Input estimation as a qualitative trend analysis problem

Christian Thürlimann[a,b], Kris Villez[a]

[a] *Eawag, Department Process Engineering, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland*
[b] *Institute of Environmental Engineering, ETH Zürich, CH-8093 Zürich, Switzerland*

Abstract

The study of techniques for qualitative trend analysis (QTA) has been a popular approach to address challenges in fault diagnosis of engineered processes. Such challenges include the lack of reliable extrapolation of available models and lack of representative data describing previously unseen circumstances. Many of these challenges appear in biological systems even when normal operation can be assumed. It is for this reason that QTA techniques have also been proposed for the purpose of fault detection, automation, and dynamic modeling. In this work, we adopt a shape-constrained spline function method for the purpose of unknown input estimation. Thanks to data collected at laboratory-scale in a biological reactor for urine nitrification, this novel approach has been demonstrated successfully for the first time.

Keywords: global optimization, input estimation, oxygen uptake rate, qualitative trend analysis, wastewater treatment

*Kris Villez
Email address: kris.villez@eawag.ch (Kris Villez)

Acronyms

Acronym	Full expression
DO	dissolved oxygen
LTI	linear time-invariant
MHE	moving horizon estimation
OUR	oxygen uptake rate
SCS	shape-constrained splines
QTA	qualitative trend analysis

1. Introduction

Routine execution of on-line process data analysis is a challenging task for many processes. The use of models to extract valuable information from the available data is often known as *soft-sensing* and several such methods for have been developed. Widely-known methods include the Kalman filter and its extensions (e.g., Romanenko & Castro, 2004; Kravaris et al., 2013; Prakash et al., 2014). These techniques provide a systematic approach to the construction of such soft-sensors on the basis of dynamic process models. Factors affecting the success include the completeness of available process understanding, whether or not measured variables include or describe the key process states comprehensively, and whether the process undergoes important changes over time. To obtain a useful model, two modeling approaches are distinguished. The first consists of white-box modeling and is based models which reflect the mechanistic understanding of the process. Successful application of soft-sensors based on white-box models requires completeness, accuracy, and precision of the applied model. If this is not met, systematic deviations, i.e. bias, should be expected between the extracted estimates and their true values. When a reliable white-box model is not available, one may choose to take the black-box route. In this case, one uses historical data to empirically define the relationships between *(i)* data that is available cheaply and reliably and *(ii)* information that is difficult to obtain directly. Unfortunately, many black-box models (e.g., neural nets, regression trees, support vector machines) lack transparency. As a result, such models may not be trusted to provide information for safety- or quality-critical decisions (see e.g., Liu, 2007; Wang et al., 2010). In addition, black-box models often suffer from large estimation errors when extrapolated. Choosing between white-box and black-box approaches often entails a trade-off between these aspects. Quite naturally, several authors have proposed a mixed approach, i.e. grey-box modeling, to represent the process mechanistically in as much as possible while representing the lesser known parts of the process as a black-box model.

In a number of situations, one may simultaneously lack detailed process understanding as well as sufficient data to properly define any of the traditional models described above. This is true for many processes and has led to the development and application of coarse-grained qualitative modeling and simulation techniques (Venkatasubramanian et al., 2003). Such methods are deliberately imprecise which leads to predictions that can be trusted (reliability) despite large uncertainties. Despite this imprecision, this still enables causal reasoning and decision-making, e.g. (e.g., Kuipers, 1989; Maurya et al., 2003; Bredeweg et al., 2009; Kansou & Bredeweg, 2014). In the process engineering literature, the qualitative approach has been advocated mainly for the purpose of fault diagnosis and is primarily implemented in the form of qualitative trend analysis (QTA, Bakshi & Stephanopoulos, 1994; Rengaswamy & Venkatasubramanian, 1995; Dash et al., 2004; Charbonnier et al., 2005; Gamero et al., 2006; Charbonnier & Gentil, 2007; Maurya et al., 2010; Villez et al., 2012, 2013; Gamero et al., 2014). The main motivation is that both process understanding and data describing the dynamics of processes subject to rare events are typically extremely limited. The same can often be said even for normal conditions for complex biological processes. When so, qualitative approaches also become valuable outside of the fault diagnosis niche, e.g. for process data mining (Stephanopoulos et al., 1997; Villez et al., 2007). More recent work has pushed the application boundary even further by enabling fault detection (Villez & Habermacher, 2016), image analysis (Derlon et al., 2017), model structure identification (Mašić et al., 2017), data reconciliation (Srinivasan et al., 2017), and process automation (Villez et al., 2008; Thürlimann et al., 2015) on the basis of the QTA philosophy.

Existing methods for QTA are useful to describe the qualitative features (e.g., maxima, minima, inflection points) of a data series. In contrast, we provide a new approach to QTA which describes the qualitative features of a process input signal which cannot be measured directly. To this end, the process itself is represented by a piece-wise linear time-invariant (LTI) model. The analyzed measurement data series is assumed to be univariate, which is typical in the QTA literature apart from a few exceptions (e.g., Flehmig & Marquardt, 2006, 2008). The unknown input signal is represented as a shape constrained spline function. Estimating the parameters of this input signal, i.e. the spline coefficients, by fitting the complete model to process data forms the focus of this study.

The method is applied for estimation of the oxygen uptake rate in an intermittently fed stirred tank reactor for urine nitrification (Udert & Wächter, 2012; Fumasoli et al., 2016). This process has been developed as part of a system to recover resources, in this case a fertilizer, from source-separated wastewater streams. In the urine nitrification process, the oxygen uptake rate (OUR) reflects the respiration rate of the ammonia oxidizing bacteria and the nitrite oxidizing bacteria in the process. One aims to achieve a low respiration rate at the end of each cycle, i.e. right before new untreated urine is fed to the reactor. Estimates of the OUR can thus be used to maximize the efficiency of the process. This is very similar to conventional aerobic sequencing batch reactors for wastewater treatment (e.g., Yoong et al., 2000). Estimates of the OUR are also essential for wastewater characterization (e.g., Spanjers & Vanrolleghem, 1995; Spérandio & Etienne, 2000; Choubert et al., 2013), model identification (e.g., Vanrolleghem & Spanjers, 1998; Petersen et al., 2001; Ferrai et al., 2010), and automation (e.g., Spanjers et al., 1996; Yoong et al., 2000; Gernaey et al., 2001). Most typically, one obtains the OUR at infrequent time points by fitting a linear line to a short series of dissolved oxygen concentration measurements obtained during an unaerated phase. The underlying idea is that the oxygen measurement series are described well by a linear trend, whose slope reflects the respiration rate in the selected time window. This approach means that the OUR is not available continuously and that nonlinear effects of aeration and sensor dynamics are deliberately ignored. With the proposed method, these assumptions are not necessary and the OUR is available as a continuous process input estimate. In addition, the method allows estimating the kinetic parameters of the aeration system and the sensor simultaneously, thus providing additional information regarding the state of the components of the monitored system. We demonstrate the method with data obtained in a single batch cycle and describe the opportunities that lie ahead.

2. Materials and Methods

All symbol definitions required in this text are given in Table 2.

Table 2: Symbol definitions

Symbol	Description
Θ	Feasible set for $\boldsymbol{\theta}$
Ω	Feasible set for $\boldsymbol{\beta}$
$\boldsymbol{\beta}$	Spline function coefficients
δ_k	Input noise at knot k
ϵ_i	Measurement error at sample i
σ_δ	Input noise standard deviation
σ_ϵ	Measurement error standard deviation
$\boldsymbol{\tau}$, τ_c, τ_y	Time constants (for concentration, for measurement)
$\boldsymbol{\theta}$	Transitions
D	Degree of the spline function
E	Number of episodes
I	Total number of samples
K	Number of spline knots
S	Number of process states

$\boldsymbol{S}$	Matrix describing the shape constraints
T	Number of transitions
i	Measurement sample index
k	Spline index
$\boldsymbol{a}_t$	Spline basis function evaluated at t
$\boldsymbol{c}_t$	Convoluted spline basis function evaluated at t
c_{DO}	Dissolved Oxygen (state)
$\underline{\boldsymbol{b}}$, $\underline{b}$	Left-side interval bounds
$\overline{\boldsymbol{b}}$, $\overline{b}$	Right-side interval bounds
d	derivative index
e	episode index
$\boldsymbol{f}$	Rate of change
$\boldsymbol{g}$	Measurement gains
r_{OUR}	oxygen uptake rate (OUR)
$\boldsymbol{s}_0$	Initial state vector
$\boldsymbol{s}$	State vector
t, t_i	Time (at sample i)
$\boldsymbol{u}$, u	Known binary input
$v_0^{(d)}$	Initial values for the unknown process input signal
v, $v^{(d)}$	Unknown process input (dth derivative)
w	Integrand
y	Measurement
y_{DO}	Dissolved oxygen (noise-free measurement)
$\tilde{y}_{DO}$	Dissolved oxygen measurement

2.1. Basic model

Data-generating model – Theory. In this work, we aim to describe measurement time series of finite length with the following generative model:

$$\dot{\boldsymbol{s}} = \boldsymbol{f}_t(\boldsymbol{s}, \boldsymbol{u}, v) \tag{1}$$

$$\tilde{y}_i = \boldsymbol{g}^T \ \boldsymbol{s}(t_i) + \epsilon_i \tag{2}$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon) \tag{3}$$

with $\boldsymbol{s} = \boldsymbol{s}(t)$, $\boldsymbol{u} = \boldsymbol{u}(t)$, $v = v(t)$.

The above model is a continuous-time state-space model composed of a set of ordinary differential equations which generates noisy measurements ($\tilde{y}_i$) at distinct sampling times (t_i, $i = 1, \ldots, I$). We further assume that *(i)* the ordinary differential equations are piece-wise LTI in the S state variables ($\boldsymbol{s}$) and the uncontrolled input ($v(t)$), and *(ii)* that the controlled inputs ($\boldsymbol{u}(t)$) are piece-wise constant. In what follows, the parameters of the piece-wise linear LTI system are given as a vector $\boldsymbol{\tau}$.

The univariate input ($v(t)$) is assumed to be described well by a signal consisting of K piece-wise polynomial segments of degree D. Each kth polynomial starts at time t_k and ends at time t_{k+1} ($t_1 = 0, t_k < t_{k+1}, t_{K+1} = t_I, k = 1, ..., K$). Every derivative up to the $D-1$th derivative of this signal is continuous over the whole domain ($0 \leq t \leq t_I$). Such a signal is obtained by simulating the following model:

$$\dot{v}(t) = \dot{v}^{(0)}(t) = v^{(1)}(t),\ v^{(0)}(0) = v_0^{(0)} \tag{4}$$

$$\dot{v}^{(d)}(t) = v^{(d+1)}(t),\ v^{(d)}(0) = v_0^{(d)},\ 1 \leq d < D \tag{5}$$

$$v^{(D)}(t) = \delta(t) = \delta_k,\ t_k \leq t < t_{k+1} \tag{6}$$

In the above, δ_k determines the Dth derivative in the kth segment and can be interpreted as a piece-wise constant input disturbance. The values for $v_0^{(d)}$ ($0 \leq d \leq D-1$) are the initial conditions for the signal and its derivatives. If δ_k is a white noise signal then the simulated signal $v(t)$ is an auto-correlated signal. $v(t)$ is defined completely by the δ_k sequence ($k = 1, \ldots, K$) and the initial conditions $v_0^{(d)}$ $(d = 0, \ldots, D-1)$, which total $D + K$ in number. Of practical importance is that $v(t)$ is equivalent to a spline function. As in general spline function theory, the t_k are referred to as *(spline) knots*. Since spline functions are linear in their parameters, $v(t)$ can be equivalently expressed as:

$$v(t) = \boldsymbol{a}_t^T \ \boldsymbol{\beta} \tag{7}$$

with $\boldsymbol{a}_t$ the $D+K$ spline basis functions evaluated at time t $(t_1 \leq t \leq t_I)$ and $\boldsymbol{\beta}$ the parameters, named *spline coefficients* (de Boor, 1978; Ramsay & Silverman, 2005). Similarly, each of the spline function's derivatives can be expressed as a linear function of the same spline coefficients, however using a set of modified basis functions, $\boldsymbol{a}_t^{(d)}$:

$$v^{(d)}(t) = (\boldsymbol{a}_t^{(d)})^T \ \boldsymbol{\beta},\ 1 \leq d \leq D \tag{8}$$

The complete model can be described as a sequential process with three steps. The first step produces $v(t)$ with the disturbance input $\delta(t)$. The second step takes $v(t)$ as a disturbance input and $u(t)$ as a control input to produce $\boldsymbol{s}(t)$. The last step produces the measurements $\tilde{y}_i$. This is depicted graphically in Fig. 1.

Data-generating model – Application. In this work, we use the developed method to estimate an OUR signal from dissolved oxygen (DO) concentration measurement time series in an intermittently fed continuously stirred tank reactor. The applied model is:

$$\begin{bmatrix} \dot{c}_{DO} \\ \dot{y}_{DO} \end{bmatrix} = \begin{bmatrix} \frac{u}{\tau_c} \ (c_{DO,sat} - c_{DO}) - r_{OUR}(t) \\ \frac{1}{\tau_y} \ (c_{DO} - y_{DO}) \end{bmatrix} = \boldsymbol{f}\left(\begin{bmatrix} c_{DO} \\ y_{DO} \end{bmatrix}, u, r_{OUR} \right) \tag{9}$$

$$\tilde{y}_{DO,k} = y_{DO}(t_i) + \epsilon_i \tag{10}$$

$$\epsilon_i \sim \mathcal{N}\left(0, \sigma_\epsilon\right). \tag{11}$$

c_{DO} and y_{DO} are the DO concentration (state) and the noise-free DO concentration measurement while u is a binary (0/1) control variable which determines

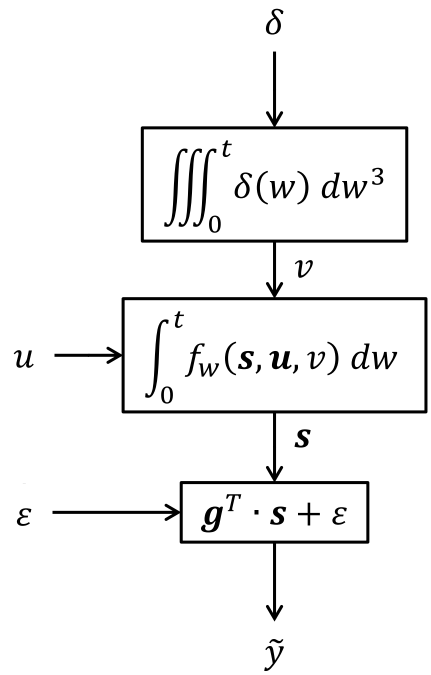

Figure 1: Schematic representation of the generative model. An discrete-time input disturbance signal δ is integrated to produce the piece-wise polynomial signal v ($D = 3$). This signal enters $\boldsymbol{f}_w(\boldsymbol{s}, \boldsymbol{u}, v)$ as an uncontrolled input, together with the controlled input (u). This produces the state vector $\boldsymbol{s}$ which is further processed into a noisy measurement ($\tilde{y}$).

whether the process is aerated $(u = 1)$ or not $(u = 0)$. τ_c is a time constant describing the dynamic effect of aeration while τ_y is a time constant describing the oxygen sensor response. We consider the OUR an unknown input disturbance ($v(t) = r_{OUR}(t)$) and apply the following additional definitions: $\boldsymbol{g} = \left[\begin{array}{cc} 0 & 1 \end{array}\right]^T$, $\boldsymbol{s} = \left[\begin{array}{cc} c_{DO} & y_{DO} \end{array}\right]^T$, $\boldsymbol{y} = y = y_{DO}$, $\boldsymbol{u} = u$, $\boldsymbol{\tau} = \left[\begin{array}{cc} \tau_c & \tau_y \end{array}\right]^T$. Accordingly, the model (9)-(11) can be written in the form of (1)-(3).

The $v(t)$ input signal is described by a cubic spline function ($D = 3$). In Fig. 2, an example of its constituting basis functions $(\boldsymbol{a}_t)$ are displayed for demonstration purposes with a knot added at every 256th measurement sample.

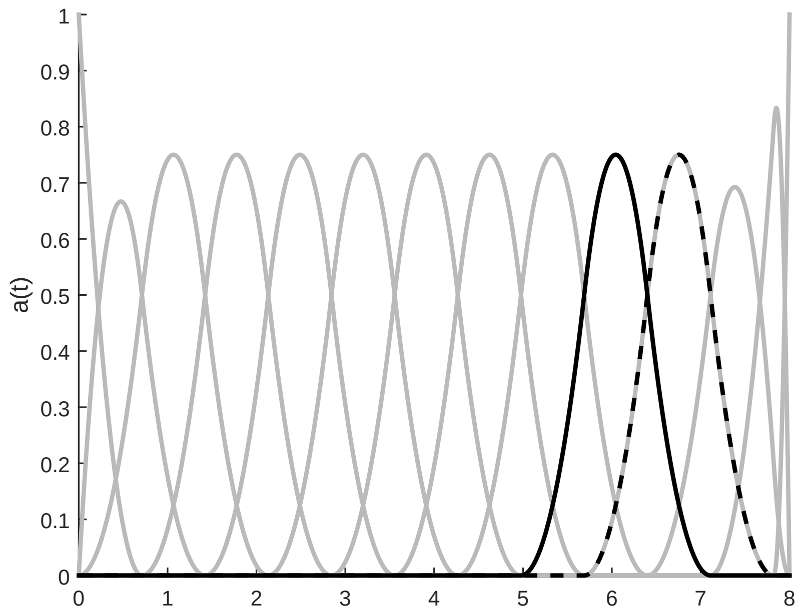

Figure 2: Cubic spline function basis – Knots are placed at every 256th sampling point (every 42'40") over a span of 8H. This means that there are 11 piece-wise polynomial intervals ($K = 11$) and 14 basis functions ($K + D = 14$). The basis functions are shown in grey, except the 10th (full black line) and the 11th (dashed black line). Apart from the first two first and the last two basis functions, every basis function is translated copy of the third basis function.

Integral form – Theory. In integral form, the above model becomes:

$$\tilde{y}_i = \tilde{y}(t_i) = \boldsymbol{g}^T \left(\boldsymbol{s}_0 + \int_0^t \boldsymbol{f}(\boldsymbol{s}(w), \boldsymbol{u}(w), v(w))\ dw \right) + \epsilon_i \tag{12}$$

$$\boldsymbol{s}_0 = \boldsymbol{s}(0) \tag{13}$$

Importantly, the piece-wise LTI nature of the process and the piece-wise polynomial nature of the input disturbance means that this integral can be rewritten as follows:

$$\tilde{y}_i = \tilde{y}(t_i) = {\boldsymbol{c}_{t_i}}^T \begin{bmatrix} \boldsymbol{s}_0 \\ \boldsymbol{\beta} \end{bmatrix} + \epsilon_i \tag{14}$$

with $\boldsymbol{c}_{t_i}$ a vector within which the first S columns describe the system's response at t_i to the initial conditions whereas the remaining $D+K$ columns describe the response of the measurement $\tilde{y}_i$ to $v(t)$ from 0 to t_i. The latter columns are obtained by convolution of each of the (piece-wise polynomial) spline basis functions with the piece-wise LTI response. This is executed analytically thanks to the fact that the response of an LTI system to a polynomial input can be described as a linear combination of unit (pulse/step/ramp/parabola/...) responses with the linear combination defined by the polynomial coefficients.

Integral form – Application. In Fig. 3, one can see the $D+K$ convoluted basis functions obtained with the spline function displayed in Fig. 2 as well as the S basis functions describing the effect of the initial conditions. In the top panel one can see the basis functions obtained without aeration ($\forall t : u(t) = 0$). The integrating nature of the process (see (12)) is particularly obvious in this figure as the basis functions in Fig. 2 are unimodal curves and the convoluted basis functions in Fig. 3 are monotonically increasing curves with a single inflection point. Moreover, the basis functions in Fig. 2 that are translated copies of each other results in convoluted basis functions in Fig. 3 that are also translated copies of each other. In the bottom panel, one can see the basis functions obtained when $u(t)$ switches multiple times between 0 and 1. In this case, the convoluted basis functions decay towards zero in periods where $u(t) = 1$. Indeed, due to aeration the effect of prior oxygen consumption disappears as time progresses. Naturally, this is only the case when the aeration is on. This is due to the fact that the $u(t)$ signal acts as a modulating signal. Because $u(t)$ is aperiodic, the original regular nature of the spline basis functions, including the translative properties discussed above, are lost.

2.2. Problem statement

In what follows next, all parameters except $\boldsymbol{s}_0$ and $\boldsymbol{\beta}$ are considered known. Thereafter the case where $\boldsymbol{\tau}$ is also unknown is considered. The control inputs $u(t)$ are assumed known perfectly and $v(t)$, $\boldsymbol{s}(t)$, and $\boldsymbol{y}(t)$ are unknown. Our

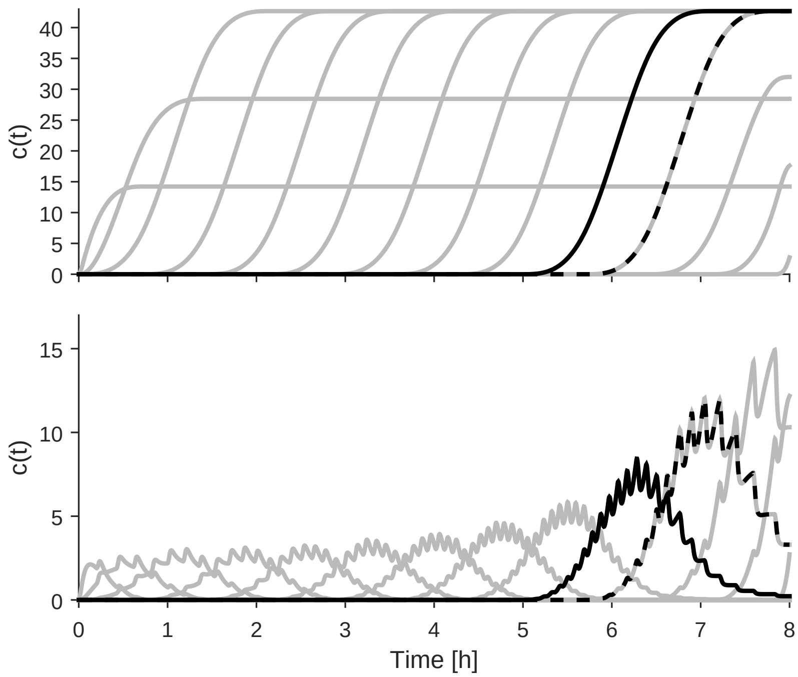

Figure 3: Convoluted cubic spline function basis – (top) Convoluted cubic spline functions in the case $u(t)$=0. The new basis functions are obtained by simple integration of the original cubic spline functions. This leads to a quartic M-spline basis. Spline basis functions that are translated versions of each other remain so after convolution. (bottom) Convoluted cubic spline functions in the switching binary input case. In this case, the convolution leads to a more complex pattern showing the effects of the on-off controller. Due to an irregular pattern of the $u(t)$ signal, the convolution does not preserve translation property anymore.

primary interest lies with the estimation of $v(t)$ by finding the best-fitting spline coefficients $\boldsymbol{\beta}$.

2.3. Method 1: Conventional input estimation

Conventionally, input estimation relies on an additional assumption regarding the input disturbances. It is typical to assume that the values of δ_k are sampled from a zero-mean normal distribution (white noise) with a given standard deviation σ_δ:

$$\delta_k \sim \mathcal{N}(0, \sigma_\delta) \tag{15}$$

Under such circumstances, one can compute maximum-likelihood estimates of the coefficients of $\boldsymbol{\beta}$ by solving the following optimization problem:

$$\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{s}}_0 = \arg\min_{\boldsymbol{\beta}, \boldsymbol{s}_0} \sum_i^I \left(\frac{y_i - \tilde{y}_i}{\sigma_\epsilon} \right)^2 + \sum_k^K \left(\frac{\delta_k}{\sigma_\delta} \right)^2 \tag{16}$$

subject to

$$y_i = \boldsymbol{c}_{t_i}{}^T \begin{bmatrix} \boldsymbol{s}_0 \\ \boldsymbol{\beta} \end{bmatrix} \tag{17}$$

$$\delta_k = \left(\boldsymbol{a}_{t_k}^{(D)} \right)^T \boldsymbol{\beta} \tag{18}$$

This optimization problem is a quadratic program with linear inequality constraints defining a non-empty set for $\boldsymbol{\beta}$. Therefore, the unique globally optimal solution can be computed analytically. In practice, the standard deviations σ_ϵ and σ_δ may not be known. In such cases, it is typical to use the ratio of variances $\lambda = \frac{\sigma_\epsilon^2}{\sigma_\delta^2}$ as a tuning parameter during estimation. For a given λ, the obtained solution will be the same regardless of the exact values of σ_ϵ and σ_δ. Setting λ to a higher (lower) value means that the variance of the disturbance inputs (δ_k) is penalized more (less) than the variance of the model prediction errors ($y_i - \tilde{y}_i$), further leading to a smoother (rougher) estimate of the δ_k input series. When applied so, λ becomes a smoothing parameter which is fine-tuned to balance a good model fit to the measurements against smoothness of the estimated signals. This idea is commonly referred to as regularization and is well-known in regression (e.g., ridge regression, Marquardt, 1970) and model-based observer tuning (e.g., Åkesson et al., 2008).

2.4. Method 2: Input estimation with shape constraints

Shape constraints – Theory. In this work, we propose an alternative strategy which is based on the assumption that one knows the shape of the signal $v(t)$

but not its expected distribution. More specifically, we assume that the shape of $v(t)$ is defined by E contiguous episodes within which the signs of its derivatives do not change. The derivative signs of the derivatives are given as a matrix $\boldsymbol{S}$ with $\boldsymbol{S}(e, d+1)$ corresponding to the sign of the dth derivative in the eth episode. The $T = E - 1$ locations in time where one episode ends and the next episode starts are known as *transitions* and are given as $\boldsymbol{\theta}$. The desired shape $\boldsymbol{S}$ is assumed known. In contrast, the transitions $\boldsymbol{\theta}$ are not known a priori and are therefore added to the estimation problem. The solution for $\boldsymbol{\beta}$ is found by solving the following fitting problem:

$$\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{s}}_0\right) = \arg \min_{\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{s}_0} g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{s}_0) = \sum_{i=1}^{I} \left|\tilde{y}_i - y_i\right|^2 \tag{19}$$

subject to

$$y_i = \boldsymbol{c}_{t_i}{}^T \begin{bmatrix} \boldsymbol{s}_0 \\ \boldsymbol{\beta} \end{bmatrix} \tag{20}$$

$$v(t) = \boldsymbol{a}_t^T \ \boldsymbol{\beta} \tag{21}$$

$$v^{(d)}(t) = \frac{\partial^d}{\partial t^d} v(t) = \boldsymbol{a}_t^{(d)^T} \ \boldsymbol{\beta} \tag{22}$$

$$v^{(d)}(t) \begin{cases} \leq 0, \text{if } t \in \left[\underline{b_e}, \overline{b_e}\right] \wedge \boldsymbol{S}(e, d+1) = +1 \\ = 0, \text{if } t \in \left[\underline{b_e}, \overline{b_e}\right] \wedge \boldsymbol{S}(e, d+1) = 0 \\ \geq 0, \text{if } t \in \left[\underline{b_e}, \overline{b_e}\right] \wedge \boldsymbol{S}(e, d+1) = -1 \end{cases} \tag{23}$$

$$\begin{aligned} \underline{\boldsymbol{b}} &= \begin{bmatrix} \underline{b_1} & \underline{b_2} & \cdots & \underline{b_e} & \cdots & \underline{b_E} \end{bmatrix} \\ &= \begin{bmatrix} t_1 & \theta_1 & \cdots & \theta_{t-1} & \cdots & \theta_T \end{bmatrix} \end{aligned} \tag{24}$$

$$\begin{aligned} \overline{\boldsymbol{b}} &= \begin{bmatrix} \overline{b_1} & \overline{b_2} & \cdots & \overline{b_e} & \cdots & \overline{b_E} \end{bmatrix} \\ &= \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_t & \cdots & t_I \end{bmatrix} \end{aligned} \tag{25}$$

$$\boldsymbol{\theta} \in \Theta \tag{26}$$

The objective function (19) is quadratic in $\boldsymbol{\beta}$. (20) and (21) are linear constraints and the remaining constraints, (22)-(26), determine the shape constraints. Because $v(t)$ is described by a spline function, these shape constraints can be reformulated as a finite number of equality and inequality constraints which together describe the feasible space for $\boldsymbol{\beta}$ as a semi-definite cone, Ω. For details on how to do this we refer to Nesterov (2000); Papp & Alizadeh (2014); Villez et al. (2013). Consequentially, a simpler formulation of the optimization problem is:

$$\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}\right) = \arg\min_{\boldsymbol{\beta},\boldsymbol{\theta}} g(\boldsymbol{\beta}, \boldsymbol{\theta}) \tag{27}$$

subject to

$$g(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{i=1}^{I} |\tilde{y}_i - y_i|^2 \tag{28}$$

$$y_i = \boldsymbol{c}_{t_i}{}^T \begin{bmatrix} \boldsymbol{s}_0 \\ \boldsymbol{\beta} \end{bmatrix} \tag{29}$$

$$v(t) = \boldsymbol{a}_t^T \; \boldsymbol{\beta} \tag{30}$$

$$v^{(d)}(t) = \frac{\partial^d}{\partial t^d} v(t) = \boldsymbol{a}_t^{(d)^T} \; \boldsymbol{\beta} \tag{31}$$

$$\boldsymbol{\beta} \in \Omega\left(\boldsymbol{S}, \boldsymbol{\theta}\right) \tag{32}$$

$$\boldsymbol{\theta} \in \Theta \tag{33}$$

where Θ is the feasible set for $\boldsymbol{\theta}$ and where $\Omega\left(\boldsymbol{S}, \boldsymbol{\theta}\right)$ is the convex feasible set for $\boldsymbol{\beta}$, given the desired shape ($\boldsymbol{S}$) and the transitions ($\boldsymbol{\theta}$).

The above optimization problem is a (convex) semi-definite program given values for $\boldsymbol{\theta}$. In special cases, the optimization problem is a second-order cone program, or even a quadratic program (Nesterov, 2000; Villez et al., 2013; Papp & Alizadeh, 2014). The problem is however nonlinear and possibly multi-modal in $\boldsymbol{\theta}$. Fortunately however, the bounding procedures developed in Villez et al. (2013) apply just as well to this newly posed problem, meaning that globally optimal values for $\boldsymbol{\theta}$ can be found in a finite number of steps via branch-and-bound optimization (Floudas, 1999; Floudas & Gounaris, 2009; Forst & Hoffmann, 2010). The branch-and-bound algorithm is halted when all live nodes of the search tree are equal to or completely within a single sampling interval. At that time, the best available upper bound solution is selected as the optimal solution. In the branching step, the node with the lowest lower bound is always selected for further branching. Further details regarding the implementation of this algorithm, including the applied bounding procedures, can be found in the *Appendix*.

Importantly, the above model does not require knowledge of the input and output disturbance standard deviations (σ_ϵ, σ_δ), let alone the smoothing parameter λ. Instead, the applied shape constraints, defined by $\boldsymbol{S}$, are used to deliver a smoothed estimate of $v(t)$. The shape constraints can be interpreted as a prior for the spline coefficients. When so, the optimization procedure delivers the corresponding maximum a posteriori estimates. This is similar in philosophy to the model identification method proposed in Vertis et al. (2016) and the data reconciliation approach proposed in Srinivasan et al. (2017). A notable difference however is that the input estimation only requires knowing the sequence of trends or shapes ($\boldsymbol{S}$) while the transitions ($\boldsymbol{\theta}$) are estimated. In contrast, Srinivasan et al. (2017) and Vertis et al. (2016) assume that *(i)* both the sequence of trends and transitions are known priori or *(ii)* that they can be determined by visual inspection.

Shape constraints – Application. In the studied case, the OUR signal is described well as a concave episode followed by a convex episode (E =2). In addition, the OUR decreases in the second episode. Thus, one writes:

$$\boldsymbol{S} = \begin{bmatrix} ? & ? & -1 & ? \\ ? & -1 & +1 & ? \end{bmatrix} \tag{34}$$

with $\boldsymbol{\theta} = \theta$ describing an inflection point at θ. The symbol ? is used to indicate unspecified signs. This inflection point corresponds to the change from exogenous to endogenous respiration conditions in the studied process.

2.5. Method 3: Joint input and parameter estimation with shape constraints

We now consider the case where the values for $\boldsymbol{\tau}$ are unknown. When so, the following optimization problem needs to be solved to obtain the best-fit $\boldsymbol{\beta}$:

$$\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\tau}}, \hat{\boldsymbol{s}}_0\right) = \arg \min_{\boldsymbol{\beta},\boldsymbol{\theta},\boldsymbol{\tau},\boldsymbol{s}_0} g(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{s}_0) \tag{35}$$

subject to

$$g(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{i=1}^{I} |\tilde{y}_i - y_i|^2 \tag{36}$$

$$y_i = \boldsymbol{c}_{t_i}{}^T \begin{bmatrix} \boldsymbol{s}_0 \\ \boldsymbol{\beta} \end{bmatrix} \tag{37}$$

$$v(t) = \boldsymbol{a}_t^T \ \boldsymbol{\beta} \tag{38}$$

$$v^{(d)}(t) = \frac{\partial^d}{\partial t^d} v(t) = \boldsymbol{a}_t^{(d)^T} \ \boldsymbol{\beta} \tag{39}$$

$$\boldsymbol{\beta} \in \Omega\left(\boldsymbol{S}, \boldsymbol{\theta}\right) \tag{40}$$

$$\boldsymbol{\theta} \in \Theta \tag{41}$$

All the above statements regarding (27)-(33) remain valid for this expanded optimization problem. Furthermore, the problem is now nonlinear and possibly multi-modal in $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$. In this work, it is solved by combining the DIRECT method with the branch-and-bound optimization (Jones et al., 1993; Finkel & Kelley, 2004, 2006). More specifically, the DIRECT method iteratively proposes values $\boldsymbol{\tau}$. For each proposed vector $\boldsymbol{\tau}$, the branch-and-bound algorithm is executed as discussed above to find the corresponding values for $\hat{\boldsymbol{\theta}}$. The DIRECT algorithm is a heuristic yet deterministic approach to global optimization. The DIRECT method does not provide guaranteed global optimality, unlike the branch-and-bound algorithm. For each candidate vector $\boldsymbol{\tau}$, the problem is solved to find the globally optimal values for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ conditional to $\boldsymbol{\tau}$. In this case, global optimality of $\hat{\boldsymbol{\theta}}$ is guaranteed conditional to the obtained values for $\hat{\boldsymbol{\tau}}$, which are however not guaranteed to be globally optimal themselves.

2.6. Experimental data

The analyzed data is collected in a laboratory-scale (12 L) intermittently-fed stirred tank reactor for urine nitrification. The reactor is operated in a cyclic manner with each cycle consisting of a short feeding stage and a no-feed stage. The feed consists of source-separate urine collected at Eawag with No-Mix toilets (Larsen et al., 2001). The oxygen level is controlled continuously by means of a bang-bang controller (on-off control, Levine, 1996)) with 5.0 and 5.5 mg $O_2 \cdot L^{-1}$as lower and upper control limits. Full nitrification is achieved by maintaining the pH above 6.5 with automated base addition (5 M NaOH). The oxygen and pH controllers are implemented by means of a WAGO PLC with a sampling rate below 1s. Every 10s, the accumulated valve opening time, the valve state, and the oxygen concentration measurement are registered together with a corresponding time stamp. The total length of the batch was 17h. The complete data vector consists of $I = 2881$ samples and covers a period of 8 hours, starting at 5h45' and ending at 13h45'. This period includes the point in time when all of the available ammonia and nitrite nitrogen is oxidized. The main reason to include only a segment of the available data is the large computational demands associated with the SCS model fitting. This challenge and possible ways to attack it are discussed below.

3. Results

3.1. Experimental data

The top panel of Fig. 4 displays the valve state and the measured oxygen as a function of time. One can see that the cycle starts with a fairly long aerated phase as the oxygen concentration slowly approaches the upper control limit. This is followed by a sequence of unaerated and aerated phases within which the oxygen concentration decreases and increases. The decreasing and increasing trends do not immediately follow the switching between aerated and unaerated phases. Some overshooting and undershooting is clearly visible. The overshooting (undershooting) tends to increase (decrease) as time progresses. This is explained as a consequence of a decreasing oxygen uptake rate. The bottom panel shows the ratio of the time length of each unaerated phase to the time between the start time of the considered unaerated phase and the next unaerated phase as a function of time. As time progresses, this ratio increases from close to zero (mostly aerated time) to close to one (mostly unaerated time). At 4h45' after the first considered measurement sample, the unaerated phase length is approximately the same as the aerated phase length. The observed profile of this ratio is thus also indicative of a decreasing oxygen uptake rate.

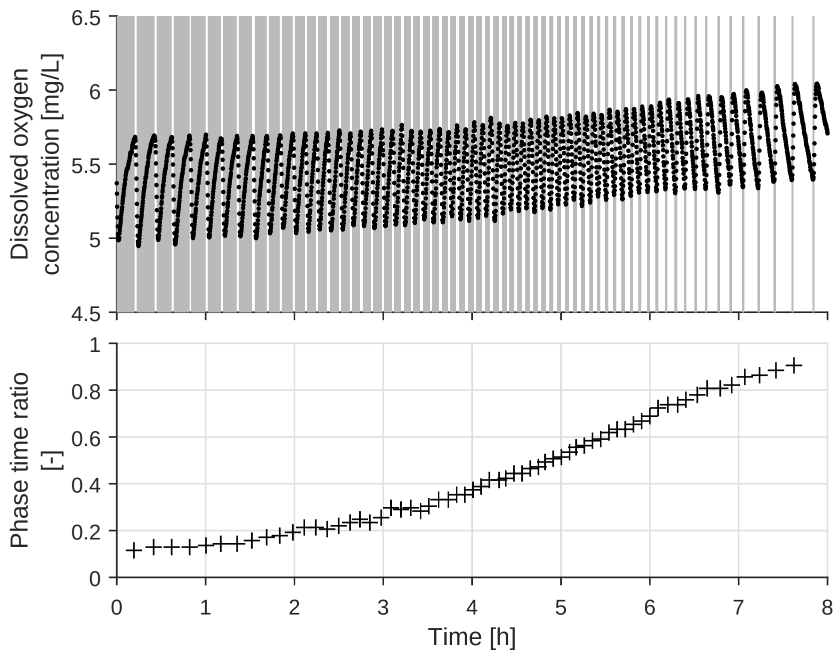

Figure 4: Experimental data – (top) Dissolved oxygen concentration measurements as a function of time. Aerated phases are indicated with grey shading. (bottom) Ratio of the time length of the unaerated phases to the time length between start times of unaerated phases as a function of time.

3.2. Results with method 1

An estimate of the input signal, $v(t)$, is first obtained by solving the problem described in (16)-(18). To this end, the model described by (9)-(11) is completely defined by $\boldsymbol{\tau} = \begin{bmatrix} \tau_c & \tau_y \end{bmatrix} = \begin{bmatrix} 3 & 1 \end{bmatrix} min^{-1}$. Note that these values for $\boldsymbol{\tau}$ are deliberately chosen to be close to the optimal values obtained with method 3 (see below).

The spline function $v(t)$ is defined with $D = 3$ and $k = \frac{i+1}{2} - 1$ ($k = 1, \ldots, K; K = \lceil \frac{I+1}{2} \rceil - 1$)). This means the spline knots are placed at every second measurement sampling time t_i. Consequentially, $v(t)$ is described as a spline function of degree $D = 3$ with K =1441 polynomial segments. $\boldsymbol{\beta}$ therefore contains $D + K = 1444$ spline coefficients. The smoothing parameter $\lambda = \frac{\sigma_\epsilon^2}{\sigma_\delta^2}$ was tuned to deliver the same root mean squared residual (RMSR) for the model prediction errors $(y_i - \tilde{y}_i)$ as obtained with method 2 (see below). This way, the amount of regularization achieved by tuning λ is similar to the regularization obtained with the application of shape constraints. In the top panel of Fig. 5 one can see that the obtained $y(t)$ profile matches the saw-tooth pattern of the DO measurements well. The OUR, shown in the middle panel of Fig. 5, can be described as a saw-tooth pattern as well, with values above and below zero. This pattern does not correspond to what is generally expected from an OUR signal. For instance, the OUR should be positive at all times and is usually a decreasing function of time. In addition, the bottom panel shows that the residuals between DO measurements and DO predictions are clearly auto-correlated. The model generally under-predicts the DO concentration at times where the aeration is switched off and over-predicts the DO concentration at times where the aeratio is switched on again.

3.3. Results with method 2

The model used with method 1 is now used again with method 2. The optimization problem (27)-(33) is solved with the shape constraints discussed for the example discussed above (34). In Fig. 6 the progress of the branch-and-bound algorithm is shown by visualizing the retained sets for θ at every iteration. After 14 iterations, the algorithm converged to within 1 measurement sampling interval and is halted. The corresponding best fit of the obtained model $(y(t))$ is shown in the top panel of Fig. 7 and corresponds to an inflection point at $\theta = 3h57'56$". One can see that the obtained $y(t)$ profile follows the saw-tooth pattern of measurements closely, as was the case with method 1. The OUR $(v(t))$ signal shown in the middle panel appears very different however. Importantly, one can see that the OUR curves match the desired shape, namely decreasing over the whole domain, concave in the 1st episode, and convex in the

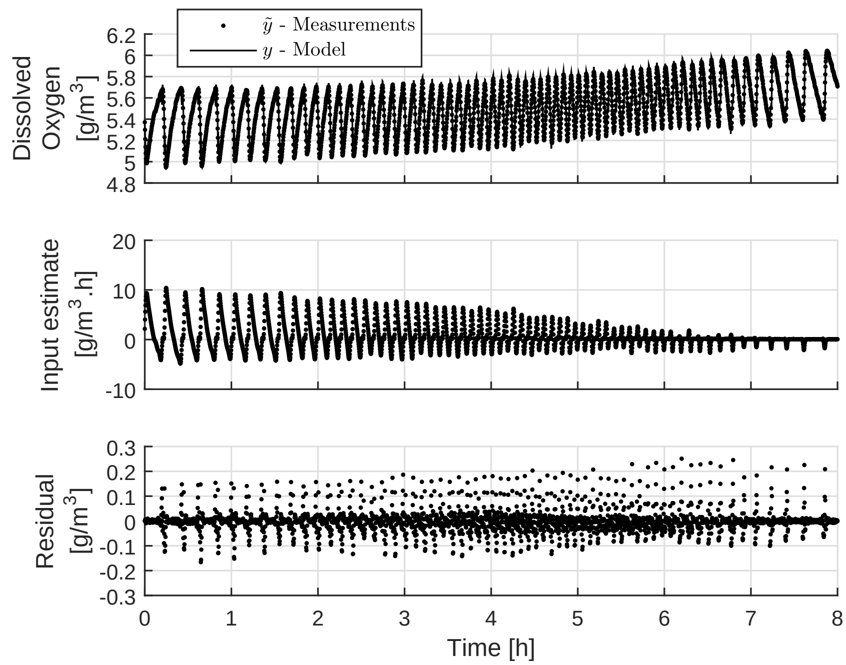

Figure 5: Regularized input estimation (method 1) – (top) Dissolved oxygen concentration measurements ($\tilde{y}(t)$) and fitted model predictions ($y(t)$) as a function of time. (middle) Input signal estimate ($v(t)$). (bottom) Predictions residuals as a function of time.

2nd episode. The bottom panel of Fig. 7 shows the residuals. These lie between -0.2 and +0.2 mg $O_2 \cdot L^{-1}$. The overall RMSR is 0.0563 mg $O_2 \cdot L^{-1}$. This is considered to reflect a reasonable fit to the data. A further improvement of the fit is however likely if not only θ but also the values of τ are optimized, as is discussed next.

3.4. Simultaneous input and parameter estimation

The branch-and-bound method is applied to find the optimal values of θ conditional to given values for $\boldsymbol{\tau}$. This optimization is nested in a DIRECT routine which proposes values for $\boldsymbol{\tau}$. The obtained fit of the model is slightly better than the one obtained before since the RMSR is now 0.0550 mg $O_2 \cdot L^{-1}$. Fig. 8 displays the estimates for $v(t)$ and $u(t)$ as well as the residuals, akin to

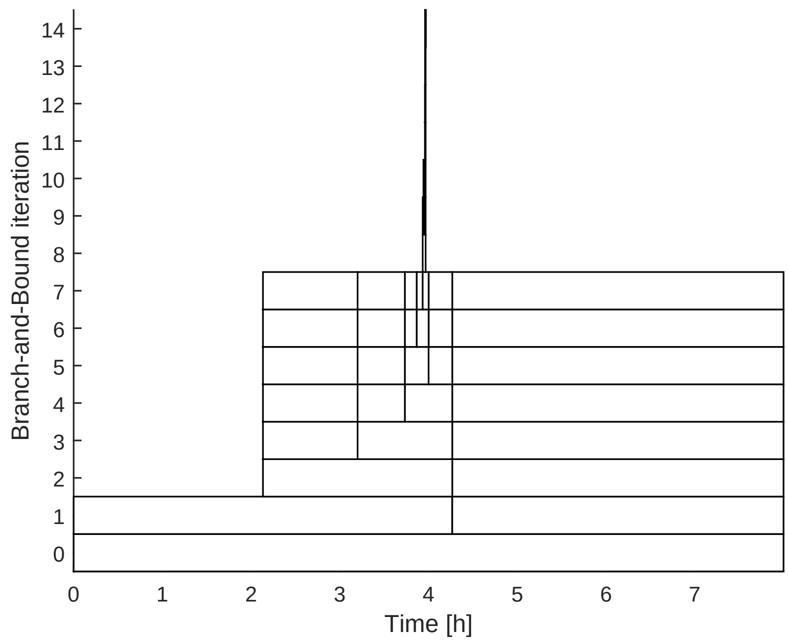

Figure 6: Progress of the branch-and-bound algorithm as a function of the branch-and-bound iteration number. At every iteration (bottom to top), the retained sets in the solution tree are shown. After 14 iterations, the optimal solution is found at 3h58'.

Fig. 7 yet now with optimal values for $\boldsymbol{\tau}$, which are $\boldsymbol{\tau} = \left[\begin{array}{cc} 2.99 & 0.893 \end{array} \right] min^{-1}$ at convergence. Interestingly, the values suggest a rather high time constant for the aeration process, equivalent to a $k_L a$ of about 20 h^{-1}. The estimated time constant for the oxygen sensor ($0.893 min^{-1}$) is reasonably fast but not negligible. The corresponding transition is found at $\theta = 3h58'30''$. This is very close to the value obtained previously with values for $\boldsymbol{\tau}$ that deviate from their optimum. This suggests that the estimate of θ is rather insensitive to the values for $\boldsymbol{\tau}$.

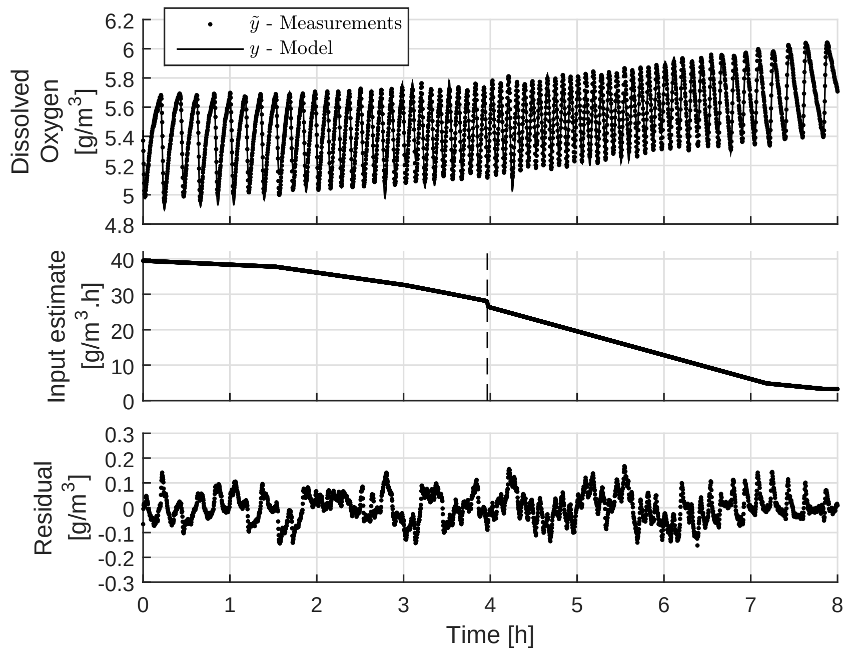

Figure 7: Shape constrained input estimation (method 2) – (top) Dissolved oxygen concentration measurements ($\tilde{y}(t)$) and fitted model predictions ($y(t)$) as a function of time. (middle) Input signal estimate ($v(t)$). (bottom) Predictions residuals as a function of time.

4. Discussion

4.1. Main achievements

In this study, an SCS-based method is used for input estimation and simultaneous input and parameter estimation and compared to a more conventional approach based on regularized fitting. It is shown that the SCS-based method leads to a nonlinear optimization problem which can however be solved to global optimality in a deterministic manner. The obtained estimation procedure corresponds to maximum a posteriori estimation if *(i)* the measurement noise is Gaussian and independently and identically distributed and *(ii)* the shape constraints are interpreted as defining a prior likelihood for the input signal. Most importantly, it was shown that the conventional approach leads to an OUR sig-

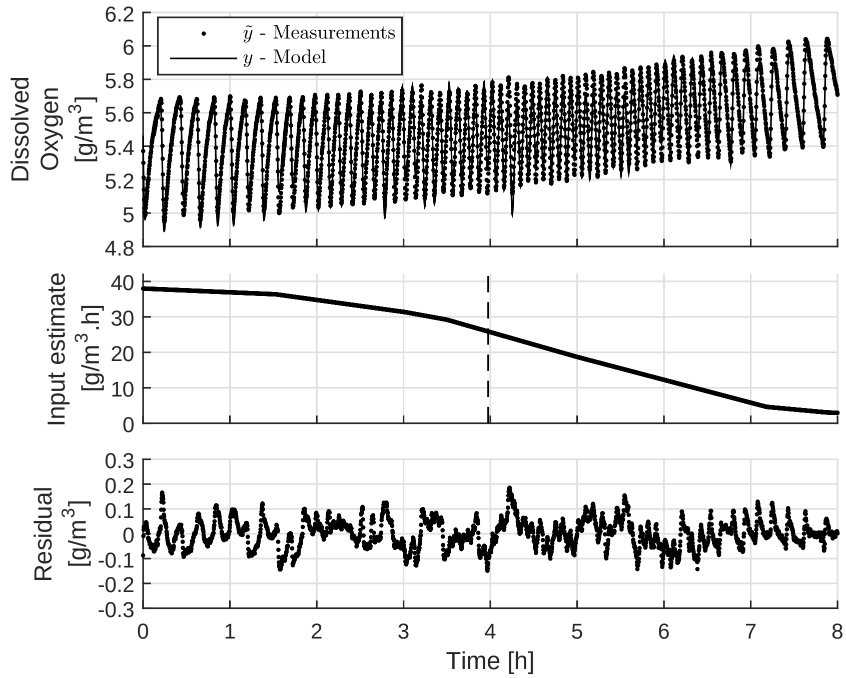

Figure 8: Simultaneous input and parameter estimation (method 3) – (top) Dissolved oxygen concentration measurements ($\tilde{y}(t)$) and fitted model predictions ($y(t)$) as a function of time. (middle) Input signal estimate ($v(t)$). (bottom) Predictions residuals as a function of time.

nal estimate that is hard to interpret, let alone trust. This is believed to be due the inability of the smoothing approach to account for model-reality mismatch. In contrast, the method based on shape constraints does not suffer from a lack of transparency and thereby leads to a sensible estimates of the process input disturbances, mainly by incorporating prior knowledge via the imposed shape constraints.

In addition, the method based on shape constraints has been extended further to enable simultaneous input and parameter estimation. This is shown possible through combination of the branch-and-bound algorithm and the DIRECT algorithm. Both the input estimation method and the simultaneous input and parameter estimation method are demonstrated with data obtained in a laboratory-scale reactor for urine nitrification. Using such experimental data shows that the proposed method can be used in realistic experimental conditions.

4.2. Benefits of the proposed method

The original SCS method is restricted to the direct analysis of univariate signals. This means that the estimated signal shape is required to correspond to the shape of the analyzed signal. With this work, this requirement has been lifted. Indeed, the estimated shape of the input signal does not need to match the shape of the measured signal.

The chosen approach bears some similarity to the QTA of principal scores as studied in Maurya et al. (2005) given that principal component analysis is used to uncover latent or hidden signals in measured data. In contrast to this study, our method is based on a mechanistic model and enables QTA by analyzing the measured data in a single step. This bears some similarity to the work in Flehmig & Marquardt (2008), even if the latter study is focused on linear trend identification. This decoupling makes it possible to estimate a slowly changing input signal that is subject to a fast process. This is especially valuable if the fast-changing process is not of primary interest for process monitoring, diagnosis, or control. This is the case for many biological processes, where the interesting dynamics of the biological process (e.g., OUR) are buried in a fast-changing signal (e.g., DO) generated by feedback controllers that maintain macroscopic variables in a desired range.

In comparison to traditional estimation of the OUR, based on infrequent OUR estimation at the end of each non-aerated phase, the proposed method offers several advantages. These include:

- All available data is used for estimation, thus likely increasing the precision of the estimates and allowing the use of lack-of-fit statistics to check for anomalous process conditions (see Villez & Habermacher, 2016)

- The parameters describing the aeration system and sensor dynamics can be estimated simultaneously with the OUR, meaning that one can monitor both the aeration system and the sensor response time with a single model and estimation method.

- The OUR signal can be integrated analytically to obtain the accumulated oxygen consumption over time. This is particularly useful for wastewater characterization where this integral is conventionally obtained by first interpolating the infrequent OUR estimates linearly (Amerlinck, 2015). Such an approximation can now be avoided.

4.3. Future work

Further study may help to improve the following aspects of the method:

- The DIRECT method used for optimization of the process parameters does not guarantee global optimality. Methods to obtain globally optimal estimates may be required if the modeled process structure and/or the estimated signal result in an objective function that has multiple local minima.

- The applied model structure was assumed to be LTI. Alternative model structures can however be proposed to further improve the obtained fit of the model. For general-purpose monitoring, the method appears satisfactory however.

- The SCS method has recently been extended for multivariate signal analysis (Derlon et al., 2017). This approach can easily be combined with the method proposed here and would enable shape-constrained estimation of a multivariate input signal. This only works if the shape of each of the fitted spline functions is the same, as is the case in Derlon et al. (2017). A more general method, permitting use of distinct shapes for each spline function, is being developed at the time of writing.

- In its current form, broad applicability is limited due to a large computational demand when the time series exceed 2000 data points and the inter-knot distance approaches the sampling interval, as in our case. This demand is partly explained by the need to compute and store large matrices (consisting of vectors $\boldsymbol{c}_{t_i}(\boldsymbol{\tau})$) during optimization. The form of the optimization problem (27)-(33) is however very similar to those solved in moving horizon estimation (MHE) methods. This signifies that an MHE approach may allow reducing the size of the optimization problems, however requiring the optimization routine to be repeated in a moving window approach.

5. Conclusions

A new method for unknown input disturbance signal estimation is presented. It is rooted in prior work on qualitative trend analysis and allows estimation of a process signal of a known shape based on a linear piece-wise time-invariant model of the process dynamics. The method is demonstrated with data obtained at laboratory-scale in a high-intensity process for resource

y from source-separated urine. The results indicate that sensible input estimation is possible while estimates of the parameters describing the dynamics of aeration system and the sensor are also produced. The method therefore appears promising as a way to maximize the information that can be extracted from typical dissolved oxygen concentration profiles in aerobic biological processes.

6. Acknowledgments

The authors thank Dan Finkel for his Matlab implementation of the DIRECT algorithm (Finkel, 2003) and the authors of Villez et al. (2013); Villez & Habermacher (2016); Derlon et al. (2017) for the SCS toolbox. This study was financed by the Swiss National Science Foundation (SNSF, Project No.: 157097).

7. Appendix

7.1. Branch-and-bound optimization

The branch-and-bound algorithm is a long-standing and broadly applicable method for deterministic global nonlinear optimization (Floudas, 1999; Floudas & Gounaris, 2009; Forst & Hoffmann, 2010). For details regarding the branch-and-bound optimization methods for SCS fitting, we refer to Villez et al. (2013); Villez & Habermacher (2016). In what follows, we only discuss the bounding procedures. This is the only element in the branch-and-bound optimization that has been changed compared to the method in Villez et al. (2013).

7.2. Bounds for input estimation

As in Villez et al. (2013); Villez & Habermacher (2016), values for $\boldsymbol{\beta}$ can be obtained by greedy or convex optimization given values for $\boldsymbol{\theta}$. Therefore, joint optimization of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ is possible by a nested strategy which obtains values for $\boldsymbol{\theta}$ by branch-and-bound optimization. Values for $\boldsymbol{\beta}$ are repeatedly obtained by optimization given $\boldsymbol{\theta}$. For estimation of $v(t)$ the same strategy is applied. The bounding procedures and their proofs are analogous to those in Villez et al. (2013). We therefore give the bounding procedures without proofs.

In what follows, we consider that during optimization the jth (hyper)rectangular set of considered values for $\boldsymbol{\theta}$, Θ_j, is described as:

$$\boldsymbol{\theta} \in \Theta_j \Leftrightarrow \underline{\boldsymbol{\theta}} \leq \boldsymbol{\theta} \leq \overline{\boldsymbol{\theta}} \tag{42}$$

with vector inequalities applied element-wise. Any feasible vector $\boldsymbol{\theta}$ within this set must satisfy the following monotonicity constraint:

$$\boldsymbol{\theta} \in \Theta_j \Leftrightarrow \forall t = 1, \ldots, T : \underline{\theta_t} \leq \theta_t \leq \overline{\theta_t} \tag{43}$$

Upper bound. An upper bound is easily found by solving (19)-(25) for $\boldsymbol{\beta}$ given any vector $\boldsymbol{\theta}$ satisfying (42)-(43). We refer to the corresponding parameter values as $\boldsymbol{\theta}^U$ and $\boldsymbol{\beta}^U$ and the objective function value as $\overline{g}\left(\Theta_j\right) = g\left(\boldsymbol{\beta}^U, \boldsymbol{\theta}^U\right)$. If no value for $\boldsymbol{\theta}$ can be found that satisfies 42 and 43, then the set Θ_j is empty and the upper bound is set equal to $\overline{g}\left(\Theta_j\right) = +\infty$.

Lower bound. If Θ_j is empty, then the lower bound is equal to $+\infty$. If Θ_j is not empty, the following relaxed optimization problem is solved:

$$\hat{\boldsymbol{\beta}}^L = \arg\min_{\boldsymbol{\beta}} \underline{g}(\boldsymbol{\beta}, \Theta_j) \tag{44}$$

subject to

$$\underline{g}(\boldsymbol{\beta}, \Theta_j) = \sum_{i=1}^{I} |\tilde{y}_i - y_i|^2 \tag{45}$$

$$y_i = \boldsymbol{c}_{t_i}(\boldsymbol{\tau})^T \ \boldsymbol{\beta} \tag{46}$$

$$v(t) = \boldsymbol{a}_t^T \ \boldsymbol{\beta} \tag{47}$$

$$v^{(d)}(t) = \frac{\partial^d}{\partial t^d} v(t) = \boldsymbol{a}_t^{(d)^T} \ \boldsymbol{\beta} \tag{48}$$

$$\boldsymbol{\beta} \in \underline{\Omega}\left(\boldsymbol{S}, \Theta_j\right) \tag{49}$$

with $\underline{\Omega}\left(\boldsymbol{S}, \Theta_j\right)$ defined as:

$$\boldsymbol{\beta} \in \underline{\Omega}\left(\boldsymbol{S}, \Theta_j\right) \quad \Leftrightarrow \begin{cases} v^{(d)}(t) & \begin{cases} \leq 0, \text{if } t \in \left[\underline{b_e}, \overline{b_e}\right] \wedge \boldsymbol{S}(e, d+1) = +1 \\ = 0, \text{if } t \in \left[\underline{b_e}, \overline{b_e}\right] \wedge \boldsymbol{S}(e, d+1) = 0 \\ \geq 0, \text{if } t \in \left[\underline{b_e}, \overline{b_e}\right] \wedge \boldsymbol{S}(e, d+1) = -1 \end{cases} \\ \underline{\boldsymbol{b}} & = \begin{bmatrix} \underline{b_1} & \underline{b_2} & \cdots & \underline{b_e} & \cdots & \underline{b_E} \end{bmatrix} \\ & = \begin{bmatrix} t_1 & \theta_1 & \cdots & \theta_{t-1} & \cdots & \theta_T \end{bmatrix} \\ \overline{\boldsymbol{b}} & = \begin{bmatrix} \overline{b_1} & \overline{b_2} & \cdots & \overline{b_e} & \cdots & \overline{b_E} \end{bmatrix} \\ & = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_t & \cdots & t_I \end{bmatrix} \end{cases} \tag{50}$$

The value for $\underline{g}(\hat{\boldsymbol{\beta}}^L, \Theta_j)$ is a valid lower bound (without proof).

References

Amerlinck, Y. (2015). *Model refinements in view of wastewater treatment plant optimization: improving the balance in sub-model detail*. Ph.D. thesis Ghent University.

Bakshi, B. R., & Stephanopoulos, G. (1994). Representation of process trends – Part III. Multiscale extraction of trends from process data. *Computers & Chemical Engineering*, *18*, 267–302.

de Boor, C. (1978). *A Practical Guide to Splines*. Springer.

Bredeweg, B., Linnebank, F., Bouwer, A., & Liem, J. (2009). Garp3 – Workbench for qualitative modelling and simulation. *Ecological Informatics*, *4(5)*, 263–281.

Charbonnier, S., Garcia-Beltan, C., Cadet, C., & Gentil, S. (2005). Trends extraction and analysis for complex system monitoring and decision support. *Engineering Applications of Artificial Intelligence*, *18*, 21–36.

Charbonnier, S., & Gentil, S. (2007). A trend-based alarm system to improve patient monitoring in intensive care units. *Control Engineering Practice*, *15*, 1039–1050.

Choubert, J.-M., Rieger, L., Shaw, A., Copp, J., Spérandio, M., Sørensen, K., Rönner-Holm, S., Morgenroth, E., Melcer, H., & Gillot, S. (2013). Rethinking wastewater characterisation methods for activated sludge systems - A position paper. *Water Science and Technology*, *67*, 2363–2373.

Dash, S., Maurya, M. R., Venkatasubramanian, V., & Rengaswamy, R. (2004). A novel interval-halving framework for automated identification of process trends. *AIChE Journal*, *50*, 149–162.

Derlon, N., Thürlimann, C. M., Dürrenmatt, D. J., & Villez, K. (2017). Batch settling curve registration via image data modeling. *Water Research*, *114*, 327–337.

Ferrai, M., Guglielmi, G., & Andreottola, G. (2010). Modelling respirometric tests for the assessment of kinetic and stoichiometric parameters on MBBR biofilm for municipal wastewater treatment. *Environmental Modelling and Software*, *25*, 626–632.

Finkel, D. E. (2003). *DIRECT optimization algorithm user guide*. Technical Report Center for Research in Scientific Computation, North Carolina State University 2.

Finkel, D. E., & Kelley, C. T. (2004). Convergence analysis of the direct algorithm. *Optimization Online*, *1-10*.

Finkel, D. E., & Kelley, C. T. (2006). Additive scaling and the direct algorithm. *Journal of Global Optimization*, *36*, 597–608.

Flehmig, F., & Marquardt, W. (2006). Detection of multivariable trends in measured process quantities. *J. Process Control*, *16*, 947–957.

Flehmig, F., & Marquardt, W. (2008). Inference of multi-variable trends in unmeasured process quantities. *Journal of Process Control*, *18*, 491–503.

Floudas, C. (1999). *Deterministic Global Optimization: Theory, Algorithms and Applications*. Kluwer Academic, Dordrecht.

Floudas, C. A., & Gounaris, C. E. (2009). A review of recent advances in global optimization. *Journal of Global Optimization*, *45*, 3–38.

Forst, W., & Hoffmann, D. (2010). *Optimization - Theory and Practice*. Springer.

Fumasoli, A., Etter, B., Sterkele, B., Morgenroth, E., & Udert, K. M. (2016). Operating a pilot-scale nitrification/distillation plant for complete nutrient recovery from urine. *Water Science and Technology*, *73*, 215–222.

Gamero, F. I., Colomer, J., Meléndez, J., & Warren, P. (2006). Predicting aerodynamic instabilities in a blast furnace. *Engineering Applications of Artificial Intelligence*, *19*, 103–111.

Gamero, F. I., Meléndez, J., & Colomer, J. (2014). Process diagnosis based on qualitative trend similarities using a sequence matching algorithm. *Journal of Process Control*, *24*, 1412–1424.

Gernaey, A. K., Petersen, B., Ottoy, J. P., & Vanrolleghem, P. (2001). Activated sludge monitoring with combined respirometric-titrimetric measurements. *Water Research*, *35*, 1280–1294.

Jones, D. R., Perttunen, C. D., & Stuckman, B. E. (1993). Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Applications*, *79*, 157–181.

Kansou, K., & Bredeweg, B. (2014). Hypothesis assessment with qualitative reasoning: Modelling the Fontestorbes fountain. *Ecological Informatics*, *19*, 71–89.

Kravaris, C., Hahn, J., & Chu, Y. (2013). Advances and selected recent developments in state and parameter estimation. *Computers & Chemical Engineering*, *51*, 111–123.

Kuipers, B. (1989). Qualitative reasoning: Modeling and simulation with incomplete knowledge. *Automatica*, *25*, 571–585.

Larsen, T. A., Peters, I., Alder, A., Eggen, R., Maurer, M., & Muncke, J. (2001). Re-engineering the toilet for sustainable wastewater management. *Environmental Science and Technology*, *35*, 192A–197A.

Levine, W. S. (Ed.) (1996). *The control handbook*. Boca Raton, New York: CRC Press, IEEE Press.

Liu, J. (2007). On-line soft sensor for polyethylene process with multiple production grades. *Control Engineering Practice*, *15*, 769–778.

Marquardt, D. W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, *12*, 591–612.

Maurya, M. R., Paritosh, P. K., Rengaswamy, R., & Venkatasubramanian, V. (2010). A framework for on-line trend extraction and fault diagnosis. *Engineering Applications of Artificial Intelligence*, *23*, 950–960.

Maurya, M. R., Rengaswamy, R., & Venkatasubramanian, V. (2003). A systematic framework for the development and analysis of signed digraphs for chemical processes - 1. Algorithms and analysis. *Industrial & Engineering Chemistry Research*, *42(20)*, 4789–4810.

Maurya, M. R., Rengaswamy, R., & Venkatasubramanian, V. (2005). Fault diagnosis by qualitative trend analysis of the principal components. *Computers & Chemical Engineering*, *83*, 1122–1132.

Mašić, A., Srinivasan, S., Billeter, J., Bonvin, D., & Villez, K. (2017). Shape constrained splines as transparent black-box models for bioprocess modeling. *Computers & Chemical Engineering*, *99*, 96–105.

Nesterov, Y. (2000). High performance optimization, applied optimization. chapter Squared functional systems and optimization problems. (pp. 405–440). Kluwer Academic Publishers volume 33.

Papp, D., & Alizadeh, F. (2014). Shape-constrained estimation using nonnegative splines. *Journal of Computational and Graphical Statistics*, *23*, 211–231.

Petersen, B., Gernaey, K., & Vanrolleghem, P. A. (2001). Practical identifiability of model parameters by combined respirometric-titrimetric measurements. *Water Science and Technology*, *43*, 347–355.

Prakash, J., Huang, B., & Shah, S. L. (2014). Recursive constrained state estimation using modified extended kalman filter. *Computers & Chemical Engineering*, *65*, 9–17.

Åkesson, B. M., Jørgensen, J. B., Poulsen, N. K., & Jørgensen, S. B. (2008). A generalized autocovariance least-squares method for kalman filter tuning. *Journal of Process Control*, *18*, 769–779.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York, NY, USA.

Rengaswamy, R., & Venkatasubramanian, V. (1995). A syntactic pattern-recognition approach for process monitoring and fault diagnosis. *Engineering Applications of Artificial Intelligence*, *8*, 35–51.

Romanenko, A., & Castro, J. A. (2004). The unscented filter as an alternative to the EKF for nonlinear state estimation: a simulation case study. *Computers & Chemical Engineering*, *28*, 347–355.

Spanjers, H., & Vanrolleghem, P. . (1995). Respirometry as a tool for rapid characterization of wastewater and activated sludge. *Water Science and Technology*, *31*, 105–114.

Spanjers, H., Vanrolleghem, P., Olsson, G., & Dold, P. (1996). Respirometry in control of the activated sludge process. *Water Science and Technology*, *34*, 117–126.

Spérandio, M., & Etienne, P. (2000). Estimation of wastewater biodegradable COD fractions by combining respirometric experiments in various So/Xo ratios. *Water Research*, *34*, 1233–1246.

Srinivasan, S., Billeter, J., Narasimhan, S., & Bonvin, D. (2017). Data reconciliation for chemical reaction systems using vessel extents and shape constraints. *Computers & Chemical Engineering*, *101*, 44–58.

Stephanopoulos, G., Locher, G., Duff, M., Kamimura, R., & Stephanopoulos, G. (1997). Fermentation database mining by pattern recognition. *Biotechnology and Bioengineering*, *53*, 443–452.

Thürlimann, C. M., Dürrenmatt, D. J., & Villez, K. (2015). Evaluation of qualitative trend analysis as a tool for automation. In *Proceedings of the 12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering (PSE2015/ESCAPE25), Copenhagen, Denmark, May 31 - June 4, 2015, 2531-2536*.

Udert, K. M., & Wächter, M. (2012). Complete nutrient recovery from source-separated urine by nitrification and distillation. *Water research*, *46*, 453–464.

Vanrolleghem, P. A., & Spanjers, H. (1998). A hybrid respirometric method for more reliable assessment of activated sludge model parameter. *Water Science and Technology*, *37*, 237–246.

Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003). A review of process fault detection and diagnosis – Part II: Qualitative models and search strategies. *Computers & Chemical Engineering*, *27*, 313–326.

Vertis, C. S., Oliveira, N. M. C., & Bernardo, F. P. (2016). Constrained smoothing of experimental data in the identification of kinetic models. *Computer Aided Chemical Engineering*, *38*, 2121–2126.

Villez, K., & Habermacher, J. (2016). Shape anomaly detection for process monitoring of a sequencing batch reactor. *Computers & Chemical Engineering*, *91*, 365–379.

Villez, K., Pelletier, G., Rosén, C., Anctil, F., Duchesne, C., & Vanrolleghem, P. A. (2007). Comparison of two wavelet-based tools for data mining of urban water networks time series. *Water Science and Technology*, *56*, 57–64.

Villez, K., Rosén, C., Anctil, F., Duchesne, C., & Vanrolleghem, P. A. (2008). Qualitative representation of trends: an alternative approach to process diagnosis and control. *Water Science and Technology*, *57*, 1525–1532.

Villez, K., Rosén, C., Anctil, F., Duchesne, C., & Vanrolleghem, P. A. (2012). Qualitative representation of trends (QRT): Extended method for identification of consecutive inflection points. *Computers & Chemical Engineering*, *48*, 187–199.

Villez, K., Venkatasubramanian, V., & Rengaswamy, R. (2013). Generalized shape constrained spline fitting for qualitative analysis of trends. *Computers & Chemical Engineering*, *58*, 116–134.

Wang, D., Liu, J., & Srinivasan, R. (2010). Data-driven soft sensor approach for quality prediction in a refining process. *IEEE Transactions on Industrial Informatics*, *6*, 11–17.

Yoong, E. T., Lant, P. A., & Greenfield, P. F. (2000). In situ respirometry in an SBR treating wastewater with high phenol concentrations. *Water Research*, *34*, 239–245.