the sampled data. Top - The sampling period. Middle - The sampled output. Bottom - The sampled input.

The estimated parameters are presented in Table 1. The SRIVC method from the CONTSID toolbox [2] is run here to give a comparison. Even though the additive noise is colored, in the estimation procedure, simpler SRIVC assumes the noise to be white. As can be seen from this table, both the SRIVC and the RIVC methods are unbiased. However, because the noise model is mis-specified, the SRIVC estimates have larger variance in estimated parameters. The proposed RIVC algorithm takes the merits of the noise modeling, so it reduces the variance of the estimates. However, it is at the expense of increasing computational load, as can be seen from this table, the RIVC

[2] see http://www.iris.cran.uhp-nancy.fr/contsid/

method needs on average 12 iterations to converge. The Bode plots of the 100 estimated RIVC models presented in Table 1 are also available in Fig.2 and Fig.3.

The gradient-based estimation procedure is known to converge to local minima. In the chosen example and initializations, this problem is not very significant, for the reason that only one parameter is estimated by the numerical search. All of estimated models have been used to compute the mean and standard deviation.

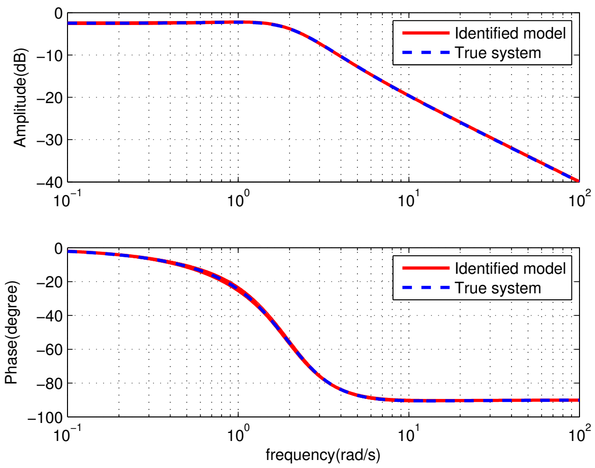


Fig. 2. Bode plots of the estimated plant models with the true system

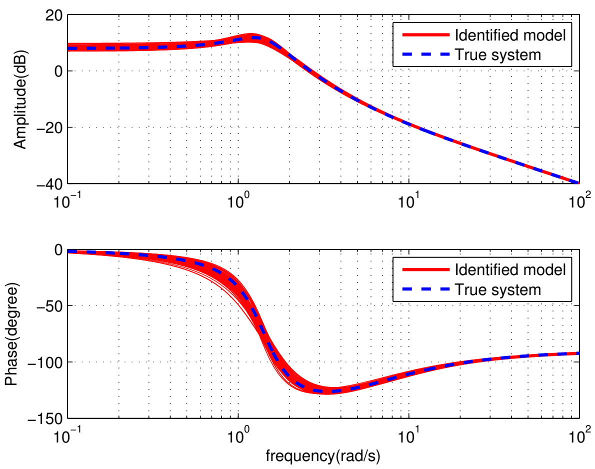


Fig. 3. Bode plots of the estimated noise models with the true system

## 7. CONCLUSION

The issue of continuous-time Box-Jenkins model identification from non-uniformly sampled data has been considered in this paper. Refined instrumental variable-based method has been developed to solve this problem, with CARMA noise estimation. The performance of the proposed scheme has been investigated by means of simulation examples.

## REFERENCES


H. Garnier, M. Gilson, P.C. Young, and E. Huselstein. An optimal IV technique for identifying continuous-time transfer function model of multiple input systems. *Control Engineering Practice*, 46(15):471–486, 2007.

H. Garnier and L. Wang (Eds.). *Identification of continuous-time models from sampled data*. Springer-Verlag, London, 2008.

R. Johansson. Continuous-time model identification and state estimation using non-uniformly sampled data. *15th IFAC Symposium on System Identification*, Saint-Malo, France, July 6-8, 2009

J. Nocedal and S. Wright. *Numerical Optimization (Second Edition)*. Springer Science+Business Media, New York, 2006.

E.K. Larsson and T. Söderström. Identification of continuous-time AR processes from unevenly sampled data. *Automatica*, 38(4):709–718, 2002.

L. Ljung and A. Wills. Issues in sampling and estimating continuous-time models with stochastic disturbances. *Automatica*, 46:925–931, 2010.

A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

T. Söderström and P. Stoica. *Instrumental variable methods for system identification*. Springer Verlag, New York, 1983.

B. Wahlberg, L. Ljung, and T. Söderström. On sampling of continuous-time stochastic processes. *Control Theory and Advanced Technology*, 9(1):99–112, 1993.

P.C. Young and H. Garnier. Identification and estimation of continuous-time data-based mechanistic (DBM) models for environmental systems. *Environmental Modelling and Software*, 21(8):1055–1072, August 2006.

P.C. Young and A.J. Jakeman. Refined instrumental variable methods of time-series analysis: Part III, extensions. *International Journal of Control*, 31:741–764, 1980.

P.C. Young, H. Garnier, and M. Gilson. Refined instrumental variable identification of continuous-time hybrid Box-Jenkins models. In *Identification of continuous-time models from sampled data*, H. Garnier and L. Wang (Eds.), 91–132, Springer-Verlag, London, 2008.

J.I. Yuz, J. Alfaro, J.C. Agüero, G.C. Goodwin. Identification of continuous-time state-space models from non-uniform fast-sampled data. *IET Control Theory and Applications*, 5(7): 842–855, 2011.

A. Dembo and O. Zeitouni. Parameter estimation of partially observed continuous-time stochastic processes via the EM algorithm. *Stochastic Processes and their Applications*, 23: 91–113, 1986.

J.C. Agüero, W. Tang, Juan I. Yuz, R. Delgado, G.C. Goodwin. Dual time–frequency domain system identification. *Automatica*, 483031–3041, 2012.

G.C. Goodwin, J.C. Agüero, M.E. Cea Garrido, M.E. Salgado, and J.I. Yuz. Sampling and sampled-data models. *IEEE Control Systems Magazine*, 34–53, October, 2013.

S. Gibson, B. Ninness. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica*, 41:1667–1682, 2005.

R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications*, Springer, New York, 2011.

F. Chen, H. Garnier and M. Gilson. Refined instrumental variable identification of continuous-time OE and BJ models from irregularly sampled data. *11th IFAC International Workshop on Adaptation and Learning in Control and Signal Processing (ALCOSP'2013)*, Caen, France, 3-5 July 2013.