LINEAR REGRESSION TECHNIQUES FOR STATE-SPACE MODELS WITH APPLICATION TO BIOMEDICAL/BIOCHEMICAL EXAMPLE

N. Khairudin, K.J. Keesman
Wageningen University, The Netherlands

Corresponding Author: K.J. Keesman, Systems and Control Group, Wageningen University
Bornsesteeg 59, 6708PD, Wageningen, The Netherlands; karel.keesman@wur.nl

Abstract. In this paper a novel approach to estimate parameters in an LTI continuous-time state-space model is proposed. Essentially, the approach is based on a so-called pqR-decomposition of the numerator and denominator polynomials of the system's transfer function. This approach allows the physical knowledge of the system to be preserved. As an illustrative example, a biomedical/biochemical process with two compartments in parallel and with first-order reaction is used. First, the process is approximated by a discrete-time state-space model. Next, after deriving the corresponding discrete-time transfer function, the rational transfer function is decomposed into pqR form and then reparametrized to obtain a set of linear regressive equations. Subsequently, the unknown linear regression parameters, which are a polynomial function of the original physical parameters, are uniquely estimated from real data of the biomedical/biochemical process using the ordinary least-squares method. This approach is favourable when there is a need to preserve physical interpretations in the parameters. Furthermore, by taking into account the original model structure, a smaller number of parameters than in the case of direct transfer function estimation may result and the identifiability property naturally appears.

1 Introduction

In literature, various parameter estimation methods have been proposed. Often, non-linear least squares methods are used to iteratively estimate parameters in models that are so-called non-linear in the parameters. However, unlike the ordinary least squares methods, the non-linear least squares methods do not guarantee a global minimum, especially in non-convex optimization problems [2]. Alternatively, for some problems it is possible to generate a linear-in-the-parameters model to approximate the system by applying a logarithmic transformation [5] or via reparametrization [3]. However, it is not always possible to apply these methods directly to a more complex model. It is also well-known that applying the classical linear regression technique on a general discrete-time LTI system poses a significant shortcoming, which is the loss of physical knowledge of the system. Usually, the estimation of unknown parameters is obtained solely by the use of input-out data relationship, as represented by the discrete-time transfer function [12,13], and thus black-box parameters result. Therefore, to preserve physical knowledge in the parameters to a large extend, a novel parameter estimation method via so-called pqR-decomposition [7] is proposed in this paper.

The main objective of this paper is to estimate parameters in an LTI state-space model structure while retaining the physical knowledge. The approach is illustrated to a biomedical/biochemical system with two compartments in parallel using real data. Firstly, the problem statement is defined in section 2. Next, the physical modelling of the biomedical/biochemical system is defined in section 3. In section 4, the methodology to obtain the exact linear regressive realization suitable for linear estimation via pqR-decomposition for the system is described. The estimation results and discussion are presented in section 5 and section 6, respectively. Finally, conclusions are drawn in section 7.

2 Problem statement

The following simple LTI continuous-time state-space model is used to illustrate some of the available methods that are currently used to estimate the parameters.

$$\sum(A,B,C) := \begin{cases} \dfrac{dx(t)}{dt} = & Ax(t) + Bu(t)\,, \quad x(0) = x_0 \\ y(t) = & Cx(t) \end{cases} \tag{1}$$

with $A = -a$, $B = 0$ and $C = 1$. Consequently, the analytical solution of equation (1) is given by $y(t) = x_0 e^{-\alpha t}$ with x_0 the initial condition. For this particular model, which in recursive form can also be written as $y(t) = y(t-1)e^{-\alpha \Delta t}$ with Δt the time step, a logarithmic transformation can be applied. Hence, the following linear regression is obtained.

$$\ln y(t) - \ln y(t-1) = -\alpha \Delta t \tag{2}$$

However, if $y(t) = x(t) + e(t)$ with $e(t)$ a noise term, the statistical properties of the transformed model might be different from those of the original model [5]. Alternatively, the equivalent discrete-time form, as shown above, may also be applied to compute the estimates via reparametrization as follows

$$\begin{aligned} y(t) &= y(t-1)e^{-\alpha \Delta t} \\ &= a_1 y(t-1) \end{aligned} \tag{3}$$

Hence, with $a_1 := e^{-\alpha \Delta t}$, equation (3) becomes a linear regression model. The value of a_1 can be estimated by applying an ordinary least-squares method. Finally, the physical parameter α can be found from $\hat{\alpha} = -\ln \hat{a}_1 / \Delta t$ where $\wedge$ denotes the estimate. Now, let us consider the following LTI SISO continuous-time state-space model structure,

$$\sum(A,B,C) := \begin{cases} \dfrac{dx}{dt} = \begin{pmatrix} a_1 & 1 \\ 1 & a_2 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u \\ y = \begin{pmatrix} 1 & 0 \end{pmatrix} x \end{cases} \tag{4}$$

Notice that for the estimation of the parameters a_1 and a_2 both the logarithm transformation and direct linear regression cannot be applied directly onto the system given by equation (4). Subspace identification [9,10] and non-linear least-squares estimation (NLS) are the two common methods used to estimate parameters in matrix A. However, applying subspace identification usually leads to a black-box state-space structure, where the use of NLS [2,3] may result in a local minima if incorrect initial estimates are specified. Furthermore, the NLS may requires substantial computational effort when a multi-start procedure is applied to find the global minimum. Thus, the aim of our study is to estimate the parameters in A and B using linear regression techniques. This paper is greatly inspired by the work on rational systems and distributed parameter systems, as presented in [3,6, 11-13].

3 Physical modelling

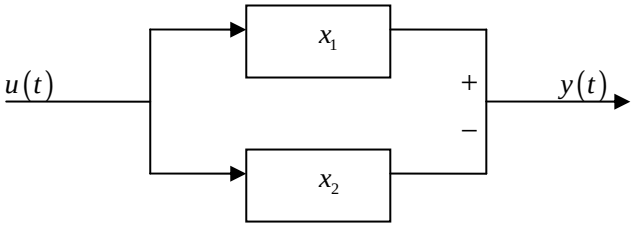

Figure 1. Two-compartmental biomedical/biochemical system.

Consider a biomedical/biochemical example that consists of two compartments in parallel with first-order reaction kinetics, as shown in Figure 1. In [8], the response to this biomedical/biochemical system is defined by a two exponential model, with initial and final output zero. The model response is given by $y(t) = ce^{a_1 t} - ce^{a_2 t}$, in which c is the initial concentration and a_1 as well as a_2 are the decay constants of the exponentials in compartment 1 and 2, respectively. This equation can also be realized by two ordinary differential equations with one observation equation. Furthermore, it is assumed that the same amount of input is supplied to both of the compartments. Thus, the concentration in the two compartments can be described by $dx_1/dt = a_1 x_1 + b_1 u$ and $dx_2/dt = a_2 x_2 + b_1 u$, respectively with $u(t)$, an impulsive input. The initial concentration in both compartment is given by $x_1(0) = x_2(0) = 0$. The output of the system is the difference between the two compartment outputs so that, $y = x_1 - x_2$. Alternatively, these first-order linear differential equations can be written into an LTI SISO continuous-time state-space form as follows

$$\sum(A,B,C) := \begin{cases} \dfrac{dx}{dt} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} x + \begin{pmatrix} b_1 \\ b_1 \end{pmatrix} u \\ y = \begin{pmatrix} 1 & -1 \end{pmatrix} x \end{cases} \tag{5}$$

where the impulse input is defined as $u(t)=\delta(t)$. Consequently, $y(t)=b(e^{a_1 t}-e^{a_2 t})$. In the next section, this physical system will be decomposed into pqR form to allow a ordinary least-squares estimation of a_1, a_2 and b_1.

4 Linear regressive realization via pqR-decomposition

The main concept of pqR-decomposition involves splitting the transfer function $G(\theta,q)$ of a system into a numerator polynomial $\tilde{N}(\theta,q)$ and a denominator polynomial $\tilde{M}(\theta,q)$ with q the forward-shift operator. To illustrate this, the continuous-time state-space model given by equation (5) is first approximated by a discrete-time state-space model using an Euler discretization scheme as follows

$$\sum_d := \begin{cases} x(k+1) = & \left(I + A(\vartheta)\Delta t\right)x(k) + B(\vartheta)\Delta t u(k) \\ y(k) = & Cx(k) \end{cases} \tag{6}$$

where k is the time index. Note that matrices A and B depend on the physical parameter vector ϑ, which contains a_1, a_2 and b_1. The corresponding transfer function of this system is given by

$$G(\theta,q)=\frac{y(k)}{u(k)}=C\left[R(A,q)\right]B(\vartheta)\Delta t \tag{7}$$

where $R(A,q) := \left[qI-\left(I+A(\vartheta)\Delta t\right)\right]^{-1}=\dfrac{\operatorname{adj}R(A,q)}{\det R(A,q)}$ and ϑ a parameter vector that contains combinations of a_1, a_2 and b_1. As will be shown later, $\theta(a,b)$ is a polynomial function in a_1, a_2 and b_1. The observation vector C is filled with an arrangement of constants and zeros. Thus, the transfer function $G(\theta,q)$ can be split into denominator and numerator polynomials as follows

$$G(\theta,q) := \frac{C\left[\operatorname{adj}R\left(\tilde{A},q\right)\right]B(\theta)\Delta t}{\det R\left(\tilde{A},q\right)}=\frac{\tilde{N}(\theta,q)}{\tilde{M}(\theta,q)} \tag{8}$$

where $\tilde{A}=I+A(\vartheta)\Delta t$. Applying this concept to the biomedical/biochemical system results in $\tilde{M}(\theta,q) := (-q+1+a_1\Delta t)(-q+1+a_2\Delta t_2)$ and $\tilde{N}(\theta,q) := b_1\Delta t^2(a_1-a_2)$. Next, these polynomials are further split into a pqR-decomposition:

$$\tilde{p}_M{}^T M\psi(q)y_k = \tilde{p}_N{}^T N\psi(q)u_k \tag{9}$$

where $\tilde{p}_M{}^T := \begin{pmatrix} p_M{}^T & 1 \end{pmatrix}$ and $\tilde{p}_N{}^T = \begin{pmatrix} p_N{}^T & 1 \end{pmatrix}$ are the parameter vectors , M and N are the regression weighting matrices and $\psi(q)y_k$, $\psi(q)u_k$ are the output and input data vector, respectively. Applying this decomposition to the biomedical/biochemical system results in

$$\tilde{M}(\theta,q) := \begin{pmatrix} a_1a_2 \\ a_2 \\ a_1 \\ 1 \end{pmatrix}^T \underbrace{\begin{pmatrix} \Delta t^2 & 0 & 0 \\ \Delta t & -\Delta t & 0 \\ \Delta t & -\Delta t & 0 \\ 1 & -2 & 1 \end{pmatrix}}_{M} \begin{pmatrix} q^0 \\ q^1 \\ q^2 \end{pmatrix} y_k \tag{10a}$$

$$\tilde{N}(\theta,q) := \begin{pmatrix} b_1a_1 \\ b_1a_2 \\ 1 \end{pmatrix}^T \underbrace{\begin{pmatrix} \Delta t^2 & 0 & 0 \\ -\Delta t^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{N} \begin{pmatrix} q^0 \\ q^1 \\ q^2 \end{pmatrix} u_k \tag{10b}$$

Then, substituting equation (10a) and equation (10b) into equation (8), while using equation (7), yields

$$\begin{pmatrix} a_1a_2 \\ a_2 \\ a_1 \\ 1 \end{pmatrix}^T \begin{pmatrix} \Delta t^2 & 0 & 0 \\ \Delta t & -\Delta t & 0 \\ \Delta t & -\Delta t & 0 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} q^0 \\ q^1 \\ q^2 \end{pmatrix} y_k = \begin{pmatrix} b_1a_1 \\ b_1a_2 \end{pmatrix} \begin{pmatrix} \Delta t^2 & 0 & 0 \\ -\Delta t^2 & 0 & 0 \end{pmatrix} \begin{pmatrix} q^0 \\ q^1 \\ q^2 \end{pmatrix} u_k \tag{11}$$

Equation (11) can be rearranged in the following linear regression form

$$\begin{pmatrix} p_N{}^T & p_M{}^T & 1 \end{pmatrix} \begin{pmatrix} N & 0 \\ 0 & -M \end{pmatrix} z_k = 0 \tag{12}$$

where $z_k = \begin{pmatrix} u_k & u_{k+1} & u_{k+2} & y_k & y_{k+1} & y_{k+2} \end{pmatrix}^T$. Finally this results in

$$\begin{pmatrix} b_1a_1 \\ b_1a_2 \\ a_1a_2 \\ a_2 \\ a_1 \end{pmatrix}^T \begin{pmatrix} \Delta t^2 & 0 & 0 & 0 & 0 & 0 \\ -\Delta t^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\Delta t^2 & 0 & 0 \\ 0 & 0 & 0 & -\Delta t & \Delta t & 0 \\ 0 & 0 & 0 & -\Delta t & \Delta t & 0 \end{pmatrix} \begin{pmatrix} q^0u_k \\ q^1u_k \\ q^2u_k \\ q^0y_k \\ q^1y_k \\ q^2y_k \end{pmatrix} = y_k - 2y_{k+1} + y_{k+2} \tag{13}$$

From equation (13), it can be pointed out that unidentifiability of the physical parameters occurs. Thus, the polynomials as a result of expanding equation (13) are reparametrized as follows

$$\begin{aligned} y_k - 2y_{k+1} + y_{k+2} &= \underbrace{\left(b_1a_1 - b_1a_2\right)}_{\theta_1}\Delta t^2u_k + \underbrace{\left(a_1 + a_2\right)}_{\theta_2}\Delta t\left(y_{k+1} - y_k\right) - \underbrace{a_1a_2}_{\theta_3}\Delta t^2y_k \\ &= \theta_1\Delta t^2u_k + \theta_2\Delta t\left(y_{k+1} - y_k\right) - \theta_3\Delta t^2y_k \end{aligned} \tag{14}$$

Re-writing equation (14) in the form of equation (12) results in a linear regressive set of equations as follows

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}^T \begin{pmatrix} \Delta t^2 & 0 & 0 \\ 0 & -\Delta t & \Delta t \\ 0 & -\Delta t^2 & 0 \end{pmatrix} \begin{pmatrix} q^0u_k \\ q^0y_k \\ q^1y_k \end{pmatrix} = y_k - 2y_{k+1} + y_{k+2} \tag{15}$$

By defining $\tilde{y}_k = y_k - 2y_{k+1} + y_{k+2}$, $\phi_k := \begin{pmatrix} \Delta t^2u_k & \Delta t\left(y_{k+1} - y_k\right) & -\Delta t^2y_k \end{pmatrix}$ and $\theta = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \end{pmatrix}^T$, we obtain a linear regression. Hence, θ can be estimated via the ordinary linear-squares method using observations from a methionine tolerance test [1], see Table 1. Since, it is common for many biomedical/biochemical systems, that the dynamic model is based on sparse data from non-repeatable experiments [1], the raw observations are interpolated to generate more samples. Data in italic font in Table 1 are the results of linear interpolation from the available observations to obtain equidistant data.

Table 1. Raw observations and interpolated data of a biomedical/biochemical process [1].

t (hour)	0	*0.25*	0.50	0.75	*1.00*	1.25	*1.50*	1.75	*2.00*	2.25
$y(t)$ (micromoles/litre)	0	*45*	90	115	*100*	85	*70*	55	*47.5*	40
$u(t)$	1	*0*	0	0	*0*	0	*0*	0	*0*	0

Then, if needed the data can even be further interpolated to predict values of observations, y, and input, u, at specific time instances. In addition to this step of 0.25 hour, as shown in Table 1, we also investigate the effect of a smaller time step Δt = 0.05 hour. Estimation results are presented in the next section.

5 Results

After applying the ordinary least-squares method to the set of linear regressive equations, we obtain results as shown in Figure 2. The solid line represents the predicted model output, while each cross represents the interpolated observations from the methionine tolerance test. The small circle represents real data from the methionine tolerant test obtained from Norton [8]. Figure 2(a) presents the results using a time interval with $\Delta t = 0.25$ hour.

Figure 2(b) shows the results related to a smaller time interval, $\Delta t = 0.05$ hour. Plots of the residuals for both cases are presented in Figure 3. The mean values of the residuals are 0.8407 and 0.1222, respectively. By looking at Figure 2, this method gives a good curve fit. However, from Figure 3(a) and Figure 3(b) it can be pointed out there are multiple points where the residuals deviate from zero, significantly. This is due to the change in direction (positive or negative) of the slope when a new (interpolated) observation becomes available. The peaks in both Figure 2(a) and Figure 2(b) cannot be accurately estimated because the linear regression structure utilizes the previous value to predict the next value. Finally, it can be concluded, that this method gives a good curve fit for this biomedical/biochemical system, especially when the sampling time is set smaller after applying linear interpolation.

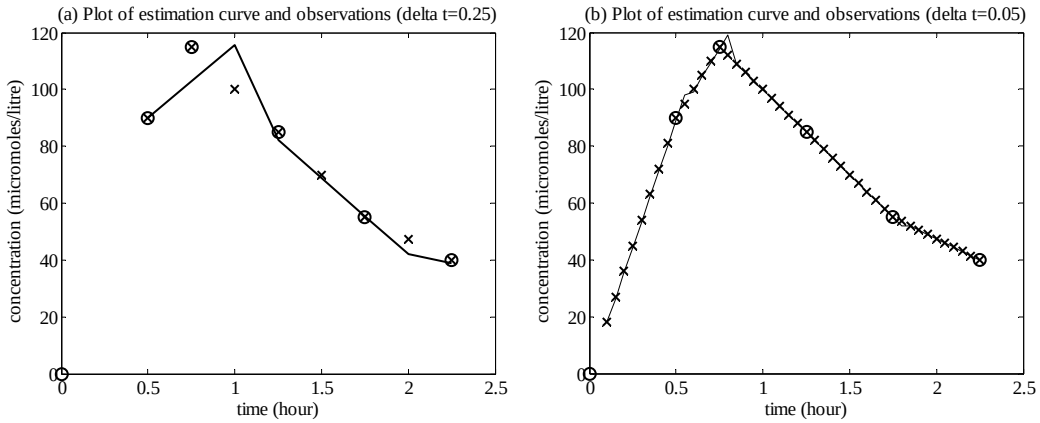

Figure 2. Estimation curve versus observations for – (a) sampling time 0.25 hour and (b) sampling time 0.05 hour.

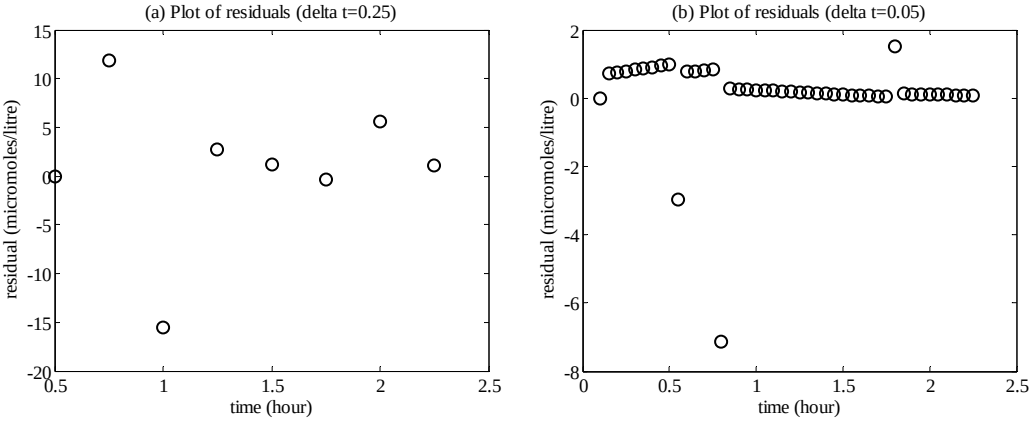

Figure 3. Plot of residuals for – (a) sampling time 0.25 hour and (b) sampling time 0.05 hour.

For Δt = 0.25 hour and after using equation (15), the final estimates are given by: $\theta = \begin{pmatrix} 436.5227 & -2.4251 & 1.6489 \end{pmatrix}^T$. Hence, these values can be used to predict the output. However, if we are interested in the physical parameter estimates, another step is needed. To obtain the physical parameter estimates, the parameter relationships in equation (14) are solved using the estimates of θ. Because of the first-order interaction ($a_1 a_2$), two solutions appear. However, one solution contains a negative b_1 and is thus discarded. The other solution is: a_1 = -1.2126-0.4226i , a_2 = -1.2126+0.4226i and $b_1 = 516.4268i$. For Δt = 0.05 hour, the final estimates are given by $\theta = \begin{pmatrix} 270.5102 & -1.5028 & 1.7470 \end{pmatrix}^T$. Again, parameter relationships in equation (14) are solved similar to the previous step and results in: a_1 = -0.7514-1.0874i , a_2 = -0.7514+1.0874i and $b_1 = 124.3861$i . Unfortunately, these solutions are not physically feasible due to the imaginary values, which are probably caused by the linear interpolation of the measurements. Hence, we will determine how sensitive the estimation result is with respect to the measurements, in the next analysis.

Let us investigate the following possibilities, i.e. $y(0.25) = 0, 5, 10, 15, ..., 85$ and repeat the linear regression estimation method via pqR-decomposition for $\Delta t = 0.25$ hour. Results are shown in Table 2.

Table 2. Linear regression estimation technique via pqR-decomposition $\Delta t = 0.25$ hour.

$y(t) = y(0.25)$ (micromoles/litre)	θ_1	θ_2	θ_3	a_1	a_2	b_1
0	1440.0	-3.1	2	-0.9156	-2.1844	1134.8790
5	1340.2	-3	2	-1	-2	1340.2
10	1237.7	-2.9	2	-1.1298	-1.7702	1932.9627
15	1132.4	-2.9	1.9	-1	-1.9	1258.2222
20	1024.1	-2.8	1.9	-1.1551	-1.6449	2090.4354
25	912.5439	-2.7254	1.8675	-1.3627-0.1027i	-1.3627+0. 1027i	4442.4658i
30	797.7738	-2.6481	1.8208	-1.3241-0.2602i	-1.3241+0. 2602i	1533.1420i
35	679.8742	-2.5705	1.7690	-1.2853-0.3422i	-1.2853+0. 3422i	993.2532i
40	559.2044	-2.4950	1.7117	-1.2475-0.3943i	-1.2475+0. 3943i	709.1759i
45	436.5227	-2.4251	1.6489	-1.2126-0.4226i	-1.2126+0. 4226i	516.4268i
50	313.1893	-2.3659	1.5805	-1.1829-0.4256i	-1.1829+0. 4256i	367.9447i
55	191.4322	-2.3247	1.5075	-1.1624-0.3955i	-1.1624+0. 3955i	241.9957i
60	74.6636	-2.3111	1.4315	-1.1556-0.3102i	-1.1556+0.3102i	120.3599i
65	-32.2214	-2.3376	1.3554	-1.2722	-1.0654	155.7958
70	-122.7122	-2.4189	1.2842	-1.6320	-0.7869	145.1958
75	-188.9868	-2.5700	1.2246	-1.9382	-0.6318	144.6699
80	-222.9857	-2.8032	1.1850	-2.2845	-0.5187	126.2827
85	-218.4403	-3.1222	1.1743	-2.6848	-0.4374	97.1957

Theoretically and by referring to Norton [8, p.81], b_1 must be positive, while a_2 and a_1 must have negative values. In addition to this, a_2 must be larger than a_1 in term of magnitude in order to get the correct estimation curve. Clearly, it can be observed from Table 2 that the estimates of a_1, a_2 and b_1 are unrealistic when $y(0.25) \geq 25$ micromoles/litre. Notice from Figure 2 and Figure 5 that, despite the physically unrealistic estimates of a_1, a_2 and b_1, the estimation curve is still able to follow the observations.

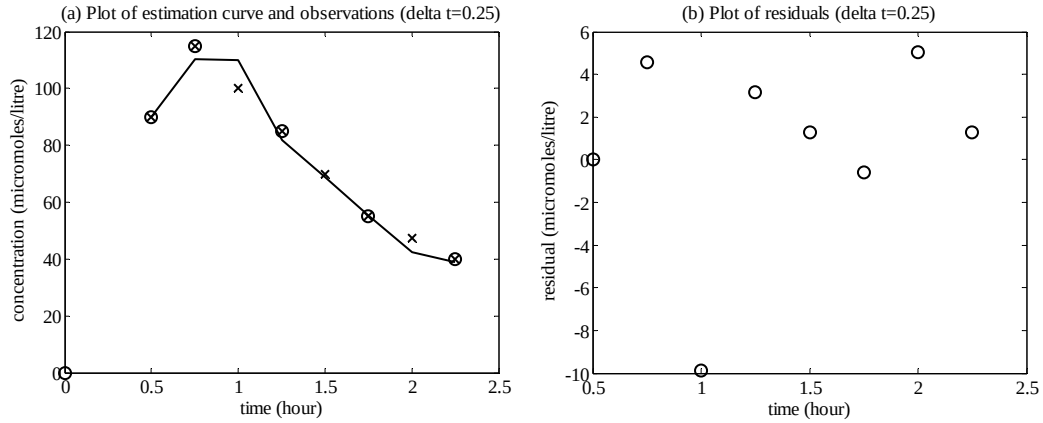

Figure 4. Estimation curve and residuals plot when $y(0.25) = 5$ for $\Delta t = 0.25$.

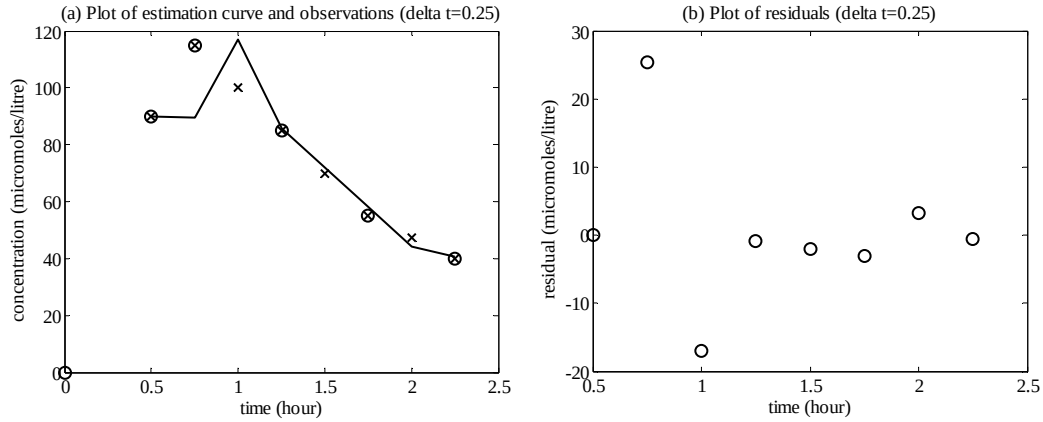

Figure 5. Estimation curve and residuals plot when $y(0.25) = 75$ for $\Delta t = 0.25$.

Our hypothesis is that the infeasible estimation results are due to the linear interpolation. Our final effort to identify the actual cause which leads to imaginary and negative estimation results is by examining the analytical solution of $\theta_1 = b_1 a_1 - b_1 a_2$, $\theta_2 = a_1 + a_2$ and $\theta_3 = a_1 a_2$ (see equation (14)). The solutions are given by

$$a_1 = -\frac{1}{2}\theta_2 \pm \frac{1}{2}\sqrt{\theta_2^2 - 4\theta_3}$$

$$a_2 = -\frac{1}{2}\theta_2 \pm \frac{1}{2}\sqrt{\theta_2^2 - 4\theta_3}$$

$$b_1 = \pm \frac{\theta_1}{\sqrt{\theta_2^2 - 4\theta_3}}$$

It can be observed that imaginary numbers appear if $\theta_3 > {\theta_2}^2 / 4$. The value of θ_1 only affects the magnitude of b_1, but it will never lead to an imaginary number.

6 Discussion

Applying linear regressive parameter estimation via pqR-decomposition gives a good curve fit between the predicted model output and the corresponding observations. This is especially true when the time interval is small. The least-squares estimation method minimizes the cost function which helps to fit the estimation curve to the observation points. On top of that, the linear regression estimation method also involves updating the next point using the previous point. This further ensures a good curve fit, even in the cases where the physical parameter estimates obtained are unrealistic as shown in Figure 2 and Figure 5. More importantly, the physical structure of the system is directly embedded in the linear regression model via the pqR-decomposition.

In an estimation study of a diffusion process, Vries [11] showed that a successful reconstruction of the physical parameters is possible. Unfortunately, for this particular study, we are not able to directly estimate the physical parameters via linear regressive parametric realization of the approximate discrete-time system, as we anticipated. Perhaps, this would be possible if more observations were available from the experiment.

However, the investigation on the sensitivity of the estimation result with respect to the measurements, led to interesting findings. It is found that $y(0.25)$ must be smaller than or equal to 20 micromoles/litre in order to obtain physically interpretable results, as shown in Table 2. This suggests that a time-delay may exist at the beginning of the experiments. It is known that a rapid oral dose of methionine in the proportion of 0.01mg/kg of body mass is administered into the subject [1]. Hence, it is likely that the administering of oral dose is causing the time-delay.

Hence, our suggestion is to take measurements at smaller sampling time especially at the beginning of the experiment. However, measuring more samples may result in higher cost for most biological experiments [1]. It is also unfortunate that in this study, the methionine tolerant test is non-repeatable. Thus, in cases such as this, our suggestion is to roughly and intuitively predict the missing points. Another possible solution is to generate more observation points which closely follow the measured data, using an appropriate model with initial guesses of the physical parameters.

It is also important to realize that linear regression estimation via pqR-decomposition takes into account the physical structure of the model. The pqR-decomposition is a direct realization of the physical model with physical interpretation. By referring to the pqR-decomposition, only three parameters are estimated after reparametrization. Typically, estimation of a corresponding transfer function related to an *n*-compartment model leads to $2n+1$ parameters. Notice that the pqR-decomposition splits the physical parameters from the numerical scheme used for discretization. One advantage of this is that we can easily check identifiability by only looking at the regression weighting matrices M and N. For the continuous-time case without a discretization step, we may also use the Laplace transform to obtain a transfer function [4]. Similarly, a pqR-decomposition, as presented in this abstract, of the transfer function $G(\theta, s)$, with *s* the Laplace variable, can be applied.

7 Concluding remarks

The novel method of pqR-decomposition leads to a physically interpretable linear regression structure from a discrete-time transfer function of an LTI system. Hence, using ordinary least-squares techniques, unique linear regression parameters, which are polynomial functions of the physical parameters, can be found. This finally results in an unbiased physical model-based predictor. However, as shown in this study, the linear regression parameters do not automatically lead to realistic physical parameter estimates. Nevertheless, the study of when

realistic estimates do occur reveals that an appropriate choice of interpolated data points is crucial and in this particular application it suggests a plausible time delay.

The linear regression realization approach, as presented in this paper, is subject to further research. In particular, other model classes and applications with suitable and sufficient experimental data will be investigated.

8 References

[1] Brown, R.F., Godfrey, K.R. and Knell, A.: *Compartmental Modelling based on Methionine Tolerance Test Data: A Case Study*. Med. & Biol. Eng. & Comput., 1979, 17, 223 - 229.

[2] Doeswijk, T.G.: *Reducing Prediction Uncertainty of Weather Controlled Systems.* PhD thesis, Wageningen Universiteit, Wageningen, The Netherlands, 2007, 9 - 10.

[3] Doeswijk, T.G. and Keesman, K.J.: *Linear Parameter Estimation of Rational Biokinetic Functions.* Water Research. Accepted for publication 2008.

[4] Godfrey, K.: *Compartmental Models and Their Application*. Academic Press, Inc., New York, 1983, 94 - 97.

[5] Johansson, R.: *System Modeling and Identification*. Prentice Hall, Englewood Cliffs, NJ, 1993.

[6] Keesman, K.J. and Doeswijk, T.G.: *Direct Least-squares Estimation and Prediction of Rational Systems*: *Application to Food Storage*. Journal of Process Control. Accepted for publication 2008.

[7] Keesman, K.J. and Khairudin, N.: *Linear regressive realization of LTI state-space models.* Submitted to: Proceedings of 15th IFAC Symposium on System Identification, 2009.

[8] Norton, J.P.: *An Introduction to Identification.* Academic Press Inc., Orlando, Florida, 1986.

[9] Van Overschee, P.: *Choice of State-Space Basis in Combined Deterministic-Stochastic Subspace Identification*. Automatica, 1995, vol 31(12), 1877-1883.

[10] Verhaegen, M. and Dewilde, P.: *Subspace model identification. Part 1: The output-error state-space model identification class of algorithms*. Int. J. Control, 1992, 56, 1187 - 1210.

[11] Vries, D.: *Estimation and Prediction of Convection-Diffusion-Reaction Systems from Point Measurements*. PhD thesis, Wageningen Universiteit, Wageningen, The Netherlands, 2008, 43 - 64.

[12] Vries, D., Keesman K.J. and Zwart, H.: *Linear Regressive Model Structures for Estimation and Prediction of Compartmental Diffusive System*. In: Proceedings of 5th International IMACS Symposium on Mathematical Modelling, Vienna, Austria, 2006.

[13] Vries D., Keesman, K.J. and Zwart, H.: *Explicit linear regression model structures for estimation, prediction and experimental design in compartmental diffusive system*. In: Proceedings of 14th IFAC Symposium on System Identification, Newcastle, Australia, 29th-31th March 2006.