

# Issues in sampling and estimating continuous-time models with stochastic disturbances <sup>★</sup>

Lennart Ljung\* Adrian Wills\*\*

\* *Division of Automatic Control, Linköpings Universitet, SE-581 80  
Linköping, Sweden (e-mail: [ljung@isy.liu.se](mailto:ljung@isy.liu.se))*  
\*\* *School of Electrical Engineering and Computer Science, University  
of Newcastle, Callaghan, NSW, 2308, Australia (e-mail:  
[adrian.wills@newcastle.edu.au](mailto:adrian.wills@newcastle.edu.au))*

**Abstract:** The standard continuous time state space model with stochastic disturbances contains the mathematical abstraction of continuous time white noise. To work with well defined, discrete time observations, it is necessary to sample the model with care. The basic issues are well known, and have been discussed in the literature. However, the consequences have not quite penetrated the practise of estimation and identification. One example is that the standard model of an observation being a snapshot of the current state plus noise independent of the state cannot be reconciled with this picture. Another is that estimation and identification of time continuous models require a more careful treatment of the sampling formulas. We discuss and illustrate these issues in the current contribution. An application of particular practical importance is the estimation of models based on irregularly sampled observations.

## 1. INTRODUCTION

A standard mathematical abstraction in control theory is the continuous time linear model with stochastic disturbances:

$$\dot{x}(t) = Ax(t) + Bu(t) + \dot{w}(t) \quad (1a)$$

$$\dot{z}(t) = Cx(t) + \dot{e}(t) \quad (1b)$$

Here  $\dot{w}(t)$  and  $\dot{e}(t)$  are stochastic disturbances. For  $x(t)$  to be a *state* in the sense that it condenses knowledge of the past, and thus becomes a Markov process, it is necessary that both  $\dot{w}(t)$  and  $\dot{e}(t)$  are unpredictable from past data. This means that they have to be *white noises*. It is well known that the mathematical description of this is somewhat advanced: These processes will have to have infinite variances, and a formal mathematical description is via stochastic integrals of their integrated versions, that are Wiener processes:

$$dx(t) = Ax(t)dt + Bu(t)dt + dw(t) \quad (2a)$$

$$dz(t) = Cx(t)dt + de(t) \quad (2b)$$

The incremental covariances of the involved processes are

$$E \begin{bmatrix} dw(t) \\ de(t) \end{bmatrix} \begin{bmatrix} dw(t) \\ de(t) \end{bmatrix}^T = \begin{bmatrix} R_1 & R_{12}^T \\ R_{12} & R_2 \end{bmatrix} dt \quad (2c)$$

See, e.g. Åström (1970) or for an account of this.

Now, in real life the time continuous output  $z(t)$  is of course not observed, and neither are any instantaneous snapshots since they would have infinite variance. Various ways to formulate a realistic discrete time version of this have been suggested, and this paper is about several aspects and issues involved in such formulation.

Our original interest in this problem came from writing code for identifying (1) from discrete time, sampled measurements of the inputs and outputs. In the course of this we ran into several issues, that have been considered in the

<sup>★</sup> This work was supported by the Swedish Research Council and the Australian Research Council.

past, but not dealt with in a comprehensive manner. It is the purpose of this paper to sort out and illustrate some of these issues.

## 2. SAMPLING

### 2.1 The state process

The state process  $x(t)$  is a well defined signal from (1a) and its values at discrete time instances  $t_k, k = 1, \dots$  can readily be found. Assume that the input is piecewise constant:

$$u(t) = u(t_k) = u_k \quad t_k \leq t < t_{k+1} \quad (3)$$

Then

$$x(t_{k+1}) = F_k x(t_k) + G_k u(t_k) + \tilde{w}(t_k) \quad (4a)$$

$$\tilde{w}_k(t) = \int_{t_k}^{t_{k+1}} e^{A(t+T-s)} dw(s) \quad (4b)$$

$$E \tilde{w}_k(t) \tilde{w}_k^T(t) = \tilde{R}_1(\delta_k) = \int_0^{\delta_k} e^{As} R_1 e^{A^T s} ds \quad (4c)$$

$$\delta_k = t_{k+1} - t_k \quad (4d)$$

$$F_k = e^{A\delta_k} \quad (4e)$$

$$G_k = P_k B \quad (4f)$$

$$P_k = A^{-1}(e^{A\delta_k} - I) \quad (4g)$$

Here  $R_1 dt$  is the incremental variance of the Wiener process  $w(t)$ , see (2c). This expression is given in many places, e.g. Sec 3.10 in Åström (1970).

Note that  $\tilde{R}_1(\delta_k)$  is the solution  $S$  to the Lyapunov equation

$$AS + SA^T + R_1 - e^{A\delta_k} R_1 e^{A^T \delta_k} \quad (5)$$

**Remark 1:** It would be proper to let the time index of  $\tilde{w}$  be  $t_{k+1}$ , since it depends on  $w(s)$  up to time  $t_{k+1}$ .

**Remark 2:** We have in (4g) used the inverse of  $A$ . Actually, the computation of  $P_k$  does not involve such

inversion, and the matrix is also well defined for singular  $A$ , but for convenience here and a few other places we use this explicit expression.

## 2.2 Discrete time observations

One idea is simply to postulate that discrete time noisy observations of the state are available, without detailing on how they are obtained from (1b):

$$y(t_k) = H_k x(t_k) + v(t_k) \quad (6)$$

This is the approach taken in Jazwinski (1970), Section 7.2, and from this a continuous-discrete Kalman filter is developed, updating estimates of the continuous state process (1a). Åström takes the same approach for identifying continuous time state space models in Åström (1980), eq (3.1)-(3.2). The same philosophy is used for DAEs in Gerdin et al. (2007).

**Remark 3:** For the filtering calculations to work, it is essential that  $v(t_k)$  is independent of  $x(t_k)$ . It is important to note that *it is impossible to have such a finite-variance observation at time  $t_k$  based on (1b)!* The reason is that in order to achieve finite variance, the signal  $z$  must be low-pass, pre-sample filtered before sampling. This filtering must occur after time  $t_k$  in order for the filtered noise-component to be independent of  $x(t_k)$ . So, a proper time index of  $y$  in (6) must be larger than  $t_k$ . Now, if our concern is to predict future values of  $y$  (as in an identification application), this is not so essential, since it is really just an issue of labeling the time stamps. But for state estimation, it may affect what is considered to be a predicted and a filtered state estimate.

An approach to describe how (6) relates to (1b) is to assume *integrated sampling*:

$$y(t_k) = \frac{1}{\Delta_k} \int_{t_k}^{t_k + \Delta_k} z(s) ds \quad (7)$$

**Remark: The proper time index of  $y$  would indeed be  $(t_k + \Delta_k)$ , since this measurement will not be available until that time!**

This is used e.g. in (10.17) in Åström (1970) (which does not divide by the time interval) and in (22.10.84) in Goodwin et al. (2001). Integrated sampling is also discussed in e.g. Goodwin et al. (1995) (which considers duality with a control problem) and Feuer and Goodwin (1996), Chapter 6 (which considers “delta-versions” of integrated sampling). These authors describe the case without input.

It is not difficult to apply (7) to (1) under the assumption (3) and obtain

$$y(t_k) = H_k x(t_k) + D_k u(t_k) + v(t_k) \quad (8a)$$

$$H_k = \frac{1}{\Delta_{T_k}} C P_k = C + C A \Delta_k / 2 + C A \Delta_k^2 / 3! + \dots \quad (8b)$$

$$D_k = \frac{1}{\Delta_k} C [A^{-2} (e^{A \Delta_k} - I) - A^{-1} \Delta_k] B = C (\Delta_k / 2 + A \Delta_k^2 / 3! + \dots) B \quad (8c)$$

$$v(t_k) = f(e(s), w(s), t \leq s \leq t_k + \Delta_k) \quad (8d)$$

Here,  $f$  is a function of the indicated signals, whose exact expression is easy to derive. It is straightforward, but somewhat laborious to compute the covariance of  $v$  and its cross covariance with  $\tilde{w}(t_k)$  in (4). See Åström (1970), Section 3.10, but the results are

$$\begin{aligned} E v(t_k) v^T(t_k) &= \tilde{R}_2(\Delta_k) \\ &= \frac{1}{\Delta_k^2} [C A^{-1} (\tilde{R}_1(\Delta_k) - A_{\Delta_k} R_1 \\ &\quad - R_1 A_{\Delta_k}^T + R_1 \Delta_k) A^{-T} C^T \\ &\quad + Z + Z^T + R_2 \Delta_k] \end{aligned} \quad (9a)$$

$$\begin{aligned} E \tilde{w}(t_k) v^T(t_k) &= \tilde{R}_{12}(\Delta_k) \\ &= \frac{1}{\Delta_k} [(\tilde{R}_1(\Delta_k) - A_{\Delta_k} R_1) A^{-T} C^T + R_{12}] \end{aligned} \quad (9b)$$

where

$$A_{\Delta_k} = A^{-1} (e^{A \Delta_k} - I) \quad (9c)$$

$$Z = C A^{-1} (A_{\Delta_k} R_{12} - R_{12} \Delta_k) \quad (9d)$$

**Remark:** The proper time index of  $v(t)$  would be  $t + \Delta_T$ .

Note that the integration in (7) over a time interval  $\Delta_k$  is inherent in most A/D converters. However, if the sampling rate of the system is  $T$ , it is normal to use a pre-sampling, lowpass filter with cutoff frequency around  $\pi/T$  before the A/D converter. This means that the filtered signal  $z$  is not changing much over the time interval  $T$ . So to capture the typical noise reducing effect of presampling filters, it is natural to use the choice

$$\Delta_k = t_{k+1} - t_k \quad (10)$$

Regardless of the time index issues,  $v(t)$  is independent of  $x(t)$  and  $\tilde{w}(t)$  is independent of  $x(t)$ .

## 2.3 Innovations Form

So, we end up with a discrete time equation

$$x(t_{k+1}) = F_k x(t_k) + G_k u(t_k) + \tilde{w}(t_k) \quad (11a)$$

$$y(t_k) = H_k x(t_k) + D_k u(t_k) + v(t_k) \quad (11b)$$

where the covariances of  $\tilde{w}$  and  $v$  are given by (5) and (9).

From these equations, the Kalman gain  $\tilde{K}_k$  can be computed via the time-varying Riccati equation in the well known manner

$$\begin{aligned} P_{k+1} &= F_k P_k F_k^T + \tilde{R}_1(\delta_k) - (F_k P_k H_k^T + \tilde{R}_{12}(\Delta_k)) \times \\ &\quad (H_k P_k H_k^T + \tilde{R}_2(\Delta_k))^{-1} (F_k P_k H_k^T + \tilde{R}_{12}(\Delta_k))^T \end{aligned} \quad (12a)$$

$$\tilde{K}_k = (F_k P_k H_k^T + \tilde{R}_{12}(\Delta_k)) (H_k P_k H_k^T + \tilde{R}_2(\Delta_k))^{-1} \quad (12b)$$

to yield the *innovations form*:

$$\hat{x}(t_{k+1}) = F_k \hat{x}(t_k) + G_k u(t_k) + \tilde{K}_k \epsilon(t_k) \quad (13a)$$

$$y(t_k) = H_k \hat{x}(t_k) + D_k u(t_k) + \epsilon(t_k) \quad (13b)$$

This representation has the property that the second order properties of the inputs and outputs are the same in (11) and (13).

By replacing  $\epsilon$  in (13a) by

$$y(t_k) - H_k \hat{x}(t_k) - D_k u(t_k)$$

from (13b), equation (13a) becomes the Kalman filter for estimating the state. The predicted output is

$$\hat{y}(t_{k+1}) = H_{k+1} \hat{x}(t_{k+1}) + D_{k+1} u(t_{k+1}) \quad (14)$$

## 2.4 Continuous Time Innovations Form

For the continuous time, time-invariant equation (1) the innovations form is analogously

$$\begin{aligned} \dot{\hat{x}}(t) &= A \hat{x}(t) + B u(t) + K \nu(t) \\ y(t) &= C \hat{x}(t) + \nu(t) \end{aligned} \quad (15)$$

where  $K$  is the steady-state Kalman gain.

### 2.5 Equidistant Sampling

Suppose now that the output is sampled with a constant sampling interval  $T : t_k = kT$ . Suppose also that the integration-time/time-constant of the sampling device (7) is constant and equal to  $\Delta$ . Then the sampled equation (11) becomes time-invariant with matrices and covariances that do not depend on  $k$ . The time varying Kalman filter/innovations form (13) will then converge to a time-invariant system

$$\hat{x}(kT + T) = F\hat{x}(kT) + Gu(kT) + \tilde{K}\epsilon(kT) \quad (16a)$$

$$y(t) = H\hat{x}(kT) + Du(kT) + \epsilon(kT) \quad (16b)$$

where the limit Kalman gain  $\tilde{K}$  is the solution to the algebraic Riccati equation:

$$P = FPF^T + \tilde{R}_1(\delta) - (FPH^T + \tilde{R}_{12}(\Delta)) \times \\ (HPH^T + \tilde{R}_2(\Delta))^{-1} (FPH^T + \tilde{R}_{12}(\Delta))^T \quad (16c)$$

$$\tilde{K} = (FPH^T + \tilde{R}_{12}(\Delta))(HPH^T + \tilde{R}_2(\Delta))^{-1} \quad (16d)$$

### 2.6 Simplistic Sampling

The computation of  $\tilde{K}$  according to (16cd) is rather complicated. It is tempting to use a simplified formula, that ignores the averaging over  $x$  in (7) and uses an assumption that the noise is piecewise constant over the sampling interval. This sounds reasonable if  $T$  is short compared to the time constants of the system. In other words it means that the continuous time innovations description (15) is sampled as if  $\nu(t)$ , like  $u(t)$  is piecewise constant, (3), yielding

$$\hat{x}(kT + T) = F\hat{x}(kT) + Gu(kt) + \bar{K}\epsilon(kT) \quad (17a)$$

$$y(t) = C\hat{x}(kT) + \epsilon(kT) \quad (17b)$$

$$F = e^{AT}; \quad G = PB; \quad \bar{K} = PK \quad (17c)$$

$$P = A^{-1}(F - I) \quad (17d)$$

In case  $F - \bar{K}C$  is not stable, the Riccati equation corresponding to  $R_1 = \bar{K}\bar{K}^T, R_{12} = \bar{K}, R_2 = I$  is solved to obtain a new stabilizing Kalman gain. We will call this *simplistic sampling* of the continuous time innovations form. This is the approach taken in the system identification toolbox, Ljung (2007). Its manual states that this is a reasonable approximation if the sampling interval is reasonably chosen w.r.t. to the system and noise dynamics.

### 2.7 System Identification

Estimation of system parameters is closely related to the estimation/prediction problem. If  $A, B, C, K$  in (15) or  $A, B, C, R_1, R_{12}, R_2$  in (1) contain unknown parameters  $\theta$ , then these can be consistently identified by minimizing the prediction error criterion

$$\sum_{k=1}^N \|y(t_k) - \hat{y}(t_k, \theta)\|^2 \quad (18)$$

or via the closely related Maximum Likelihood criterion. Here  $\hat{y}$  is the expression from (14), depending on  $\theta$  via the expressions (12, 5, 9, 8). See, e.g. Ljung (1999). The main work in minimizing (18) concerns finding the gradients of the criterion, which is based on the derivative

$$\psi(t_k) = \frac{\partial}{\partial \theta} \hat{y}(t_k, \theta) \quad (19)$$

For the time-invariant innovations representation expressions for the gradient are quite straightforward (e.g. Section 9.6 in Ljung and Glad (1994)). In the identification toolbox Ljung (2007) the gradients of the matrices  $F, G$ , etc with respect to the parameters of the continuous time model are found by numerical differentiation.

## 3. A COLLECTION OF WHY'S

### 3.1 Why Are Continuous Time Models of Interest?

The main reason why we deal with continuous time models is that physical insight and intuition are typically best represented in continuous time. Most physical modeling work ends up with a continuous time model, and in order to make use of that knowledge it is natural to parameterize a model (1) accordingly. Even though the fit to data will have to be done in discrete time as in (18), the dimension of the vector  $\theta$  reflecting continuous time dynamics will be lower in that way.

For equidistantly sampled data one could build a black-box discrete time model, that is converted to continuous time afterwards, using the inverse sampling formulas. However to comply with any physical structure of (1) this would involve another round of optimization. If there is no physical knowledge, so (1) is black-box, and the sampling time is constant, a simple route is to first build a discrete time model and then convert it to continuous time. This route is essentially equivalent to estimating the continuous model directly for this case.

In the case of unequally sampled data, the discrete time system (13) is time varying, even if the underlying continuous system is time invariant. Then this continuous time model is the natural basis to deal with the sampled observations.

### 3.2 Why Is It Essential to Have Correct Sampling Formulas?

Well, actually it is not that essential. You could have any weird sampling formula from  $A, B, C, K$  to  $F, G, H, D, \bar{K}$  that is surjective. The model fit is done in the metric of the sampled systems, so the resulting sampled system would be the best discrete time model available. The underlying parameterization of the continuous time model may however be irrelevant. Then, on the other hand, you could as well fit a discrete time model directly. So, only if you have a genuine interest in the continuous time model (see previous question), the sampling formulas are essential.

### 3.3 Why Haven't the Correct Sampling Formulas Been Extensively Used?

The sampling formulas (4)–(13) are not new, and not difficult to derive. Still the results for integrating sampling are not widely used. They are for example typically not given in textbooks and the MATLAB CONTROL SYSTEM TOOLBOX offers no implementation. (There is more attention to the conceptually related, but different, issue of piecewise linear inputs – ‘First order hold’ –, i.e. an integrator between a piecewise constant input source and the actual input.) Also, we are not aware of any explicit inverse formulas (“d2c”) for the integrating sampling case.

In a way this is a case of “double standards:” On the one hand the observations of the ideal continuous time description (1) will need lowpass filtering to achieve realistic measurements. On the other hand one is not willing

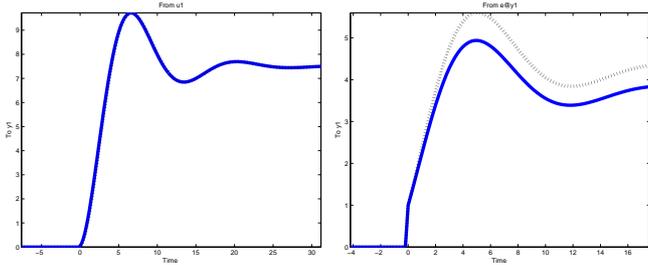


Fig. 1. Step response for DT system. Sampling interval 0.2. Solid: correct sampling. Dotted: simplistic sampling. Left: from input (the dotted and solid lines overlap in the resolution of the figure), Right: from noise source

to take the consequences of this when it comes to the effects of the input. The instantaneous sampling, giving  $H = C$  and  $D = 0$  are simpler and more attractive. The engineering rules of thumb say that a signal should be filtered by a presampling filter with cutoff frequency below the Nyquist frequency. This essentially corresponds to  $\Delta = T$  for  $t_k = kT$  in (7). For the double standards to be acceptable, one must consequently sample fast enough ( $T$  small enough) so that the effects on the dynamics of  $H \neq C$  and  $D \neq 0$  in (8) are negligible.

Another view on this, that is applicable to the equidistantly sampled case, is to build a discrete time model that describes the observed data as well as possible, and regard any sampling devices, including presampling filters as part of the model. But, again, if a continuous time model is essential, one must be careful in converting back to such a model.

#### 4. THE SAMPLING FORMULAS

For equidistant sampling, the simplistic sampling (17) is considerably simpler than the correct sampling (16). It is interesting to compare how they behave.

*Example 1.* Consider the system in innovations form

$$\begin{aligned} \dot{x} &= Ax + Bu + Ke \\ &= \begin{bmatrix} -0.357 & 1 \\ -0.243 & 0 \end{bmatrix} x + \begin{bmatrix} 0.254 \\ 1.820 \end{bmatrix} u + \begin{bmatrix} 1.287 \\ 0.786 \end{bmatrix} e \\ y &= Cx + e = [1 \ 0]x + e \end{aligned}$$

This system has a recommended sampling interval (by the CONTROL SYSTEM TOOLBOX) of 0.31 s. We consider the step responses from  $u$  and  $e$  for systems that are sampled with 0.2 s and 2 s respectively. The results are shown in Figures 1 and 2. The effects on the matrices for the different sampling rules are not insignificant. For sampling time  $T = 2$  we have

$$H = [0.6057 \ 0.7397]; \quad D = 1.1665; \quad \tilde{K} = \begin{bmatrix} 1.0279 \\ 0.2763 \end{bmatrix}$$

for (16) and

$$H = C = [1 \ 0]; \quad D = 0; \quad \tilde{K} = \begin{bmatrix} 1.2815 \\ 0.2395 \end{bmatrix}$$

for (17). Still the difference in dynamic response in Figure 2(left) is not very big.

#### 5. ESTIMATING CONTINUOUS SYSTEMS USING EQUIDISTANT SAMPLES

*Example 2.* Consider the system (20) again. Produce a “continuous time simulation” by sampling it at sampling

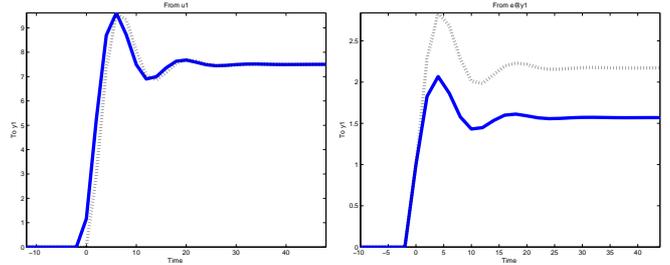


Fig. 2. Step response for DT system. Sampling interval 2. Solid: correct sampling. Dotted: simplistic sampling. Left: from input, Right: from noise source

Par	True	Est. Correct samp	Est. Simpl.samp
A(1,1)	-0.3567	-0.3511 ± 0.0129	-0.3518 ± 0.0129
A(2,1)	-0.2426	-0.2434 ± 0.0058	-0.2439 ± 0.0056
B(1,1)	0.2540	0.2169 ± 0.0572	0.0329 ± 0.0568
B(1,2)	1.8198	1.8575 ± 0.0584	1.8652 ± 0.0569
K(1,1)	1.2865	1.2556 ± 0.0415	1.1373 ± 0.0313
K(1,2)	0.7864	0.7398 ± 0.0421	0.6278 ± 0.0372

Table 1. Estimated parameter for sampling interval  $Ts = 0.2$ .  $\pm$  indicates the estimated standard deviation

interval 1 ms. Let the input be piecewise constant over time intervals of 2 s, and let the variance of  $e$  at this sampling rate be 1000. (Corresponding to a noise intensity of the continuous source of 1.) Generate 2560 seconds of such a data record (2.56 million samples). Create integrated sampled data using (7) for  $\Delta = T = 0.2$  and  $\Delta = T = 2$ . The resulting data are shown in Figure 3. A continuous time model of the same structure as (20) with six unknown parameters, corresponding to  $A(1,1)$ ,  $A(2,1)$ ,  $B(1,1)$ ,  $B(2,1)$ ,  $K(1,1)$  and  $K(2,1)$  was estimated using (18) with the correct sampling formula and the simplistic sampling formula for the two sampling intervals.

**Remark:** In connection with (7) we commented upon the time labeling of the outputs. While it is essentially a label in the filtering context, it plays a role for the alignment of inputs and outputs for the identification application. Therefore we tested the simplistic sampling both for the case of time labeling as in (7) and with a shifted version, naming  $y(t_k) = y((k+1)T)$ . The identified continuous time models with simplistic sampling for the two cases are shown in Figure 4. Clearly, such a time shift (which corresponds to an “anti-causal” shift of the input (InputDelay = -Ts)) is beneficial for the model, and we adopt this practice for the simplistically sampled model. For the correctly sampled model, there is no confusion about the time labels, since the formulas correctly account for the dependencies.

The resulting estimated parameters with their estimated standard deviations are shown in Tables 1 and 2. We see that the estimated parameters are fine when the correct sampling has been used. The true values are well inside reasonable confidence regions. The models obtained by the simplistic sampling have worse accuracy, especially for the larger sampling interval. The true values are in several cases not within reasonable confidence intervals.

A few comments are in order:

- The main reason why the simplistically sampled system gives bad estimates is that it does not provide a direct  $D$ -matrix term. The sampled data has such a term due to the integrated sampling (its value is 1.17

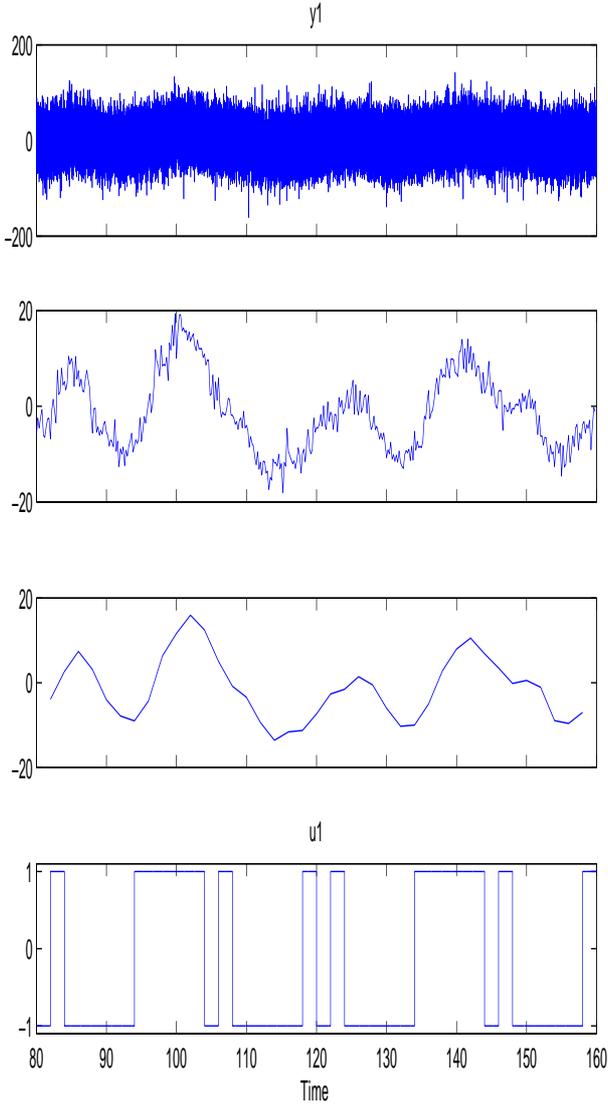


Fig. 3. The input output data. From top (1) “Continuous time output data”. (Generated by sampling interval 0.001.), Sampled output data, (2) sampling interval 0.2, (3) sampling interval 2, and (4) input data.

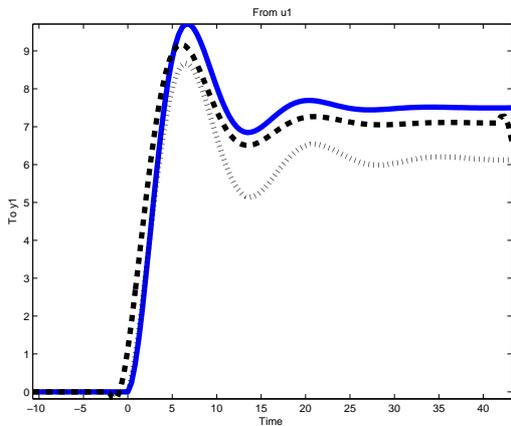


Fig. 4. Step response for identified continuous time system. Sampling interval 2. Solid: True system. Dotted: Estimate using simplistic sampling. Dashed: Estimate using simplistic sampling on anticausally shifted data.

Par	True	Est. Correct samp	Est. Simpl.samp
A(1,1)	-0.3567	-0.3611 ± 0.0151	-0.3357 ± 0.0143
A(2,1)	-0.2426	-0.2366 ± 0.0067	-0.2041 ± 0.0048
B(1,1)	0.2540	0.2623 ± 0.0897	-0.8392 ± 0.0440
B(2,1)	1.8198	1.7563 ± 0.0637	1.4489 ± 0.0505
K(1,1)	1.2865	1.1997 ± 0.2022	14.0573 ± 50.6497
K(2,1)	0.7864	0.6977 ± 0.0848	9.8242 ± 38.8295

Table 2. Estimated parameter for sampling interval  $T_s = 2$ .  $\pm$  indicates the estimated standard deviation

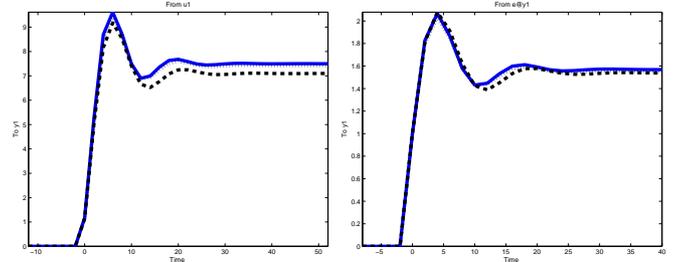


Fig. 5. Step response for identified discrete time system. Sampling interval 2. Solid: True system. Dotted: Estimate using correct sampling. Dashed: Estimate using simplistic sampling. Left: from input, Right: from noise source. The dotted lines coincide with the solid lines in this resolution.

for the 2 s sampling interval). The model suffers from not being able to explain this effect.

- The model fit is done in the metric of the sampled data, even though the parameters relate to continuous time. Comparing the discrete time responses of the models (according to their own sampling rules) with that of the correctly sampled system shows less discrepancies compared to the parametric errors, as seen in Figure 5. This is in accordance with the comments in Section 3.2.

## 6. ESTIMATING CONTINUOUS SYSTEMS FROM UNEQUALLY SAMPLED DATA

*Example 3.* Consider again the system in (20), simulated as before with a sample interval of 1ms to emulate a “continuous” system. The noise properties are the same as those in Example 2. Again produce 2560 seconds worth of such data, but this time sample at non-equidistant points in time, so that  $t_{k+1} - t_k$  is not constant over  $k$ . The average sample interval was roughly 2 seconds and the samples were distributed uniformly between 1.34 and 2.68 seconds (i.e. the ratio of largest to smallest sample interval is 2).

This “continuous” data was then sampled via (7), where we chose a constant integrating time  $\Delta_k$  equal to the minimum sample interval (1.34 seconds), i.e.  $\Delta_k = 1.34$  for all  $k$ . This corresponds to an assumption that the sampling device is not adaptive over time. This assumption can be easily removed. Figure 6 shows a snapshot of the continuous and sampled data.

With the above scheme we obtained  $N = 1280$  samples and based on these samples we estimated the parameters  $A(1,1)$ ,  $A(2,1)$ ,  $B(1,1)$ ,  $B(2,1)$ ,  $K(1,1)$  and  $K(2,1)$  of (20) based on the correct and simplistic sampling formulas. For the case of correct sampling, the Kalman filtering equations (12), (13) and (14) were used to form the predictor.

Figure 7 compares the step responses from correct and simplistic sampling models to the true model. By way of

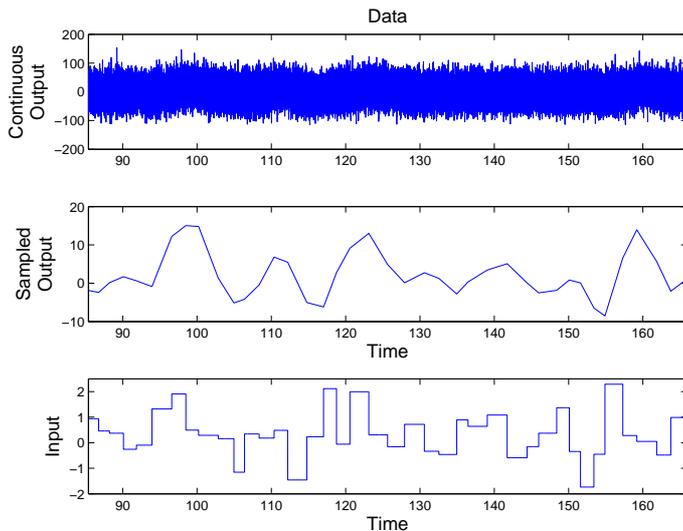


Fig. 6. Data for non-uniformly sampled example. Top: “continuous” output (0.001 second sample interval). Middle: sampled output (non-uniform sampling). Bottom: input.

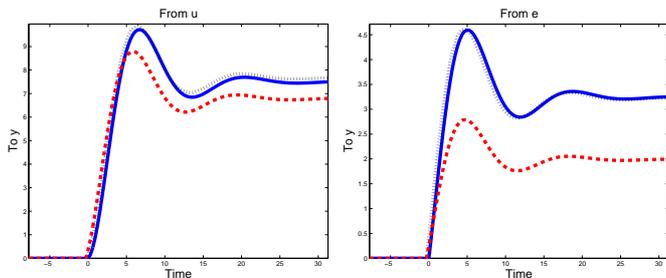


Fig. 7. Step response for identified continuous time system with non-uniformly sampled data. Solid: True system. Dotted: Estimate using correct sampling. Dashed: Estimate using simplistic sampling. Left: from input, Right from noise source.

information Table 3 shows the parameter values and their standard deviations for both sampling cases. Notice that the correct sampling gives more accurate results than the simplistic case, as expected.

Parameters	True	Correct Sampling	Simplistic Sampling
$A(1,1)$	-0.3567	$-0.3662 \pm 0.0133$	$-0.3713 \pm 0.0159$
$A(2,1)$	-0.2426	$-0.2436 \pm 0.0055$	$-0.2487 \pm 0.0076$
$B(1,1)$	0.2540	$0.2687 \pm 0.0822$	$0.9387 \pm 0.0882$
$B(2,1)$	1.8198	$1.8675 \pm 0.0500$	$1.6867 \pm 0.0621$
$K(1,1)$	1.2865	$1.4170 \pm 0.2072$	$0.7971 \pm 0.1070$
$K(2,1)$	0.7864	$0.7840 \pm 0.1068$	$0.4946 \pm 0.0570$

Table 3. Estimated parameter values for non-uniformly sampled example,  $\pm$  indicates standard deviations.

## 7. CONCLUSIONS

The main conclusion is that estimating continuous time systems requires some care and understanding of how the sampling was done. There is always some kind of averaging or lowpass filtering involved in the A/D conversion of measured data. If the sampling is fast compared to the system dynamics of interest this filtering/averaging effect may be negligible compared to the dynamics. We have shown in this contribution how to properly handle the

case of integrating sampling, which is one variant of real sampling. The effect of the estimated models may be significant for slow sampling. The use of continuous time models may be tied to an interest in physical models, “grey-box” identification. It could also be due to non-equidistantly sampled data, for which the continuous time model is the natural basis.

When dealing with a realistic sampling model, several issues arise in the choice of sampling interval  $T$ . A short interval gives higher variance of the output measurement, but this is basically compensated for by obtaining more measurements over a given time period. For any sampling formula to be correct it is also essential that the true, continuous time input can be correctly reconstructed from the sampled measurements. This may also influence the choice of  $T$ . Finally it is clear that the issues raised in this paper are also relevant for possible downsampling after the experiment.

## REFERENCES

- A. Feuer and G. C. Goodwin. *Sampling in Digital Signal Processing and Control*. Birkhauser, 1996.
- Markus Gerdin, Thomas Schön, Torkel Glad, Fredrik Gustafsson, and Lennart Ljung. On parameter and state estimation for linear differential-algebraic equations. *Automatica*, 43:416–425, March 2007.
- G. C. Goodwin, S. F. Graebe, and M. E. Salgado. *Control System Design*. Prentice Hall, Upper Saddle River, New Jersey, 2001.
- G.C. Goodwin, D. Q. Mayne, and A. Feuer. Duality of hybrid optimal regulator and hybrid optimal filter. *Int. J. Control*, 61(6):1465–1471, 1995.
- A. Jazwinski. *Stochastic Process and Filtering Theory*, volume 64 of *Mathematics in Science and Engineering*. Academic Press, New York, 1970.
- L. Ljung. *System Identification - Theory for the User*. Prentice-Hall, Upper Saddle River, N.J., 2nd edition, 1999.
- L. Ljung. *The System Identification Toolbox: The Manual*. The MathWorks Inc. 1st edition 1986, 7th edition 2007, Natick, MA, USA, 2007.
- Lennart Ljung and Torkel Glad. On global identifiability of arbitrary model parameterizations. *Automatica*, 30(2):pp 265–276, Feb 1994.
- K. J. Åström. *Introduction to Stochastic Control Theory*. Academic Press, New York, 1970.
- K. J. Åström. Maximum likelihood and prediction error methods. *Automatica*, 16:551–574, 1980.