

Identification with Stochastic Sampling Time Jitter[★]

Frida Eng^{1,2}, Fredrik Gustafsson².

Dept of EE, Linköpings universitet, SE-58183 Linköping, Sweden

Abstract

This work investigates how stochastic sampling jitter noise affects the result of system identification, and proposes a modification of known approaches to mitigate the effects of sampling jitter, when the jitter is unknown and not directly measurable. By just assuming conventional additive measurement noise, the analysis shows that the identified model will get a bias in the transfer function amplitude that increases for higher frequencies. A frequency domain approach with a continuous-time model allows an analysis framework for sampling jitter noise. The bias and covariance in the frequency domain model are derived. These are used in bias compensated (weighted) least squares algorithms, and by asymptotic arguments this leads to a maximum likelihood algorithm. Continuous-time output error models are used for numerical illustrations.

Key words: Non-uniform sampling, Sampling jitter, System identification, Stochastic systems, Maximum likelihood, Least squares estimation, Frequency domain, Parametric model

1 Introduction

Consider a deterministic signal model $s(t; \theta)$, which may depend on an observed or known input. This work studies the problem of identifying the unknown parameter vector θ , when the discrete time observations y_k requested at time $t = kT$ (T denotes the sampling interval) are subject both to the usual additive measurement noise and also stochastic unmeasurable jitter noise τ_k as part of the sampling process. That is, the observation includes the term $s(kT + \tau_k; \theta)$, which becomes a stochastic variable.

This type of non-uniform sampling may occur when uniform sampling is requested, but the sensor for one or several reasons cannot measure exactly at that time instant, and the true sampling time is either unmeasurable, or the communication protocol does not allow to transport time stamps to each measurement. Sampling jitter may also occur due to imperfect hold circuits, synchronization or other hardware problems. Not even high-

performance digital oscilloscopes are free from sampling jitter as demonstrated in Verbeyst et al. [2006]. There, a dedicated system identification experiment is developed to estimate jitter effects. The result, when a sampling time of 1.22 ps is used, is that a commercial sampling oscilloscope has a sampling jitter standard deviation of around 1 ps, that is $\approx 80\%$ of the sampling time.

The general problem of non-uniform sampling is extensively treated in literature, see Bilinskis and Mikelsons [1992] and Marvasti [2001]. In most publications, the sampling times are known, and the problem is to analyze leakage and alias effects. Another twist is to design sampling times to minimize aliasing. For stochastic sampling jitter, the distribution of $s(t + \tau_k)$ is derived in Eng and Gustafsson [2005, 2006] and Souders et al. [1990]. These results will be used and extended in this paper. In the context of jitter estimation, the sampling oscilloscope has been extensively studied, the methods include averaging over several measurements, as in Verspecht [1995], and Taylor series expansions, in Verbeyst et al. [2006], in order to estimate the variance of the jitter and use this to compensate for the jitter effects. In this paper, we consider a model based approach, where the main goal is to find an underlying continuous-time system, and the jitter density can be seen as known or given in the estimation as a bi-product.

The basic idea is as follows. The frequency domain approach is to minimize the distance between the measurement and signal model discrete time Fourier transforms

[★] Parts of this work were presented at The IFAC World Congress, 2005, in Prague, and at The IFAC Symposium on System Identification, 2006, in Newcastle.

Email addresses: frida@isy.liu.se (Frida Eng), fredrik@isy.liu.se (Fredrik Gustafsson).

¹ Corresponding author: Tel. +46 13 281365. Fax +46 13 282622.

² Both authors are with the Automatic Control group, Department of Electrical Engineering at Linköping University

(DTFT), $\| \text{DTFT}(y_k) - \text{DTFT}(s(kT + \tau_k; \theta)) \|$ with respect to the parameters θ in some suitable norm. This frequency domain approach is standard in system identification [Ljung, 1999, Pintelon and Schoukens, 2001]. A continuous-time model for $s(t)$ is used to be able to analyze the sampling jitter effects. The analysis shows that by neglecting the jitter, the Fourier transform (FT) of the signal model will suffer from an amplitude bias in $|\text{FT}(s(t))| = |S(f)|$ that increases for higher frequencies. Further, the larger jitter noise variance, the larger bias. The remedy is to compensate for the bias, and the closed form solution involves a frequency weighting in the norm, $\| \text{DTFT}(y_k) - \int S(\psi; \theta) w(f, \psi) d\psi \|$.

The outline is as follows. The system identification problem and main notation are presented in Section 2. In Section 3, the bias effect of sampling jitter noise on the frequency transform is derived, and the bias compensated least squares estimator is proposed. Section 4 derives the second order properties of the frequency transform due to jitter noise, and a weighted least squares algorithm as well as an asymptotic maximum likelihood estimator are presented. Section 5 illustrates these algorithms for several simulated numerical examples. The work is concluded in Section 6.

2 Problem Formulation

The general problem formulation looks as follows. The sensor is requested to sample uniformly, but delivers discrete time measurements corrupted by amplitude noise and sampling time jitter according to

$$y_k = s(kT + \tau_k; \theta) + v(kT + \tau_k; \theta). \quad (1a)$$

The signal term, noise term and jitter distribution can all be dependent on the unknown parameter θ and given by

$$s(t; \theta) = (g_\theta \star u)(t), \quad (1b)$$

$$v(t; \theta) = (h_\theta \star e)(t), \quad (1c)$$

$$\tau_k \in p_\theta(\tau). \quad (1d)$$

Here $u(t)$ is a known input, $e(t)$ is white noise with known characteristics, $g_\theta(t)$ denotes the system impulse response and $h_\theta(t)$ the noise dynamics. The jitter sampling noise is a sequence of independent stochastic variables with probability density function (pdf) $p_\theta(\tau)$. Both the signal, noise and sampling models can be parameterized in the unknown parameter vector θ . We will primarily focus on continuous-time systems here.

The system identification problems under consideration can be stated as estimating the parameter θ in a model

structure

$$\mathcal{M}_{OE} : g_\theta(t), h_\theta(t) = \delta(t), p_\theta(\tau) = p(\tau), \quad (2a)$$

$$\mathcal{M}_{BJ} : g_\theta(t), h_\theta(t), p_\theta(\tau) = p(\tau), \quad (2b)$$

$$\mathcal{M}_{JOE} : g_\theta(t), h_\theta(t) = \delta(t), p_\theta(\tau), \quad (2c)$$

$$\mathcal{M}_{JBJ} : g_\theta(t), h_\theta(t), p_\theta(\tau). \quad (2d)$$

Here, OE denotes the output error and BJ the Box-Jenkins model structure, respectively, where the jitter distribution is known. JOE and JBJ are the corresponding problems where also the jitter noise distribution is parameterized.

Using previous knowledge about the sampling jitter effect in the frequency domain indicates that the frequency domain approach [see for example Ljung, 1999, Pintelon and Schoukens, 2001] is suitable for identification in this case. Denote the Fourier transform of the measurements and signal model, respectively,

$$Y_d(f) = \sum_{k=0}^{N-1} y_k e^{-i2\pi f k T}, \quad (3)$$

$$S(f; \theta) = G(f; \theta) U(f). \quad (4)$$

The general problem formulation is now to minimize the distance between the measurement, $Y_d(f)$, and model, $S(f; \theta)$, in the frequency domain.

$$\hat{\theta} = \arg \min_{\theta} \int_{-\infty}^{\infty} \lambda(f; \theta) |Y_d(f) - S(f; \theta)|^2 df, \quad (5)$$

for some suitable weighting function, λ . Normally, the weights, $\lambda(f; \theta)$ are given by the inverse noise spectrum [Ljung, 1999, Pintelon and Schoukens, 2001]. We will show a few other examples later.

Given a continuous-time signal model $S_c(f)$, a well known property of the Fourier transform gives that the discrete Fourier transform, $S_d(f)$, becomes

$$S_d(f) = \int_{-\infty}^{\infty} S_c(\psi) d_N(f - \psi) d\psi. \quad (6)$$

Here $d_N(f)$ is the normalized Dirichlet kernel (also known as the aliased sinc function), defined as

$$d_N(f) = e^{-i\pi f(N-1)T} \frac{\sin(\pi f N T)}{\sin(\pi f T)}. \quad (7)$$

The local behavior of the normalized Dirichlet kernel (see Figure 1 on p. 4) describes the effects of leakage and its $1/T$ periodicity describes aliasing. For the regular case, with no sampling jitter, the correct way is to compare $Y_d(f)$ with $S_d(f; \theta)$, and using the unweighted

least squares (LS) norm in (5) over a discrete set of frequencies yields the parameter vector θ as

$$\hat{\theta}^{\text{LS}} = \arg \min_{\theta} \sum_f |Y_d(f) - \int_{-\infty}^{\infty} S_c(\psi; \theta) d_N(f - \psi) d\psi|^2. \quad (8)$$

Similarly to Pintelon and Schoukens [2005] and Gillberg and Gustafsson [2005], the frequency set under consideration is arbitrary, in order to have the freedom to highlight certain frequency regions of interest.

This problem is closely related to the approach in Gillberg and Ljung [2005a,b] for the special case of no jitter noise, and in Gillberg and Gustafsson [2005] for the case of non-uniform sampling (but deterministic known sampling times). Note also that the leakage can also be interpreted as originating from a transient in the signal model caused by an unknown initial state. An alternative dual approach based on estimating this initial state is studied in Schoukens et al. [2005] and Pintelon and Schoukens [2001].

The main idea with bias compensation is to mitigate the bias effects in $Y_d(f)$ by replacing the deterministic discrete time signal model $\int S_c(\psi; \theta) d_N(f - \psi) d\psi$ in (8) with a term, that besides sampling effects also includes the jitter noise. This term is denoted $\mu_Y(f)$, and is derived from $E[Y_d(f)]$. The purpose of this work is to analyze the bias and variance contribution from jitter sampling noise on the signal model $S_c(f; \theta)$, and modify the criterion (8) accordingly.

3 Bias Compensation

First, the bias in $Y_d(f)$ due to sampling jitter is derived. Then, the bias compensated LS (BCLS) estimate is defined. Equation (8) is here replaced with the more general expression

$$\hat{\theta} = \arg \min_{\theta} \sum_f |Y_d(f) - \mu_Y(f; \theta)|^2, \quad (9)$$

where $\mu_Y(f; \theta)$ is to be derived. In the following two subsections, the dependence on the parameter vector θ will be implicit.

3.1 Bias in $Y_d(f)$

The DTFT (3) of the sequence y_k using the sampling jitter model (1) can be written

$$\begin{aligned} Y_d(f) &= \sum_{k=0}^{N-1} y_k e^{-i2\pi f k T} \\ &= \sum_{k=0}^{N-1} (s(kT + \tau_k) + v_k) e^{-i2\pi f k T} \\ &= \hat{S}_d(f) + \underbrace{\sum_{k=0}^{N-1} v_k e^{-i2\pi f k T}}_{\hat{V}(f)}. \end{aligned} \quad (10)$$

We will restrict this investigation to OE structures and therefore $v_k = v(kT + \tau_k) = e(kT + \tau_k)$ are zero-mean independent stochastic variables. Using the inverse FT $s(t) = \int_{-\infty}^{\infty} S_c(f) e^{i2\pi f t} df$, the signal term becomes

$$\begin{aligned} \hat{S}_d(f) &= \sum_{k=0}^{N-1} s(kT + \tau_k) e^{-i2\pi f k T} \\ &= \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} S_c(\psi) e^{i2\pi\psi(kT + \tau_k)} d\psi e^{-i2\pi f k T} \\ &= \int_{-\infty}^{\infty} S_c(\psi) W(f, \psi) d\psi, \end{aligned} \quad (11)$$

where

$$W(f, \psi) \triangleq \sum_{k=0}^{N-1} e^{i2\pi(\psi - f)kT} e^{i2\pi\psi\tau_k}. \quad (12)$$

The term W is a stochastic frequency window, completely describing the jitter, leakage and aliasing effects in the signal term $S_d(f)$ of $Y_d(f)$. Note that $W(f, \psi) = d_N(f - \psi)$ in the special case of no jitter, $\tau_k = 0$. It should be stressed that this continuous-time frequency domain approach is perhaps the only way to explicitly separate the signal amplitude noise, captured in \hat{V} , and sampling jitter noise, captured in W , in a model-based approach.

3.2 First Moment of W

The moments of the transform $Y_d(f)$ have an explicit dependence on the signal transform $S_c(f)$. The mean value of $Y_d(f)$ in (3) is

$$\begin{aligned} \mu_Y(f) &= E[Y_d(f)] = E[\hat{S}_d(f)] \\ &= \int_{-\infty}^{\infty} S_c(\psi) E[W(f, \psi)] d\psi \\ &= \int_{-\infty}^{\infty} S_c(\psi) \mu_W(f, \psi) d\psi \end{aligned} \quad (13)$$

and $\mu_W = E[W]$ is given by the following lemma.

Lemma 1 (Mean Value of W) *The mean value, μ_W , with respect to τ_k , of the stochastic window, $W(f, \psi)$, defined in (12), is*

$$\mu_W(f, \psi) = \varphi(-\psi)d_N(f - \psi),$$

where $\varphi(f)$ is the characteristic function³, $\varphi(f) = E[e^{-i2\pi f\tau}]$, of τ and $d_N(f)$ is the normalized Dirichlet kernel (7).

PROOF.

$$\begin{aligned} \mu_W(f, \psi) &= E[W(f, \psi)] = \sum_{k=0}^{N-1} e^{i2\pi(\psi-f)kT} E[e^{i2\pi\psi\tau_k}] \\ &= \varphi(-\psi) \underbrace{\frac{1 - e^{i2\pi(\psi-f)NT}}{1 - e^{i2\pi(\psi-f)T}}}_{d_N(f-\psi)}, \end{aligned}$$

where we recognize the characteristic function, $\varphi(f) = E[e^{-i2\pi f\tau}]$ and the normalized Dirichlet kernel, d_N .

The characteristic function $\varphi(f) = \varphi(f; \theta)$ models damping corresponding to the sampling jitter noise. It will include possible distribution parameters from $p_\theta(\tau)$ in (2). Its properties include [see for example, Papoulis, 1991]:

- The definition gives, $|\varphi(f)| \leq \varphi(0) = 1, \forall f$.
- From the series expansion we know that if $\varphi(f) = 1$, for all f , all moments $E[(\tau_k)^n] = 0$, so there is no jitter noise,
- The series expansion also shows that the decay rate of $|\varphi(f)|$ increases with the jitter noise variance, given a class of distributions and zero-mean jitter.

The second property implies that $\mu_W(f, \psi) = d_N(f - \psi)$ for the case of no jitter noise.

3.3 Bias Compensated Least Squares Estimate

Since now $\mu_Y(f) = E[Y_d(f)]$ is completely known, we can compensate for both the leakage and jitter effect on the measurement DTFT. We include this in the parameter estimation as, cf. (9),

$$\hat{\theta}^{\text{BCLS}} = \arg \min_{\theta} \sum_f |Y_d(f) - \mu_Y(f; \theta)|^2 \quad (14a)$$

³ The usual definition from probability theory, $\varphi(u) = E[e^{iu\tau}]$, is slightly modified for notational convenience.

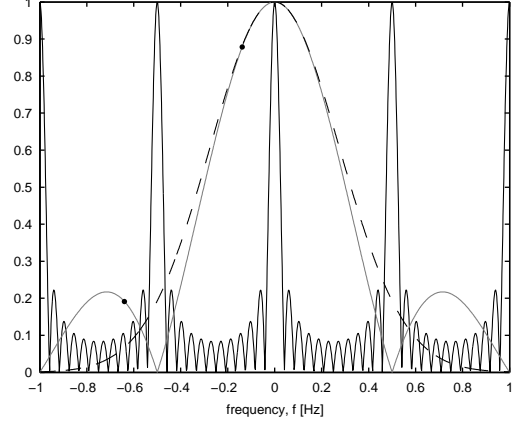


Fig. 1. Example of the normalized Dirichlet kernel (amplitude), $|d_N(f)|$ (thick), (7) with $T = 2$, and the two characteristic functions, $\varphi_\tau(f)$ in (15), (a) thin and (b) dashed.

where

$$\mu_Y(f; \theta) = \int_{-\infty}^{\infty} S_c(\psi; \theta) \varphi(-\psi; \theta) d_N(f - \psi) d\psi. \quad (14b)$$

Both the signal $S_c(f; \theta)$ and jitter pdf $p(\tau; \theta)$, and thus the characteristic function $\varphi(f; \theta)$, may depend on the unknown parameter vector θ .

3.4 Illustrations

Let us consider the implications of Lemma 1 in more detail. The Dirichlet kernel that appears as the window function in the jitter-free case is illustrated in Figure 1 (solid black). As two concrete examples of characteristic functions, consider

$$\tau \in U(-1, 1) \quad \Rightarrow \quad \varphi(f) = \frac{\sin(2\pi f)}{2\pi f} \quad (15a)$$

$$\tau \in N(0, 1/3) \quad \Rightarrow \quad \varphi(f) = e^{-2\pi^2 f^2/3}, \quad (15b)$$

which have the same variance. The amplitudes of these functions are also illustrated in Figure 1.

According to Lemma 1, the jitter noise will attenuate the amplitude of $Y_d(f)$. The attenuation is inversely linear in frequency for uniform jitter and exponentially decaying for Gaussian jitter noise. This also implies that frequencies above the Nyquist frequency are attenuated, suppressing alias. This attenuation is compensated for in the BCLS estimate.

To illustrate the jitter effects we use an academic example. Consider a single frequency signal with known amplitude and phase, but unknown frequency, f_0 , $s(t) = \sin(2\pi f_0 t)$, such that $S_c(f) = (\delta(f - f_0) - \delta(f + f_0))/2i$. We compare uniform sampling, $T = 2$, and jitter sampling, $\tau \in U(-T/2, T/2)$, for two different frequencies

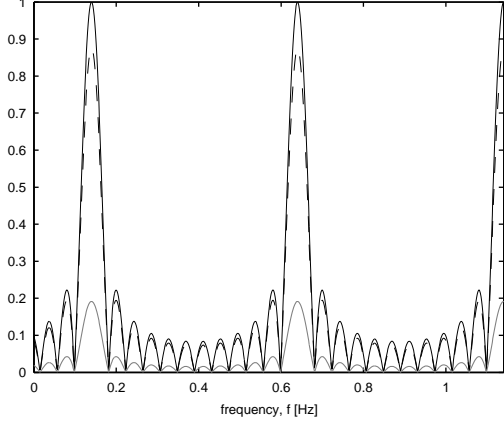


Fig. 2. $|E[Y_d(f)]|$ given in (16) are shown, (a) and (b) (thick line) overlap, while (c) (thin) and (d) (dashed) are different. The sampling jitter noise in the last two cases is given by (15a), and the damping could be found directly by studying $\varphi_\tau(f)$ in Figure 1, note the two dots.

f_0 , and get the following expressions on the positive frequency axis. Uniform sampling gives

$$Y_d(f) = \frac{1}{2i} d_N(f - 0.14), \quad f_0 = 0.14 \quad (16a)$$

$$Y_d(f) = \frac{1}{2i} d_N(f - 0.64), \quad f_0 = 0.64, \quad (16b)$$

where $d_N(f - 0.14) = d_N(f - 0.64)$. When the sampling times are corrupted by jitter noise, (13) and Lemma 1 give

$$E[Y_d(f)] = \frac{\varphi(-0.14)}{2i} d_N(f - 0.14), \quad f_0 = 0.14, \quad (16c)$$

$$E[Y_d(f)] = \frac{\varphi(-0.64)}{2i} d_N(f - 0.64), \quad f_0 = 0.64. \quad (16d)$$

The result is still periodic, but with different damping depending on the frequency.

This is also shown in Figure 2, where $|Y_d(f)|$ and $|E[Y_d(f)]|$ are plotted for the different cases in (16). The sampling jitter noise is given by (15a). From Figure 1, we know that the amplitude bias is larger for higher frequencies f_0 , $0 < \varphi(-0.64) < \varphi(-0.14) < 1$, which is why we can separate the two frequencies in this jitter sampling case.

4 Covariance Compensation

Bias compensating $Y_d(f)$ implies unbiased estimates of $S_c(f; \theta)$. To reach asymptotic efficiency, also the variance of $Y_d(f)$ is needed. In this section, both Weighted Least Squares (WLS) and Maximum Likelihood (ML) estimates are stated based on this covariance.

4.1 Covariance of $Y_d(f)$

To use these two estimators we need the covariance of Y , R_Y , which we get from the expressions in (10) as

$$R_Y(f, \psi) = R_{\hat{S}}(f, \psi) + R_{\hat{V}}(f, \psi), \quad (17)$$

with

$$\begin{aligned} R_{\hat{S}}(f, \psi) &\triangleq \text{Cov}(\hat{S}_d(f), \hat{S}_d(\psi)) \\ &= \iint S_c(\eta) \text{Cov}(W(f, \eta), W(\psi, \zeta)) S_c(\zeta)^* d\eta d\zeta, \\ &= \iint S_c(\eta) R_W(f, \eta, \psi, \zeta) S_c(\zeta)^* d\eta d\zeta. \end{aligned} \quad (18)$$

Here, $R_W(f, \eta, \psi, \zeta) \triangleq \text{Cov}(W(f, \eta), W(\psi, \zeta))$ is the covariance of the stochastic frequency window in (12). The contribution from the measurement noise is

$$R_{\hat{V}}(f, \psi) = \text{Cov}(\hat{V}(f), \hat{V}(\psi)) = \sigma^2 \sum_{k=0}^{N-1} e^{-i2\pi(f-\psi)kT},$$

since we restricted the investigation to white Gaussian measurement noise.

4.2 Covariance of $W(f)$

The covariance of $W(f)$ is

$$\begin{aligned} R_W(f, \eta, \psi, \zeta) \\ = E[W(f, \eta)W^*(\psi, \zeta)] - \mu_W(f, \eta)\mu_W(\psi, \zeta)^* \end{aligned} \quad (19)$$

and the second order moment is given by the following lemma.

Lemma 2 (Second Order Moment of W) *The second order moment of W is given as*

$$E[W(f, \eta)W^*(\psi, \zeta)] = \Phi^T(\eta, \zeta) D_N(f - \eta, \psi - \zeta).$$

where

$$\Phi(f, \psi) = \begin{pmatrix} \varphi(-f)\varphi(\psi) \\ \varphi(\psi - f) \\ \varphi(-f)\varphi(\psi) \end{pmatrix}$$

and

$$D_N(f, \psi) = \begin{pmatrix} \sum_{k=0}^{N-1} \sum_{l=0}^{k-1} e^{-i2\pi(fk-\psi l)T} \\ d_N(f - \psi) \\ \sum_{k=0}^{N-1} \sum_{l=0}^{k-1} e^{-i2\pi(-\psi k + fl)T} \end{pmatrix}.$$

The functions were defined in Lemma 1

PROOF. The second moment of the stochastic window, W , is by definition

$$\begin{aligned} E[W(f, \eta)W^*(\psi, \zeta)] &= \sum_{k,l} e^{i2\pi(\eta-f)kT} E[e^{i2\pi(\eta\tau_k - \zeta\tau_l)}] e^{-i2\pi(\zeta-\psi)lT} \\ &= \sum_k \sum_{l < k} \dots + \sum_k \sum_{l=k} \dots + \sum_k \sum_{l > k} \dots \\ &= \Sigma_1 + \Sigma_2 + \Sigma_3. \end{aligned}$$

The first term of the second moment above is

$$\begin{aligned} \Sigma_1(f, \psi, \eta, \zeta) &= \varphi(-\eta)\varphi(\zeta) \sum_{k=0}^{N-1} e^{i2\pi(\eta-f)kT} \sum_{l=0}^{k-1} e^{-i2\pi(\zeta-\psi)lT} \end{aligned}$$

since $\varphi(f) = E[e^{-i2\pi f\tau_k}]$ was the CF of τ_k and τ_k are i.i.d. The second term is the sum over $l = k$,

$$\begin{aligned} \Sigma_2(f, \psi, \eta, \zeta) &= \sum_{k=0}^{N-1} e^{i2\pi(\eta-f-\zeta+\psi)kT} \varphi(\zeta - \eta) \\ &= \varphi(\zeta - \eta) d_N(f - \eta - \psi + \zeta) \end{aligned}$$

and, finally, the third term is

$$\begin{aligned} \Sigma_3(f, \psi, \eta, \zeta) &= \varphi(-\eta)\varphi(\zeta) \sum_{l=0}^{N-1} e^{-i2\pi(\zeta-\psi)lT} \sum_{k=0}^{l-1} e^{i2\pi(\eta-f)kT} \end{aligned}$$

Identification of terms gives the result of the lemma.

The two factors, Φ and D_N , correspond to parts depending on the sampling noise and on the finite sampling, respectively, cf. $\varphi(f)$ and $d_N(f)$ in Lemma 1. It is also quite straightforward to verify $R_W(f, \eta, \psi, \zeta) = 0$ when there is no sampling jitter present, since in that case, $E[W(f, \eta)W^*(\psi, \zeta)] = d_N(f - \eta)d_N(\psi - \zeta)^* = \mu_W(f, \eta)\mu_W(\psi, \zeta)^*$, and (19) evaluates to zero.

4.3 Weighted Bias Compensated Least Squares

The jitter noise will perturb some frequencies more than others, which can be mitigated with the Weighted Least Squares (WLS) approach [see for example Kailath et al., 2000, Ch. 2], where the weights, $\lambda(f)$, are given by the covariance. Furthermore, the correlation between different frequencies needs to be taken into account. That is, (9) is here extended to

$$\begin{aligned} \hat{\theta}^{\text{WBCLS}} &= \arg \min_{\theta} \{ (\mathbf{Y} - \mu_{\mathbf{Y}}(\theta))^* \mathbf{R}_{\mathbf{Y}}(\theta)^{-1} (\mathbf{Y} - \mu_{\mathbf{Y}}(\theta)) \}, \end{aligned} \quad (20)$$

where $\mathbf{R}_{\mathbf{Y}} = I$ corresponds to (9). Here

$$\mathbf{Y} = (Y_d(f_1), \dots, Y_d(f_n))^T,$$

$\mu_{\mathbf{Y}} = E[\mathbf{Y}]$ from (13) and Lemma 1, and $\mathbf{R}_{\mathbf{Y}} = \text{Cov}(\mathbf{Y})$ as given in (17)–(19) and Lemma 2. The dependence on the parameter θ is given from the model, cf. (4).

4.4 Maximum Likelihood

The central limit theorem together with the fact that the sampling noise τ_k and the measurement noise v_k are i.i.d., Equations (10) and (11) indicate that both \hat{V} and \hat{S}_d are asymptotically Gaussian, and hence

$$\mathbf{Y} \in \text{As}\mathcal{N}(\mu_{\mathbf{Y}}, \mathbf{R}_{\mathbf{Y}}), \quad (21)$$

where $\text{As}\mathcal{N}$ denotes the asymptotic Gaussian distribution. When this holds, the asymptotic Maximum Likelihood (ML) estimator, see for example, Ljung [1999, Ch. 7] or Fisher [1922], can be defined as

$$\begin{aligned} \hat{\theta}^{\text{AsML}} &= \arg \min_{\theta} \{ (\mathbf{Y} - \mu_{\mathbf{Y}}(\theta))^* \mathbf{R}_{\mathbf{Y}}(\theta)^{-1} (\mathbf{Y} - \mu_{\mathbf{Y}}(\theta)) \\ &\quad + \ln \det \mathbf{R}_{\mathbf{Y}}(\theta) \}, \end{aligned} \quad (22)$$

with the vector \mathbf{Y} , the mean $\mu_{\mathbf{Y}}$ and covariance $\mathbf{R}_{\mathbf{Y}}$ given in Section 4.3.

4.5 Implementation Issues

In the upcoming examples in Section 5, a grid based search is implemented to avoid problems with local minima. However, a promising alternative would be to use numerical methods to minimize the estimators in (20) and (22), like gradient or Gauss-Newton searches as described in Ljung [1999]. To avoid potential problems with local minima in this case, the parameter vector can be initialized by standard methods for OE models [Ljung, 1999] neglecting the jitter noise. This should work fine for moderate levels of jitter noise.

One way to avoid the matrix inversion of $\mathbf{R}_{\mathbf{Y}}$ is to consider only frequencies $f = n/NT$ for integer valued n . The idea is that the Fourier transform $Y_d(f)$ is asymptotically uncorrelated [Ljung, 1999] for these frequencies, which would simplify the estimators considerably. For instance, the asymptotic ML estimator from (22) becomes

$$\begin{aligned} \hat{\theta}^{\text{AsML}} &= \arg \min_{\theta} \sum_n \left\{ \frac{|Y_d(n/NT) - \mu_{\mathbf{Y}}(n/NT; \theta)|^2}{R_{\mathbf{Y}}(n/NT, n/NT; \theta)} \right. \\ &\quad \left. + \ln R_{\mathbf{Y}}(n/NT, n/NT; \theta) \right\}. \end{aligned} \quad (23)$$

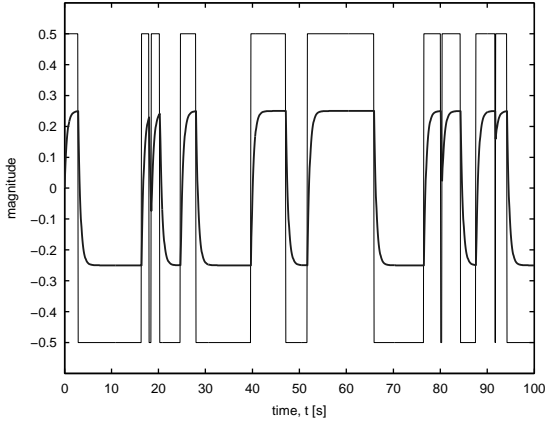


Fig. 3. Example of a PRBS signal ($u(t)$ thin) and the output from $G(s) = 1/(s + 2)$, ($s(t)$ thick), in continuous-time.

5 Examples

Three numerical examples are presented. First, a known first order OE model is used to illustrate that the jitter noise distribution is identifiable. Then, a \mathcal{M}_{OE} and a \mathcal{M}_{JOE} example, both with second order systems, are presented to illustrate the main results.

To find the best parameter vector θ the maximum likelihood estimator (22) will be used. The DTFT is calculated for the frequencies

$$f_k = k/NT, \quad k = 0, \dots, \lfloor N/2 \rfloor - 1, \quad (24)$$

i.e., we get the DFT. In the examples, we will also use the bias compensated least squares estimator (14) to compare with.

The input $u(t)$ is a Pseudo Random Binary Sequence (PRBS) signal, which jumps between -0.5 and 0.5 at time intervals of exponentially distributed lengths with a mean value of 1 s. The measurement noise $v(t)$ is zero-mean Gaussian with standard deviation 0.01. The number of measurements, N , is set to 200.

5.1 Known First Order OE Model with Unknown Jitter pdf

For this example, we chose

$$\mathcal{M} : g_\theta(t) = \mathcal{L}^{-1} \left(\frac{1}{s+2} \right), h_\theta(t) = \delta(t), p_\theta(\tau), \quad (25)$$

where \mathcal{L} denotes the Laplace transform operator. Here, only $p_\theta(t)$ is parameter dependent, and the system is known. The input and measurement noise were defined above. In Figure 3, an example of the input $u(t)$ and output $y(t)$ is given for the above settings.

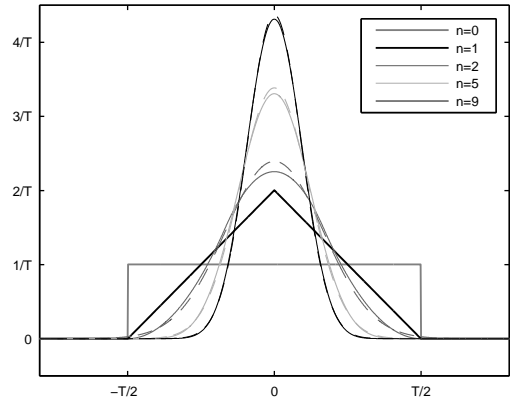


Fig. 4. The solid lines show the pdf, $p_\theta(\tau)$, corresponding to different values of $\theta = (n, T/2)^T$, when the CF is $\varphi(f) = \text{sinc}(\frac{fT}{n+1})^{n+1}$. For the higher values of n , the dashed lines show the Gaussian bells with the same variance as the parameterized pdf. The difference is almost invisible in this plot for $n = 9$.

The focus is on finding the distribution of the sampling noise. We choose a parameterization based on a sum of $n + 1$ uniformly distributed variables in the interval $[-a/(n+1), a/(n+1)]$. That is, the support of the pdf, $p_\theta(\tau)$, is $[-a, a]$, and its parameters are $\theta = (n, a)^T$. The characteristic function of $p_\theta(\tau)$ can be written as

$$\varphi_\theta(f) = \left(\frac{\sin(\frac{2\pi fa}{n+1})}{\frac{2\pi fa}{n+1}} \right)^{n+1} = \text{sinc} \left(\frac{2fa}{n+1} \right)^{n+1}, \quad (26)$$

since a sum of independent stochastic variables corresponds to convolving their pdfs, which in turns corresponds to multiplying their characteristic functions. The expected value is 0 and the variance is $\sigma^2 = \frac{1}{n+1} \frac{a^2}{3}$. Setting $n = 0$ yields a rectangular distribution, $n = 1$ a triangular distribution, and increasing n makes the distribution converge to a normal distribution, with decreasing variance. Thus, a higher n indicates a more narrow distribution. See Figure 4 for a few examples of the pdf and corresponding (of equal variance) Gaussian distribution.

Now, we want to test both jitter covering the whole interval $[-T/2, T/2]$, as well as only a portion of it; change the order but keep the variance fixed; and also look outside the model class. Therefore, we use $p_\theta(\tau)$ given by the CF in (26), for four sets of true parameters $[n^0, a^0] = [0, T/2]$, $[0, T/2\sqrt{3}]$, $[2, T/2]$, and $[2, T/2\sqrt{3}]$, with sampling time $T = 1$. This gives the variance of the jitter as $(\sigma^0)^2 = (T/2\sqrt{3})^2$, $(T/6)^2$, $(T/6)^2$ and $(T/6\sqrt{3})^2$, respectively. We also let the true pdf be a zero-mean Gaussian distribution with the same variance as the two last cases above, namely $\sigma^2 = (T/6)^2$ and $\sigma^2 = (T/6\sqrt{3})^2$.

Table 1

Results for identification of model (25). Mean value and standard deviation of $\hat{\theta} = (\hat{n} \hat{a})^T$ and $\hat{\sigma} = \hat{a}/\sqrt{3(\hat{n} + 1)}$ for six sets of true parameters, θ^0 . In each of the six blocks, the top row shows the true parameter values. The following two rows are the results for BCLS estimation, mean value and standard deviation. The last two rows are the corresponding results for ML estimation. The two lowest blocks have true pdf's outside the model class, which is why no true values of n and a are given.

	n	a	σ
θ^0	0	0.5	$1/\sqrt{12} \approx 0.289$
$E[\hat{\theta}^{\text{BCLS}}]$	0.2145	0.4683	0.2599
$\text{Std}[\hat{\theta}^{\text{BCLS}}]$	0.7360	0.0678	0.0489
$E[\hat{\theta}^{\text{AsML}}]$	0.0437	0.4801	0.2719
$\text{Std}[\hat{\theta}^{\text{AsML}}]$	0.0789	0.0233	0.0170
θ^0	0	$1/\sqrt{12} \approx 0.289$	$1/6 \approx 0.167$
$E[\hat{\theta}^{\text{BCLS}}]$	1.1818	0.3759	0.1682
$\text{Std}[\hat{\theta}^{\text{BCLS}}]$	1.6815	0.1186	0.0649
$E[\hat{\theta}^{\text{AsML}}]$	0.7073	0.3660	0.1671
$\text{Std}[\hat{\theta}^{\text{AsML}}]$	0.8184	0.0662	0.0224
θ^0	2	0.5	$1/6 \approx 0.167$
$E[\hat{\theta}^{\text{BCLS}}]$	1.1386	0.3752	0.1676
$\text{Std}[\hat{\theta}^{\text{BCLS}}]$	1.6608	0.1088	0.0592
$E[\hat{\theta}^{\text{AsML}}]$	0.6420	0.3766	0.1732
$\text{Std}[\hat{\theta}^{\text{AsML}}]$	0.6439	0.0753	0.0250
θ^0	2	$1/\sqrt{12} \approx 0.289$	$1/(6\sqrt{3}) \approx 0.096$
$E[\hat{\theta}^{\text{BCLS}}]$	1.6816	0.2618	0.1018
$\text{Std}[\hat{\theta}^{\text{BCLS}}]$	1.9972	0.1553	0.0607
$E[\hat{\theta}^{\text{AsML}}]$	2.4947	0.3015	0.1011
$\text{Std}[\hat{\theta}^{\text{AsML}}]$	2.1080	0.0866	0.0205
θ^0	—	—	$1/6 \approx 0.167$
$E[\hat{\theta}^{\text{BCLS}}]$	1.8175	0.3267	0.1281
$\text{Std}[\hat{\theta}^{\text{BCLS}}]$	2.0357	0.1607	0.0678
$E[\hat{\theta}^{\text{AsML}}]$	1.0115	0.3589	0.1515
$\text{Std}[\hat{\theta}^{\text{AsML}}]$	1.0415	0.0855	0.0297
θ^0	—	—	$1/(6\sqrt{3}) \approx 0.096$
$E[\hat{\theta}^{\text{BCLS}}]$	1.8807	0.1958	0.0752
$\text{Std}[\hat{\theta}^{\text{BCLS}}]$	2.1966	0.1425	0.0530
$E[\hat{\theta}^{\text{AsML}}]$	3.4188	0.2696	0.0802
$\text{Std}[\hat{\theta}^{\text{AsML}}]$	2.4361	0.0877	0.0219

The identification procedure now follows:

- The mean value $\mu_Y(f)$ is given by (13) and Lemma 1, together with the model specifications given in (25) and (26).
- The covariance, $R_Y(f, f)$, can be deduced from Lemma 2.
- The parameter estimates, $\hat{\theta}$, are given from (14) for the BCLS estimate and from (22) for the asymptotic ML estimate.
- After minimization, an estimate of the standard deviation of the sampling noise is given by $\hat{\sigma} = \hat{a}/\sqrt{3(\hat{n} + 1)}$ for each run.

For the six different pdfs, we evaluate the mean value and the standard deviation of the estimated parameters, $\hat{\theta}$, based on 128 different realizations of the sampling noise sequence and the input. The result is shown in Table 1, for both the AsML and BCLS estimate, using the model in (25). We also compute the estimate of the standard deviation σ , implied by \hat{n} and \hat{a} . We can conclude the following:

- The standard deviation σ for the sampling noise is accurately estimated, while the exact parameter constellation, $(n \ a)^T$, is harder. Simulations with more data are not included here. However, for $N = 400$ the estimation of σ is slightly better, and for some cases the AsML estimates of n and a are also improved. For $n = 1000$ no further improvement is shown. The parameterization of $\varphi_\theta(f)$ would have to be improved, for cases where identification of the exact distribution is important.
- The distribution with $n^0 = 0$, $a^0 = 0.5$, rectangular over the whole interval, has best results.
- The AsML estimate has the smallest error for all cases, indicating that the covariance R_Y varies significantly along the frequency axis.
- True pdfs outside the model class (last two tests) pose no problem to the estimation of σ .

Despite the small number of measurements, the estimation of the jitter noise standard deviation works nice. The performance is comparable with earlier results where covariance estimation was the main object [Verbeyst et al., 2006, Verspecht, 1995]. The results in this section are promising for further investigation, when jitter noise properties are sought. However, we continue with the focus of this work, model estimation.

5.2 Unknown Second Order OE Model with Known Jitter pdf

Now, let us focus on the linear system, $g(t)$, that is, the \mathcal{M}_{OE} problem in (2a). We illustrate the algorithms for a second order system, specified by its transfer function

Table 2

Results for identification of model (27). Mean value and standard deviation of $\hat{\theta} = (\hat{k}_0 \hat{z} \hat{p}_r \hat{p}_i)^T$ for the three estimates given by (8), (14) and (22).

Model (27)	k_0	p_r	p_j
θ^0	6.25	-1.5	2
$E[\hat{\theta}^{\text{LS}}]$	5.8468	-1.5876	1.9508
$\text{Std}[\hat{\theta}^{\text{LS}}]$	0.1723	0.0448	0.0045
$E[\hat{\theta}^{\text{BCLS}}]$	6.2536	-1.4981	2.0019
$\text{Std}[\hat{\theta}^{\text{BCLS}}]$	0.0479	0.0385	0.0415
$E[\hat{\theta}^{\text{AsML}}]$	6.2499	-1.5008	2.0000
$\text{Std}[\hat{\theta}^{\text{AsML}}]$	0.0480	0.0343	0.0401

G. The model is

$$\mathcal{M} : g_\theta(t), h_\theta(t) = \delta(t), p_\theta(\tau) = p(\tau). \quad (27a)$$

The system $g_\theta(t) = \mathcal{L}^{-1}(G_\theta(s))$ is parameterized as

$$G_\theta(s) = \frac{k_0}{(s - (p_r + ip_i))(s - (p_r - ip_i))} \quad (27b)$$

$$\theta = (k_0 \ p_r \ p_i)^T \quad (27c)$$

with the true parameters being $\theta^0 = (6.25, -1.5, 2)^T$ during this run. The sampling noise distribution, $p(\tau)$, is known and given by (26) with $n = 0$ and $a = T/2$. The input and measurement noise are the same as in Section 5.1.

The parameters are now given by the same identification procedure as described on p. 8. The result for the AsML and BCLS estimation is shown in Table 2. We also compare with the results when no bias-compensation is used (LS in (8)).

The following conclusions can be drawn from the results:

- Estimates from BCLS and AsML are unbiased with small estimated variances ($< 3\%$).
- The difference between BCLS and AsML is not as obvious as in the former identification setup, although all parameters are estimated slightly better using AsML.
- As has been seen before, in both analysis and other literature, the LS estimate introduces a significant bias.

The bias is even more visible when studying the magnitude of the identified system. Figure 5 shows $|G_{\theta^0}(i2\pi f)|$ compared to $|G_\theta(i2\pi f)|$ with $\theta = E[\hat{\theta}^{\text{LS}}]$. Using the parameters from BCLS or AsML does not give a visible difference from the true system.

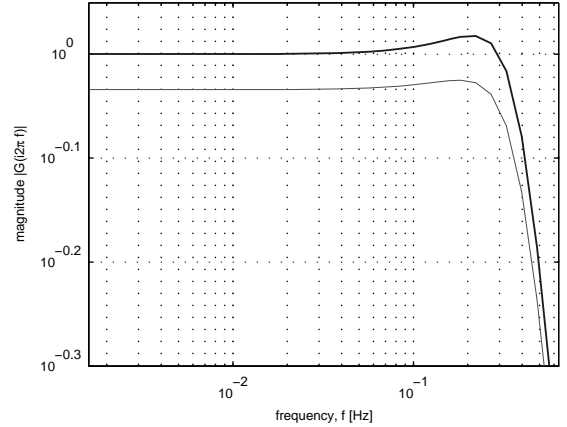


Fig. 5. A comparison of $|G_\theta(i2\pi f)|$ (thin line), given by (27) and $\theta = E[\hat{\theta}^{\text{LS}}]$ from Table 2, with the true $|G_{\theta^0}(i2\pi f)|$ (thick line). When using $\theta = E[\hat{\theta}^{\text{BCLS}}]$ or $\theta = E[\hat{\theta}^{\text{AsML}}]$, the difference from the true system is not visible.

5.3 Unknown Second Order OE Model with Unknown Jitter pdf

The previous example is extended by letting the jitter noise pdf be unknown. That is, consider the \mathcal{M}_{JOE} problem in (2c). The system $g_\theta(t)$ is given by (27) and the sampling noise pdf, $p_\theta(\tau)$, is given by (26), with $\theta = (n \ a \ k_0 \ p_r \ p_i)^T$ and the true parameters being $\theta^0 = (2, T/2\sqrt{3}, 6.25, -1.5, 2)^T$. The input and measurement noise are the same as in Section 5.1.

From the two previous sections, we know that n and a were hard to estimate, whereas k_0 , p_r and p_j posed no problem. We expect to see similar things for the joint problem. The parameters are now given by the same identification procedure as described on p. 8. The chosen parametrization enables full separation of the dynamic model and the jitter noise, since the true system output, $S_c(f)$, is only multiplied by the CF of the jitter noise, $\varphi(f)$, and not mixed.

The identification results are given in Table 3. The two estimates are good for all parameters, and no significant difference can be seen between BCLS and AsML in this case. It is interesting to note that the result is better for estimating n and a than for the corresponding pdf when the system was known, Section 5.1. Both here and in the previous section, the reduced AsML estimate (23) was used, for computational reasons.

6 Conclusions

The effects of the stochastic jitter noise τ_k in the model $y_k = s(kT + \tau_k; \theta) + v_k$ should not be neglected and modeled as included in the measurement noise v_k . The bias compensated least squares estimator describes the

Table 3

Mean value and standard deviation of $\hat{\theta} = (\hat{n} \hat{a} \hat{k}_0 \hat{z} \hat{p}_r \hat{p}_i)^T$ in the model (27) with jitter noise as defined in (26).

	n	a	k_0	p_r	p_j	σ
θ^0	2	$\frac{T}{2\sqrt{3}} \approx 0.2887$	6.25	-1.5	2	$\frac{1}{6\sqrt{3}} \approx 0.0962$
$E[\hat{\theta}^{\text{BCLS}}]$	1.8281	0.2960	6.2609	-1.4980	2.0044	0.1061
$\text{Std}[\hat{\theta}^{\text{BCLS}}]$	1.0357	0.0324	0.0331	0.0262	0.0289	0.0179
$E[\hat{\theta}^{\text{AsML}}]$	2.2578	0.2918	6.2513	-1.5045	1.9964	0.0947
$\text{Std}[\hat{\theta}^{\text{AsML}}]$	0.8534	0.0329	0.0334	0.0221	0.0278	0.0072

remedy for the jitter effects:

$$\hat{\theta}^{\text{BCLS}} = \arg \min_{\theta} \sum_f |Y_d(f) - \int_{-\infty}^{\infty} S(\psi; \theta) \varphi(-\psi; \theta) d_N(f - \psi) d\psi|^2.$$

This criterion minimizes the LS distance between measurement DTFT and the model's FT, taking leakage and alias effects into account via a convolution with the normalized Dirichlet kernel $d_N(f - \psi)$. The characteristic function $\varphi(-\psi; \theta)$ of the jitter noise implies a scaling of the signal model $S_c(f)$. Neglecting this scaling factor results in a bias.

A second order compensation of jitter effects leads to the WBCLS estimator (20) and the asymptotic maximum likelihood estimator (22). The estimators were illustrated on simulated examples for continuous-time OE system identification, with known or with parametrized jitter distribution. The results are very promising and a clear motivation for developing dedicated algorithms for the case of jitter noise. For example, efficient Gauss-Newton algorithms would be useful, where also unknown noise models (Box-Jenkins structures) could be targeted.

References

- I. Bilinskis and A. Mikelsons. *Randomized Signal Processing*. Prentice Hall, 1992.
- F. Eng and F. Gustafsson. System identification using measurements subject to stochastic time jitter. In *IFAC World Congress*, Prague, Czech Republic, July 2005.
- F. Eng and F. Gustafsson. Bias compensated least squares estimation of continuous time output error models in the case of stochastic sampling time jitter. In *IFAC Symp. on Syst. Ident. (SYSID)*, Newcastle, Australia, Mar. 2006.
- R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A.*, 222:309–368, 1922. URL <http://www.jstor.org>.
- J. Gillberg and F. Gustafsson. Frequency-domain continuous-time AR modeling using non-uniformly sampled measurements. In *IEEE Int. Conf. on Acoust., Speech, Signal Processing*, Philadelphia, Pennsylvania, USA, 2005.
- J. Gillberg and L. Ljung. Frequency-domain identification of continuous-time ARMA models from sampled data. In *IFAC World Congress*, Prague, Czech Republic, July 2005a.
- J. Gillberg and L. Ljung. Frequency-domain identification of continuous-time output error models from sampled data. In *IFAC World Congress*, Prague, Czech Republic, July 2005b.
- T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, 2000.
- L. Ljung. *System Identification – Theory for the User*. Prentice Hall, 2 edition, 1999.
- F. Marvasti, editor. *Nonuniform Sampling: Theory and Practice*. Kluwer Academic Publishers, 2001.
- A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 3 edition, 1991. ISBN 0-07-048477-5.
- R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. IEEE Press, 2001.
- R. Pintelon and J. Schoukens. ML identification of closed-loop systems in a specified frequency band. In *IFAC World Congress*, Prague, Czech Republic, July 2005.
- J. Schoukens, Y. Rolain, and R. Pintelon. Analysis of windowing/leakage effects in frequency response function measurements. In *IFAC World Congress*, Prague, Czech Republic, July 2005.
- T. M. Souders, D. R. Flach, C. Hagwood, and G. L. Yang. The effects of timing jitter in sampling systems. *IEEE Trans. Instrum. Meas.*, 39(1):80–85, Feb. 1990. ISSN 0018-9456. doi: 10.1109/19.50421.
- F. Verbeyst, Y. Rolain, J. Schoukens, and R. Pintelon. System identification approach applied to jitter estimation. In *Instrumentation and Measurement Technology Conference*, pages 1752 – 1757, Sorrento, Italy, Apr. 2006. doi: 10.1109/IMTC.2006.235202.
- J. Verspecht. *Calibration of a Measurement System for High Frequency Nonlinear Devices*. PhD thesis, Vrije Universiteit Brussel, Belgium, nov 1995.