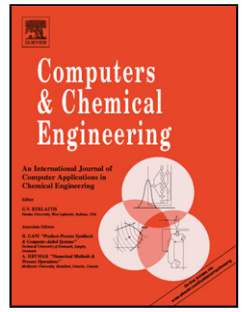
Accepted Manuscript



Title: Shape Anomaly Detection for Process Monitoring of a Sequencing Batch Reactor

Author: Kris Villez Jonathan Habermacher

PII: S0098-1354(16)30108-9
DOI: http://dx.doi.org/doi:10.1016/j.compchemeng.2016.04.012
Reference: CACE 5430

To appear in: *Computers and Chemical Engineering*

Received date: 8-9-2015
Revised date: 8-3-2016
Accepted date: 7-4-2016

Shape Anomaly Detection for Process Monitoring of a Sequencing Batch Reactor

Highlights

- Discontinuous shape constrained spline function are fitted to global optimality
- Shape constrained spline functions are used for fault detection for the first time
- The newly proposed fault detection method outperforms principal component analysis

Shape Anomaly Detection for Process Monitoring of a Sequencing Batch Reactor

Kris Villez[a,*], Jonathan Habermacher[a,b]

[a]Department of Process Engineering, Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland
[b]Institute of Environmental Engineering, ETH Zürich, CH-8093 Zürich, Switzerland

Abstract

Anomaly detection is critical to process modeling, monitoring, and control, since successful execution of these engineering tasks depends on access to validated data. Classical methods for data validation are quantitative in nature and require either accurate process knowledge, large representative data sets, or both. In contrast, a small section of the fault diagnosis literature has focused on qualitative data and model representations. The major benefit of such methods is that imprecise but reliable results can be obtained under previously unseen process conditions. This work continues with a line of work focused on qualitative trend analysis which is the qualitative approach to data series analysis. An existing method based on shape-constrained spline function fitting is expanded to deal explicitly with discontinuities and is applied here for the first time for anomaly detection. An experimental test case and a comparison with the principal component analysis method bear out the benefits of the qualitative approach to process monitoring.

Keywords: Anomaly detection, batch process monitoring, fault identification, principal component analysis, qualitative trend analysis, statistical process control

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

Preprint submitted to Comput. Chem. Eng. - SI:PSE2015/ESCAPE25 *March 8, 2016*

1. Introduction

Need for Anomaly Detection. The advent of increasingly intense data collection strategies for industrial processes suggests that increasing regulatory and efficiency requirements can be met by data-driven methods to model, monitor, and automate engineered process systems. However, data-driven computer-based technologies can only be successful if the data quality produced is guaranteed to be sufficient for automated decision-making and if the optimized process behaves in predictable ways. The data quality of biological processes can be severely deteriorated in many ways, including inadvertent human errors (*e.g.* calibration errors) and naturally occurring phenomena, ranging from events such as the passage of bubbles and particles over long-term processes such as film formation (*e.g.* biofilm growth, deposition, scaling) to sensor aging (*e.g.* corrosion). The processes themselves do not necessarily exhibit normal conditions either. Process faults commonly identified in biological processes include the toxicity effects of inlet streams and changes in microbial community composition or biochemical expression. Successful modeling, monitoring, and automation thus depend on effective tools for detecting anomalous data (Nopens et al., 2007; Thomann, 2008; Rieger et al., 2010; Dürrenmatt & Gujer, 2012; Spindler & Vanrolleghem, 2012).

Available Methods. A vast literature focuses on the automated detection, isolation, and identification of faults in actuators, processes, and sensors. These techniques are most commonly based on a (quantitative) model which describes data obtained under normal conditions of process operation. An important distinction can be made between techniques using models based on first principles (also known as mechanistic or white-box models, Venkatasubramanian et al., 2003c) and techniques using empirical data-based models (*i.e.* black-box models, Venkatasubramanian et al., 2003b). White-box models are particularly useful when the monitored process is understood to the point of allowing reliable mathematical models of it to be constructed. Black-box models are recommended for cases where the process understanding is limited and large representative data sets are available. Both supervised (*e.g.* classification) and unsupervised methods (*e.g.* clustering, principal component analysis) are popular. Unfortunately, both process understanding and historical data sets are often severely limited, especially for biological processes. Furthermore, extrapolation of the model in time can be challenging due to incipient changes and stochastic variations in the monitored process. Traditional methods for fault detection are seldom applicable without the need for substantial efforts to collect data and/or model the monitored process.

The Promise of Qualitative Methods. A smaller section of the literature presents qualitative methods as a valuable set of alternatives to the above-mentioned set of methods (Venkatasubramanian et al., 2003a). These methods are based on abstract, coarse-grained representations of data series and process dynamics. White-box models, such as qualitative differential equations (Kuipers, 1994) or signed directed graphs (Maurya et al., 2003), can again be identified. These are used to represent process dynamics qualitatively by focusing on the signs of process states and/or one or more of their rates of change, rather than their exact values. This deliberate lack of precision in the resulting model predictions leads to a high reliability of the resulting predictions even when extrapolated far from the conditions under which the model was identified. However, detailed process understanding is still a requirement since the qualitative models have so far been obtained only by abstracting from a quantitative dynamic model describing normal operating conditions, which is assumed to be available. Qualitative trend analysis (QTA) methods constitute the black-box equivalent (Maurya et al., 2007). In this case, data series of continuous variable measurements are represented by means of episodes, which characterize segments of the series in terms of the signs of one or more derivatives (Maurya et al., 2007). Such abstraction can facilitate the recognition of historical data patterns despite unpredictable variations in the exact data values. Most of the available methods are unsupervised in nature, *i.e.* without specification of the expected patterns. Due to the relatively recent emergence of this field, QTA methods are mostly based on an intuitive recombination of existing quantitative techniques (*e.g.* Dash et al., 2004; Villez, 2015).

Current Limitations of Qualitative Methods. (Villez et al., 2013) proposed a formal globally optimal deterministic optimization approach to QTA by recasting the pattern recognition problem as the maximum likelihood fitting of a shape-constrained splines (SCS) function. Solving this problem to a globally optimal level comes at large computational cost. For this reason, a faster and approximate method called qualitative path estimation (QPE) is developed by Villez (2015), offering similar performance at minimal computational cost. Both SCS and QPE methods are currently limited as *(i)* the qualitative patterns which ought to be recognized need to be specified before execution of the algorithm, *(ii)* the analysis is limited to univariate data series, and *(iii)* discontinuous trends cannot be accounted for in a systematic manner. To the authors' knowledge, it is impossible to modify the QPE method to remove this last limitation (see Villez, 2015). In this work, therefore, the existing SCS method is modified to support QTA in the presence of discontinuous trends.

This Study. In addition to the modifications of the SCS method, this article also describes for the first time how the SCS method provides a lack-of-fit statistic which can be used for fault detection. The analogy of this approach to the use of the Q or squared prediction error (SPE) statistic commonly used in fault detection based on principal component analysis (PCA, Jackson & Mudholkar, 1979; Kresta et al., 1991) is demonstrated below. Furthermore, this work compares the anomaly detection performances of both SCS and PCA. This article continues with *Materials and Methods*, in which the applied data models, anomaly detection methods, and the proposed performance evaluation are initially explained, followed by a description of the analyzed data and their purpose in this study. In the *Results* section, all the results obtained are discussed in detail while the *Discussion* section provides an in-depth analysis. This study is summarized in the last section, namely *Conclusions*.

2. Materials and Methods

The modified SCS method and PCA as applied here are initially described. This is followed by a description of the studied data sets. An overview of the acronyms and typographical conventions used as well as a list of symbols are given in Appendix A (Tables A.1, A.2, and A.3).

2.1. Methods

Two methods are used for anomaly detection. The first one is a modification of the existing SCS method, while the second one is based on PCA. Both methods result in the computation of a lack-of-fit statistic, namely a sum of squared residuals (SSR). In both cases, this statistic is used to detect anomalous data as explained at the end of this subsection.

2.1.1. Shape-Constrained Splines

Shape-constrained function fitting is applied here as a way of detecting significant deviations between the shape of a data series and a predefined shape reflecting normal conditions. The following paragraphs show how this problem can be formulated mathematically and solved numerically.

Definitions and Notation. In analogy to previous work, the following definitions are used here:

Episode. An episode is an argument interval over which the signs of a function or data series and/or a selection of their derivatives do not change. It is defined by a primitive, a start time, and an end time.

Primitive. A primitive is a unique combination of signs for a value of a function and/or one or more of its derivatives. Each primitive is usually referred to by means of an arbitrarily chosen character. The sign of the first and second derivatives of a cubic spline function are of interest in this work. The primitives are called triangular primitives when the signs for both the first and second derivative are specified (Cheung & Stephanopoulos, 1990). The correspondence between the signs of the derivatives and the characters is given in Fig. 1 and is the same as in Villez et al. (2013).

Qualitative Sequence. A qualitative sequence (QS) is a series of primitives. Such a QS is used to describe the assessed or expected shape of a function or data series. A QS does not include the argument locations (transitions) at which a change in primitive is expected or observed.

Qualitative Representation. A qualitative representation (QR) is a complete description of the expected or observed shape of a function or time series and consists of a QS and values for the argument values of the corresponding transitions.

Transition. A transition is defined as the argument location where one primitive changes to the next.

Any QS is defined mathematically by means of integers, $s_{e,j}$ ($s_{e,j} \in \{-1, 0, +1\}$), with e indicating the index of the primitive in the QS ($e \in \{1, 2, \ldots, n_e\}$) and j indicating the considered derivative ($j \geq 0$). An unknown or unspecified sign is symbolized with a question mark (?), similar to previous work (Villez et al., 2013). In all cases studied in this work, only triangular primitives are used so that the sign values of the cubic spline function and its third derivative are ? for all episodes. These signs are combined in matrix form as follows, with r being the highest derivative under consideration:

$$\boldsymbol{S} = \begin{bmatrix} s_{1,0} & s_{1,1} & \ldots & s_{1,j} & \ldots & s_{1,r} \\ s_{2,0} & s_{2,1} & \ldots & s_{2,j} & \ldots & s_{2,r} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{e,0} & s_{e,1} & \ldots & s_{e,j} & \ldots & s_{e,r} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{n_e,0} & s_{n_e,1} & \ldots & s_{n_e,j} & \ldots & s_{n_e,r} \end{bmatrix} \tag{1}$$

The transitions between primitives are given as a vector:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \ldots & \theta_{n_t} \end{bmatrix}^T \tag{2}$$

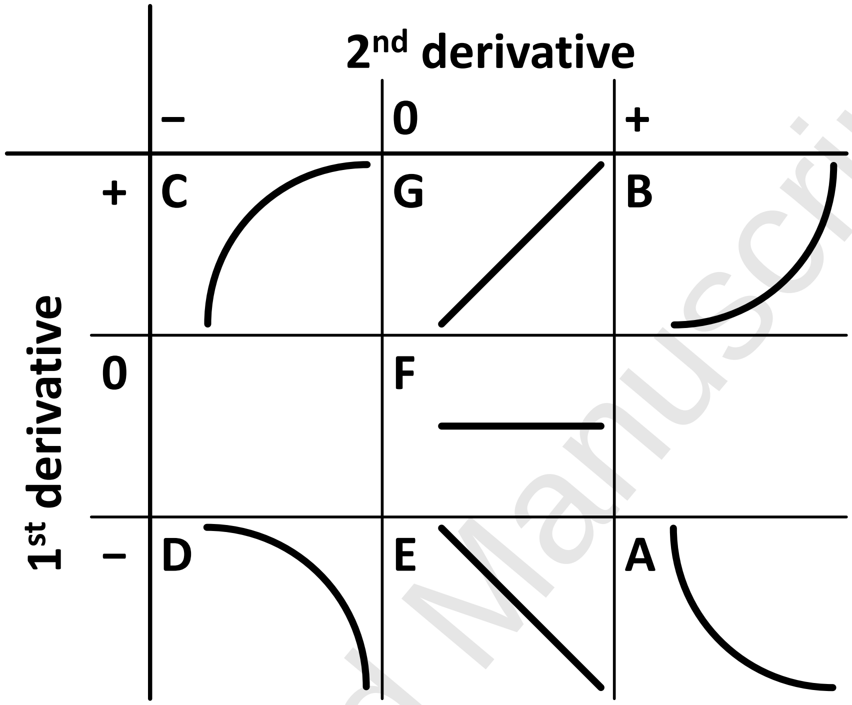


Figure 1: Primitives. The above scheme includes all triangular primitives defined on the basis of the sign of the 1^{st} and/or the 2^{nd} derivative.

with $n_t = n_e - 1$. A number, n_d, of transitions are known to imply a discontinuity in one or more derivatives which are otherwise continuous. These are defined as follows:

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_1 & \delta_2 & \dots & \delta_{n_d} \end{bmatrix}^T \tag{3}$$

and constitute a subset of θ:

$$\boldsymbol{\delta} \subseteq \boldsymbol{\theta} \tag{4}$$

The maximal degrees for the derivatives which are still continuous in $\boldsymbol{\delta}$ are given as

$$\boldsymbol{c}_{var} = \begin{bmatrix} c_{var,1} & c_{var,2} & \dots & c_{var,n_d} \end{bmatrix}^T \tag{5}$$

Problem Formulation. The fitting of shape-constrained spline functions is formulated mathematically as follows. Consider that a sequence of n data pairs, (x_i, y_i), is given as a vector of arguments ($\boldsymbol{x}$) and a matching vector of measurement values ($\boldsymbol{y}$):

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_i & \dots & x_n \end{bmatrix}^T \tag{6}$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_i & \dots & y_n \end{bmatrix}^T \tag{7}$$

The problem of fitting a shape-constrained univariate function can then be written in a general form as follows:

$$\min_{\boldsymbol{\beta},\boldsymbol{\theta}} g(\boldsymbol{\beta}) = g(\boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{y}) \tag{8}$$

subject to:

$$\boldsymbol{\beta} \in \Omega(\boldsymbol{S}, \boldsymbol{\theta}) \tag{9}$$

$$\boldsymbol{\theta} \in \Theta \tag{10}$$

where $\boldsymbol{\beta}$ is a function parameter vector; $\Omega(\boldsymbol{S}, \boldsymbol{\theta})$ is the set consisting of all vectors $\boldsymbol{\beta}$ for which the resulting function satisfies the qualitative representation defined by $\boldsymbol{S}$ and $\boldsymbol{\theta}$; and Θ is the feasible set of transitions. Further definitions are as shown in Table A.3. As discussed shortly above and as in prior SCS-based work, it is assumed that $\boldsymbol{S}$ is available a priori, either by means of expert reasoning or by computer-based qualitative reasoning (*e.g.* Kuipers, 2001; Bredeweg et al., 2009). The feasible set for each transition (θ_t) equals the function domain, subject to the isotonicity of their consecutive values:

$$\boldsymbol{\theta} \in \Theta \quad \Leftrightarrow \quad \forall t \in \{1, 2, \dots, n_t\} \ : \ x_1 \leq \theta_t \leq \theta_{t+1} \leq x_n \tag{11}$$

The shape constraints, described by the set $\Omega(\boldsymbol{S}, \boldsymbol{\theta})$, consist of equality and inequality constraints for the value of the fitted function and/or one or more of its derivatives. These constraints are expressed as functions of the argument

(x) and the function parameters ($\boldsymbol{\beta}$). As they are valid over intervals of the function domain (episodes), their number can be infinite. Fortunately, certain function families permit such an infinite number of constraints to be equivalently expressed as a finite number of constraint equations involving only the function parameters. This is possible for univariate spline functions, as demonstrated in Papp & Alizadeh (2014). The fitted functions are thus restricted to belong to the family of spline functions. Under these conditions, the above problem can be solved deterministically and globally by means of the branch-and-bound algorithm (Villez et al., 2013).

In existing work (Villez et al., 2013), the spline knots are considered known a priori. This means that discontinuous behavior of the otherwise continuous function and/or its derivatives implied by shape constraints is not permitted. Indeed, some examples of QSs (*e.g.* *FC*) imply a discontinuity of the function and/or one or more derivatives. Such discontinuous behavior can only be achieved by placing additional knots with multiplicity at the corresponding transition (Ramsay & Silverman, 2005). The function and its derivatives remain continuous in other parts of the function domain. Since the transitions are parameters to be optimized, this implies that a fraction of the spline knots is unknown a priori. It follows that the bounds proven in Villez et al. (2013) cannot be applied when the QS implies discontinuous behavior. This restriction is removed with the formulation and proof of new bounds in Appendix B. These new bounds require that the optimization problem can be written as follows:

$$\begin{aligned}\min_{\boldsymbol{\beta},\boldsymbol{\theta}} g(\boldsymbol{\beta}) &= g(\boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{y}) \\ &= \sum_i |y_i - f(\boldsymbol{\beta}, x_i)|^p + \sum_{j=0}^{j=r} \lambda_j \int_{x_1}^{x_n} \left| f^j(\boldsymbol{\beta}, v) \right|^{q_j} \, dv \qquad (12)\end{aligned}$$

with constraints as given above (Eqs. 9-10) and the definitions shown in Table A.3. The powers p and q_j are larger than or equal to one. The above objective function consists of separable penalty functions for the function fit and, possibly, the smoothness of the function and/or a number of its derivatives.

The global solution to the above shape-constrained spline fitting problem can be found to an arbitrary level of precision by a deterministic global search algorithm since the following apply (sufficient conditions):

1. The set $\Omega(\boldsymbol{S}, \boldsymbol{\theta})$ is convex for any given $\boldsymbol{S}$ and $\boldsymbol{\theta}$ (Papp & Alizadeh, 2014).
2. The objective function g is convex in the parameters $\boldsymbol{\beta}$ (Papp & Alizadeh, 2014).
3. Bounds to the objective function g exist and can be computed for any subset of Θ. This last condition is proven in Appendix B.

For details of the deterministic global optimization method for shape-constrained spline fitting, including necessary proofs, we refer to Appendix B. For the purpose of batch process monitoring, the above optimization problem is solved for every new data series $(\boldsymbol{x}, \boldsymbol{y})$ generated by the considered batch process. In this work, all penalty coefficients, λ_j, are equal to zero so that the objective function value after optimization corresponds to the minimal SSR given the spline function and the imposed shape constraints. This SSR is further referred to as SSR_{SCS}.

2.1.2. Principal Component Analysis

PCA is a well-known method for data dimension reduction and can be used for anomaly detection under specific assumptions, *e.g.* that all analyzed data samples are drawn independently from the same distribution. PCA for anomaly detection is executed in two phases. First, a PCA model is calibrated by analysis of a data matrix consisting of historic data samples which are considered normal. In a second step, the model obtained is used for confirmatory testing of newly obtained samples. As the PCA model and its use for fault detection are described at great length in the literature (see *e.g.* Jackson & Mudholkar, 1979; Joliffe, 2002), the following text focuses on the essentials.

Phase 1 - Calibration. The calibration data set is given as a matrix ($\boldsymbol{Y}_{cal}$) with n rows corresponding to variables and m columns corresponding to data samples. A centered data set ($\boldsymbol{Y}_{cal,C}$) is obtained by subtracting the mean vector from each matrix column. This matrix is decomposed into a matrix consisting of principal score vectors ($\boldsymbol{T}_{cal}$) and one consisting of loading vectors ($\boldsymbol{P}$). This is done here by means of singular value decomposition. This decomposition can be written as:

$$\boldsymbol{Y}_{cal,C} = \boldsymbol{P} \cdot \boldsymbol{T}_{cal} \tag{13}$$

It is well known that the loading vectors, *i.e.* the columns of $\boldsymbol{P}$, correspond to the eigenvectors of the empirical maximum likelihood covariance matrix estimate ($X \cdot X^T/m$) computed for the calibration data set. Similarly, the variances of the principal scores (rows of $\boldsymbol{T}_{cal}$) are equal to the eigenvalues of the same covariance matrix. By definition, the loading vectors and principal scores are ordered in decreasing order of eigenvalues. To achieve dimensional reduction, a number of principal components (PCs) with the smallest eigenvalues are removed from the model. This leads to the following equation where $\bar{\boldsymbol{T}}_{cal}$ and $\bar{\boldsymbol{P}}_{cal}$ comprise the retained part of the model and $\boldsymbol{R}$ represents the residuals:

$$\boldsymbol{Y}_{cal,C} = \bar{\boldsymbol{P}} \cdot \bar{\boldsymbol{T}}_{cal} + \boldsymbol{R} \tag{14}$$

It can be shown that the column vectors of $\bar{\boldsymbol{P}}$ describe the least-squares optimal plane approximating the centered data for a given number of PCs (Schuermans et al., 2005). A challenging task in PCA is the determination of the number of PCs (Joliffe, 2002). A simple scree plot of the eigenvalues is found to suffice in this study.

Phase 2 - Confirmatory Analysis. In a second phase, new samples are projected onto the PCA model. This is done by computing each centered sample ($\boldsymbol{y}_{test,C}$) by subtracting the means computed in phase 1. The principal scores are then computed as follows, thanks to the orthonormal properties of the loading vectors:

$$\bar{\boldsymbol{t}}_{test} = \boldsymbol{y}_{test,C} \cdot \bar{\boldsymbol{P}}^T \tag{15}$$

The quality with which the PCA model describes the new data samples is measured by the following sum of squared residuals (SSR_{PCA}), also known as the

Q statistic (Jackson & Mudholkar, 1979) and the SPE statistic (Kresta et al., 1991):

$$SSR_{PCA} = \left(\boldsymbol{y}_{test,C} - \bar{\boldsymbol{P}} \cdot \bar{\boldsymbol{t}}_{test}\right)^T \cdot \left(\boldsymbol{y}_{test,C} - \bar{\boldsymbol{P}} \cdot \bar{\boldsymbol{t}}_{test}\right) \tag{16}$$

The higher SSR_{PCA} is, the lower is the chance that the analyzed data are produced according to the PCA model. SSR_{PCA} can thus be used as a statistic for the automatic detection of anomalous data patterns. This is discussed in more detail below.

2.1.3. Anomaly Detection

Both the PCA model and the SCS model result in an SSR computed individually for each data sample. To use these SSR values for anomaly detection, an upper control limit (UCL) is specified to define the classification boundary for the anomaly detection problem (Montgomery, 2005). When the computed SSR is above (below) this UCL, the analyzed sample is considered anomalous (normal). In the case of PCA, an UCL can be computed on the basis of reliable approximations of the distribution of the SSR statistic and be given a proposed false positive rate (FPR, frequency of anomaly detections for normal data). This requires a multivariate normal distribution for the residuals to be assumed and identified (Jackson & Mudholkar, 1979; Kresta et al., 1991). In the case of SCS, no such approximations exist. To allow a fair comparison of these methods, the performance of the SSR statistics is instead evaluated by means of the receiver-operator-characteristic (ROC, Fawcett, 2006) which plots the true positive rate (TPR, frequency of anomaly detections for abnormal data) as a function of the FPR for different values of the UCL. These frequencies are computed and plotted by using every value obtained for SSR once as the UCL. This approach also avoids the arbitrary effects of an a priori specified FPR on the evaluation of the proposed methods.

2.2. Data Sets

Two data sets are used in this study. The first consists of a single univariate time series and is merely used to demonstrate the modified SCS method. The second consists of a larger set of time series obtained in a batch process for biological wastewater treatment. The latter set is used to demonstrate and compare the anomaly detection performance with of both the SCS and PCA methods. Both data sets are described below and are included in *Supplementary Materials.*

2.2.1. Data Set 1: Refinery Data

The refinery data set is taken from Ramsay & Silverman (2005) and contains 193 tray level measurements from an oil refinery distillation column. These measurements are recorded equidistantly in time and were used to showcase the introduction of the discontinuous behavior of knots with multiplicity and the identification of functional differential equations (Ramsay & Silverman, 2005). They are used here to demonstrate how a shape-constrained spline function can be fitted to global optimality when the enforced QS implies a discontinuity.

2.2.2. Data Set 2: Oxidation-Reduction Potential (ORP) Data

The second data set consists of 1684 univariate time series collected in a sequencing batch reactor (SBR) for aerobic wastewater treatment. This SBR consists of a reactor tank in a two-tank reactor setup which includes an experimental side-stream reactor (SStR) operated as a continuously stirred tank reactor. The SBR is used for aerobic treatment of sewage and the SStR for aerobic digestion of excess sludge. Each SBR cycle lasts six hours and is operated in the following fixed sequence of stages: *(i)* pumping of sludge from the SStR to the SBR and liquor from the SBR to the SStR (7 min.), *(ii)* addition of fresh wastewater under anoxic conditions (10 min.), *(iii)* aerated reaction phase (285 min.), and *(iv)* sludge withdrawal, settling, and decanting (58 min.). Each new cycle starts immediately after decanting. A complete description of the setup can be found in Habermacher et al. (2015).

The selected data consist of the first 513 oxidation-reduction potential (ORP) measurements collected every 10 seconds in each batch cycle. The time series thus represent the first 85 minutes of the batch cycle which includes the first two stages and the first hour of the aerobic stage. These data series exhibit a typical shape with discontinuities in the first and second derivatives at distinct locations. It is this behavior at the beginning of normal batch cycles that inspired the methodological development of the proposed shape-constrained spline fitting method.

To benchmark the anomaly detection methods, two experts were asked to classify each time series as explained by either normal or anomalous functioning of the process and the ORP sensor. The first expert is the second author of this paper and the second one is one of his research advisors. Their classification was recorded by means of a customized visualization which presented the time series to each expert separately and in a random order. Each inspected time series was shown in white against a black backdrop while all other time series for that operational period were shown in dark gray in the same image. The program used for this visualization and response recording can be found in the *Supplementary Materials*. As the two experts did not agree in all cases, a joint session was held in which the time series with conflicting classifications were shown simultaneously to both experts in order to obtain a consensus classification wherever feasible. After this joint session, 1564 cycles are classified as normal and 96 as abnormal. No consensus could be reached for 24 cycles. The data series and reference classification results can be found in the *Supplementary Materials*. Only the time series for which a consensus between the two experts was reached (1660 cycles) are used for comparative analysis.

2.3. Software

All computations were executed using Matlab (R2014b, The MathWorks Inc., 2014). The SCS method could be realized thanks to the use of convex optimization software (MOSEK ApS, 2012), a functional data analysis toolbox (Ramsay & Silverman, 2002), and an updated version of the SCS toolbox (Villez et al., 2013). In view of reproducibility, all data and programs necessary to repeat the data analysis and produce all figures in this work can be found in the *Supplementary Materials*. All software created newly for this work is published under an open-source license and released simultaneously with this publication.

3. Results

The results for the refinery time series are initially reported in the following. The results obtained with the ORP measurements are subsequently discussed.

3.1. Data Set 1: Refinery Data

The refinery time series is shown in Fig. 2. It can be seen that the QS *FC* is a reasonable abstraction for this series. This *FC* sequence implies discontinuous behaviour at the transition between the two constituting primitives. Indeed, a change from a zero-valued 1st and 2nd derivative to a strictly positive 1st derivative combined with a negative 2nd derivative can only be achieved by discontinuous behavior of the 1st and 2nd derivative.

To demonstrate the shape-constrained spline fitting method, a shape-constrained cubic spline function ($r = 3$) with fixed knots placed at every second data point is fitted to the data in the least squares sense. To this end, the powers and smoothness penalty coefficients are set as follows:

$$p = 2 \tag{17}$$

$$\forall j \in \{0, \ldots, r\} : \lambda_j = 0 \tag{18}$$

The function is constrained to exhibit the assumed *FC* shape. This can be expressed by the following sign matrix (cfr. Eq. 1 & Table 1):

$$\boldsymbol{S} = \begin{bmatrix} ? & 0 & 0 & ? \\ ? & +1 & -1 & ? \end{bmatrix} \tag{19}$$

An additional knot with multiplicity 3 is placed at the transition to provide the desired discontinuity of the 1st and 2nd derivative. This means that the function itself is the highest function derivative which remains continuous in this added knot, in mathematical terms:

$$\boldsymbol{\kappa}_{var} = \kappa_1 = \boldsymbol{\theta} = \theta = \boldsymbol{\delta} = \delta \tag{20}$$

$$\boldsymbol{c}_{var} = c_{var,1} = 0 \tag{21}$$

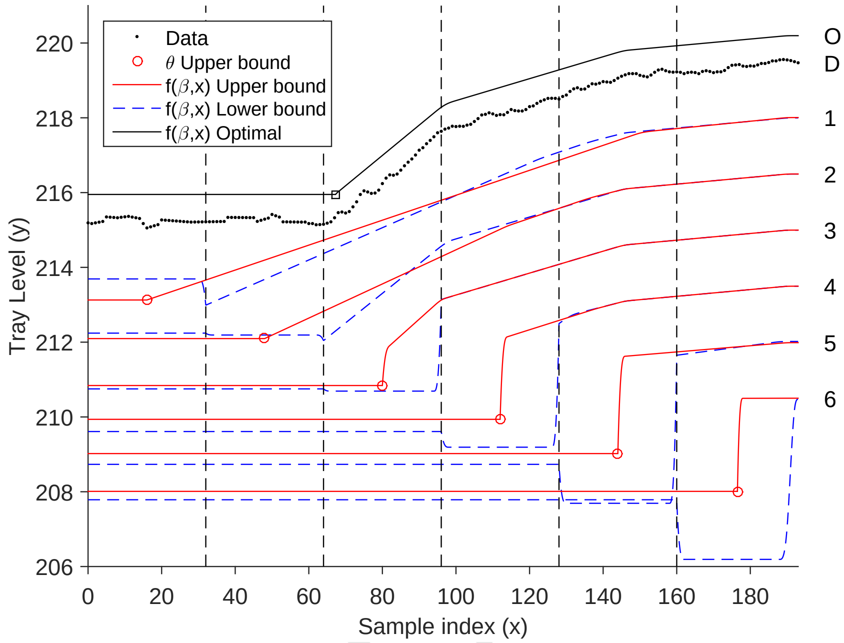


Figure 2: Refinery data set - Data series and fitted functions as a function of time. The data series (D) are shown in their original scale. All functions are shown with an offset to facilitate visualization. Lines below the data series correspond to the functions fitted to compute the upper and lower bounds for six contiguous intervals for θ, indicated by vertical dashed lines and indexed 1 to 6 from top to bottom. Red circles indicate the transitions as applied for the upper bound computation ($\hat{\theta}^{QP}$). The full black line (O) above the data series corresponds to the globally optimal shape-constrained spline function. The global optimum for θ is indicated by a black square at 67.2813.

Optimal fitting requires solving for $\boldsymbol{\beta}$ and θ. The problem, as indicated above, is convex in $\boldsymbol{\beta}$ for a given value of θ and is then solved efficiently by means of interior-point optimization. This was executed for a grid of values for θ spaced equidistantly over the domain of the spline function. The resulting objective function (SSR) is shown in Fig. 3. Such a brute force approach *(i)* is naturally inefficient, *(ii)* does not guarantee global optimality, and *(iii)* does not scale well with the number of transitions to be optimized.

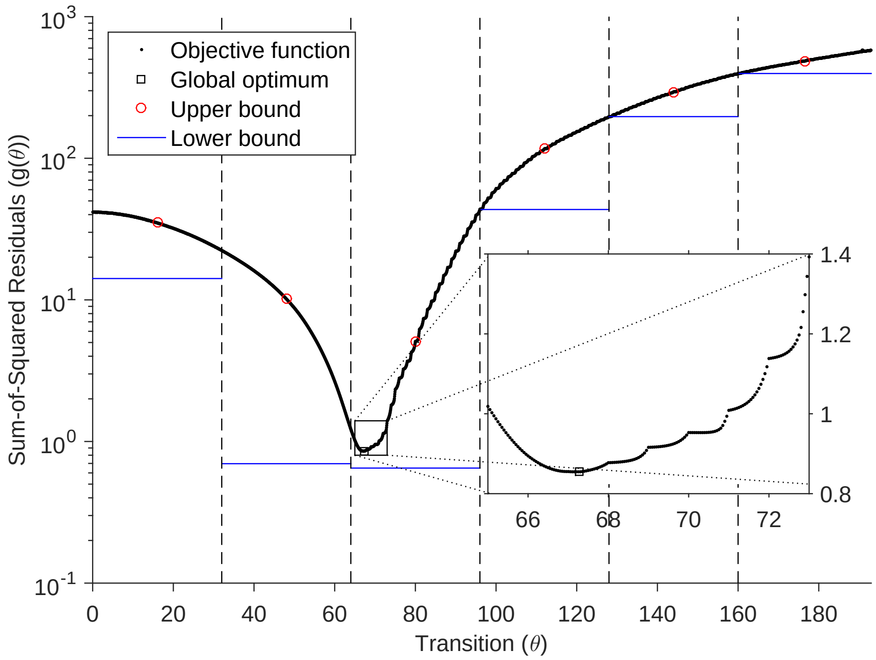


Figure 3: Refinery data set - *(i)* Objective function as a function of the transition (θ) evaluated in an equidistant grid with steps of 0.05 and *(ii)* upper and lower bounds to the objective function for six intervals. The upper and lower bounds to the objective function are found consistent by visual inspection.

The shape-constrained spline fitting problem is solved in a better way by means of the branch-and-bound algorithm. The suitability of this algorithm depends primarily on the validity of the applied bounding procedures. This validity is demonstrated first. To this end, the bounds to the objective function are computed for six different solution sets for the transition, namely the following contiguous intervals of the argument range: $[0, 32]$, $[32, 64]$, $[64, 96]$, $[96, 128]$, $[128, 160]$, and $[160, 193]$. For each of these intervals, the upper bound solution for the transition ($\hat{\theta}^{QP}$) equals the center of the interval (see Appendix B). The function parameters ($\boldsymbol{\beta}$) are optimized for both the upper and lower bound (see Appendix B, Section B.2). The corresponding spline functions are shown in Fig. 2. Fig. 3 plots the corresponding upper and lower bounds to the objective function. In every interval the lower bound for every interval is lower than any value obtained for the objective function in the considered interval. At the same time, the upper bound is effectively equal or higher than at least one objective function value in the considered interval. This demonstrates the bounding procedures given in Appendix B.

The top panel of Fig. 4 visualizes the execution of the branch-and-bound algorithm to find the optimal value for θ by displaying the solution sets generated by this branch-and-bound algorithm as a function of its iteration count. After a total of 24 steps, the last live node corresponds to an interval of width 0.003125 (1/32) and the optimization algorithm is halted. The best upper bound solution is found for $\theta = 67.2813$. The lowest values for the upper and lower bounds among the live nodes are shown as function of the algorithm iterations in the bottom panel of Fig. 4. Here one can see that the lower bound converges monotonically to its final value. In contrast, the upper bound does not decrease monotonically. This is because our implementation of the algorithm does not keep memory of the upper bound solutions. The corresponding globally optimal spline function is shown in Fig. 2.

3.2. Data Set 2: Oxidation-Reduction Potential (ORP) Data

3.2.1. Visual Data Inspection

A subset of 21 normal ORP time series which span the complete data set is shown in Fig. 5. Thanks to offset visualization, it is easy to see that they are all very similar in shape. More specifically, the displayed time series can be described in a rough fashion by means of the QS EAC with a discontinuity in the first and second derivatives at both transitions. The transitions between the episodes are close to changes in batch stages within the first 80 minutes of each cycle, namely the change from the liquor exchange stage to the feeding stage (7') and from the feeding stage to the aerobic oxidation stage (17'). It can also be seen that the curvature of the A primitive becomes less pronounced in batches at the end of the data set. The same data are shown in their original scale in the *Supplementary Materials* (Fig. S.1). The latter figure clearly suggests a mean

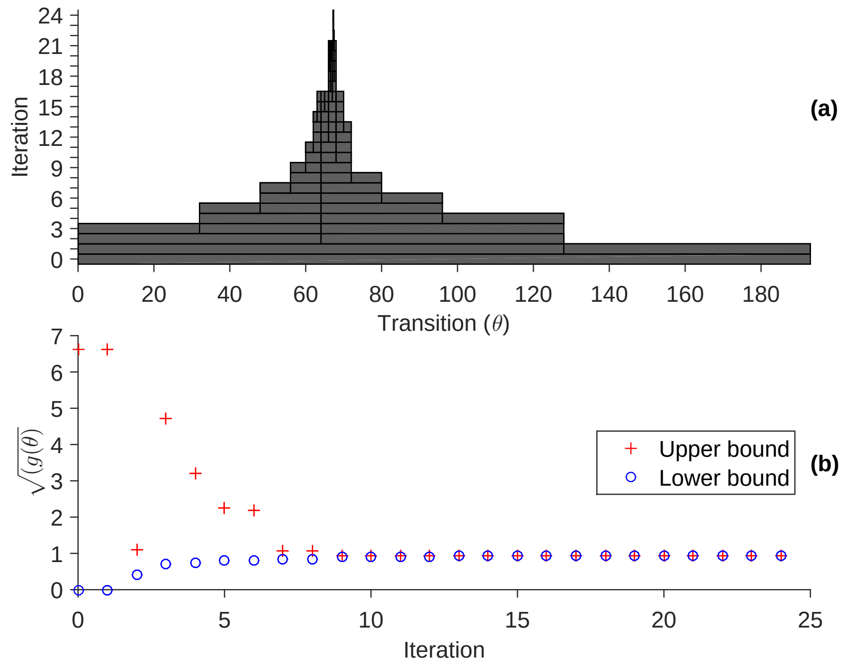


Figure 4: Refinery data set - Visualization of the progress of the branch-and-bound algorithm. *(a)* Solution tree. For each iteration of the algorithm, live leaf nodes are shown as shaded rectangles. *(b)* Upper and lower bounds to the objective function as computed during execution of the branch-and-bound algorithm.

shift occurring between cycles 400 and 480. Detailed inspection (not shown) indicates that the mean shift occurs between cycles 410 and 411, which is when the ORP sensor was maintained and calibrated.

3.2.2. Shape-Constrained Spline Function Fitting

For the ORP data series, a natural cubic spline function with knots in every data sample is fitted in the least squares sense. The objective function is thus

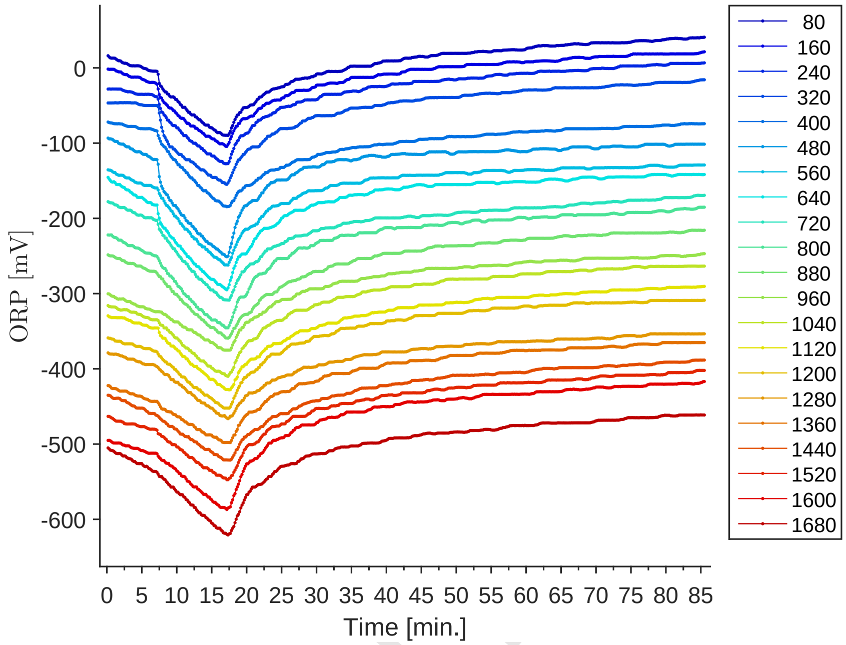


Figure 5: Exemplary data series for normal operating conditions. Data of every 80^{th} are displayed. The time series for cycle 80 is in its original scale. All other time series are shown with an offset for convenient visualization. An EAC sequence appears to be a good qualitative description of these time series.

parameterized as follows:

$$\boldsymbol{\kappa}_{fix} = \boldsymbol{x} \tag{22}$$

$$\forall j \in \{0, \ldots, r\}: \ \lambda_j = 0 \tag{23}$$

$$p = 2 \tag{24}$$

$$r = 3 \tag{25}$$

This means that the fitted functions would be perfectly fitting interpolating spline functions without the application of shape constraints. The QS EAC

corresponds to the following sign matrix:

$$\boldsymbol{S} = \begin{bmatrix} ? & -1 & 0 & ? \\ ? & -1 & +1 & ? \\ ? & +1 & -1 & ? \end{bmatrix} \tag{26}$$

Both transitions are associated with discontinuities in both the first and second derivative which require a knot with multiplicity 3 to be placed in the transitions. This is described as follows in mathematical terms:

$$\boldsymbol{\kappa}_{var} = \boldsymbol{\theta} = \boldsymbol{\delta} = \begin{bmatrix} \kappa_1 & \kappa_2 \end{bmatrix}^T = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}^T = \begin{bmatrix} \delta_1 & \delta_2 \end{bmatrix}^T \tag{27}$$

$$\boldsymbol{c}_{var} = \begin{bmatrix} c_{var,1} & c_{var,2} \end{bmatrix}^T = \begin{bmatrix} 0 & 0 \end{bmatrix}^T \tag{28}$$

The optimal solution for $\boldsymbol{\theta}$ is obtained by means of the branch-and-bound algorithm explained in Appendix B. Fig. 6 displays the executed branching steps of the algorithm when executed for batch 37. The algorithm is halted when all dimensions of the live nodes (subsets) are smaller than 0.125 (= 1/8). This occurs after 29 branching steps. The minimum for $g(\boldsymbol{\beta}, \boldsymbol{\theta})$ is equal to 433 mV^2 and is found at $\hat{\boldsymbol{\theta}} = \begin{bmatrix} 7' \, 5'' & 17' \, 18'' \end{bmatrix}^T$. Fig. 7 shows the spline function obtained. It fits the data very well and the resulting QR matches the earlier visual inspection well. The residuals are visibly small and some auto-correlation is apparent. This is most visible after the second transition, matching the start of the aerated stage closely. Within this stage, an on-off (*i.e.* bang-bang) controller actively controls the dissolved oxygen concentration leading to corresponding oscillations in the ORP signal. These oscillations are typically small and are not analyzed further within this work. Despite the coarse approximation of the ORP signal that results, good detection performances are reported below.

3.2.3. Sum-of-Squared-Residuals and Anomaly Detection

The optimization of the shape-constrained spline function as described above is repeated for every time series in the data set. The top panel of Fig. 8 displays the resulting SSRs as a function of the batch cycle index (SSR_{SCS}). This

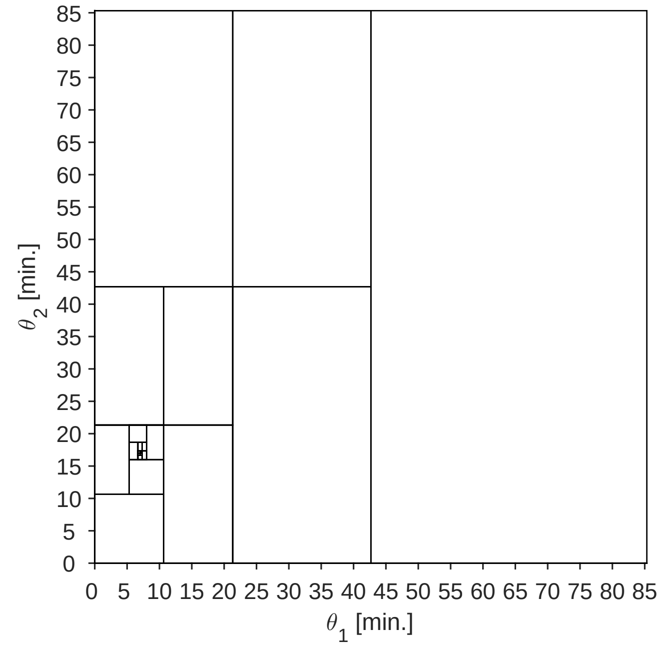


Figure 6: Visualization of the branch-and-bound algorithm. Each vertical (horizontal) line represents a branching step which splits a parent node along θ_1 (θ_2).

statistic ranges between 241 mV^2 and $4.9 \cdot 10^5$ mV^2. The maximum SSR_{SCS} value obtained for a normal cycle is 977 mV^2. An initial assessment of the performance of this method is obtained by setting the UCL equal to the latter value. This is the lowest possible limit leading to a zero FPR. Of the 96 abnormal time series, 56 are then positively detected (TPR: 58%).

The bottom panel of Fig. 8 shows the SSR statistic obtained with PCA modeling (SSR_{PCA}). The calibration set consists of the data from the first 100 normal batch cycles. Two PCs were selected on the basis of scree plots (*Supplementary Materials*, Fig. S.2). This PCA model captures 96.6% of the total variance of the calibration set and its loading vectors are shown in the *Supplementary Materials*, Fig. S.3).

The SSR_{PCA} statistic ranges from 446 mV^2 to $9 \cdot 10^6$ mV^2. The maximum

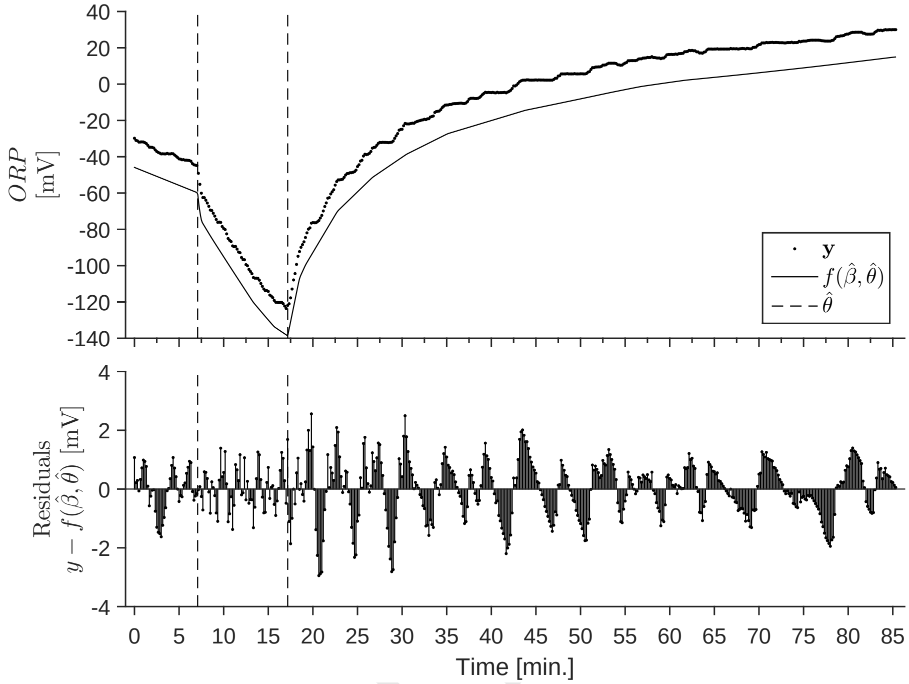


Figure 7: Optimal shape-constrained spline fitting for batch 37. Top: data ($\boldsymbol{y}$, original scale) and fitted function ($f(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$, shown with offset). Bottom: Residuals.

SSR for the normal cycles is $7.37 \cdot 10^5$ mV^2. For each normal cycle, the SSR_{PCA} is higher than the corresponding SSR_{SCS}, indicating that the SCS data model fits the normal data better than the selected PCA model can. This is not surprising because the SCS model has 513 parameters (spline coefficients) which can be adjusted whereas the PCA model, once calibrated, has only two principal scores. An interesting phenomenon occurs at batch 411, where the SSR_{PCA} suddenly rises from $2.16 \cdot 10^4$ mV^2 to $3.47 \cdot 10^5$ mV^2. Beyond batch 411, the SSR_{PCA} remains high. This is explained by the mean shift caused by sensor maintenance discussed above. Indeed, as the PCA model was calibrated with data obtained before this maintenance event, it is unlikely that this model can represent data after such an event.

Using the maximum SSR_{PCA} value for the normal cycles as the UCL, eight

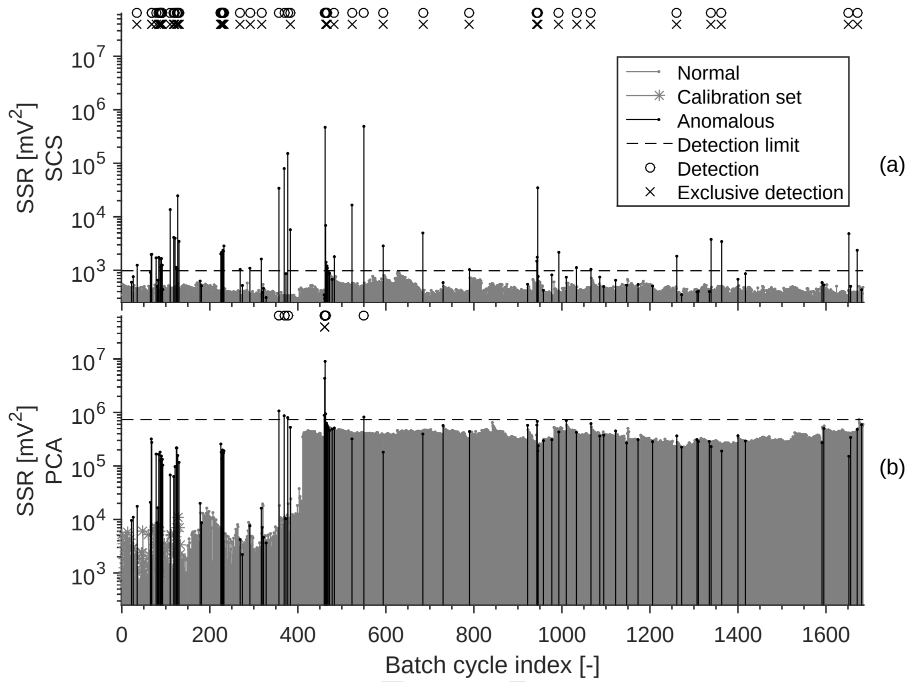


Figure 8: Anomaly detection with a single detection limit for the sum-of-squared residuals statistic. Top: shape-constrained spline fitting - Bottom: principal component analysis. The detection limit is the lowest limit giving no false alarms.

positive detections (TPR: 8.3%) are obtained at the same FPR (0%). Interestingly, these detections include two cycles not detected by the SCS method above. The remaining six cycles are detected by both methods. This also means that 50 out of 56 positive detections with SCS are not obtained with the PCA method. The time series corresponding to positive detections by SCS, PCA, or both methods are displayed separately in the *Supplementary Materials* (Fig. S.4-S.6).

In order to account for the observed mean shift at batch cycle 411, the PCA-based detection is repeated by using a separate PCA model for the cycles before and after the ORP sensor maintenance event. The PCA model for the first 410 time series remains the same as before (Subset 1). The PCA model for the remaining time series (Subset 2) is obtained by selecting the first 100 normal time series following batch 410 as the calibration set. A model with two principal components (PCs) was again selected on the basis of scree plots (see *Supplementary Materials*, Fig. S.7-S.8). This model captures 94.8% of the total variance. The resulting SSR statistic is shown in the bottom panel of Fig. 9. Within subset 1, the resulting SSR_{PCA} statistic ranges from 446 mV^2 to $1.07 \cdot 10^6$ mV^2 and the maximal value for normal batches is $3.77 \cdot 10^4$ mV^2. Within subset 2, SSR_{PCA} ranges from 440 mV^2 to $2.59 \cdot 10^7$ mV^2 and exhibits a maximum for the normal cycles at $1.05 \cdot 10^5$ mV^2. All values for SSR_{PCA} are higher than the corresponding values for SSR_{SCS}, except for a single time series (batch 491). This indicates that the PCA models still deliver a worse fit than the SCS method in general.

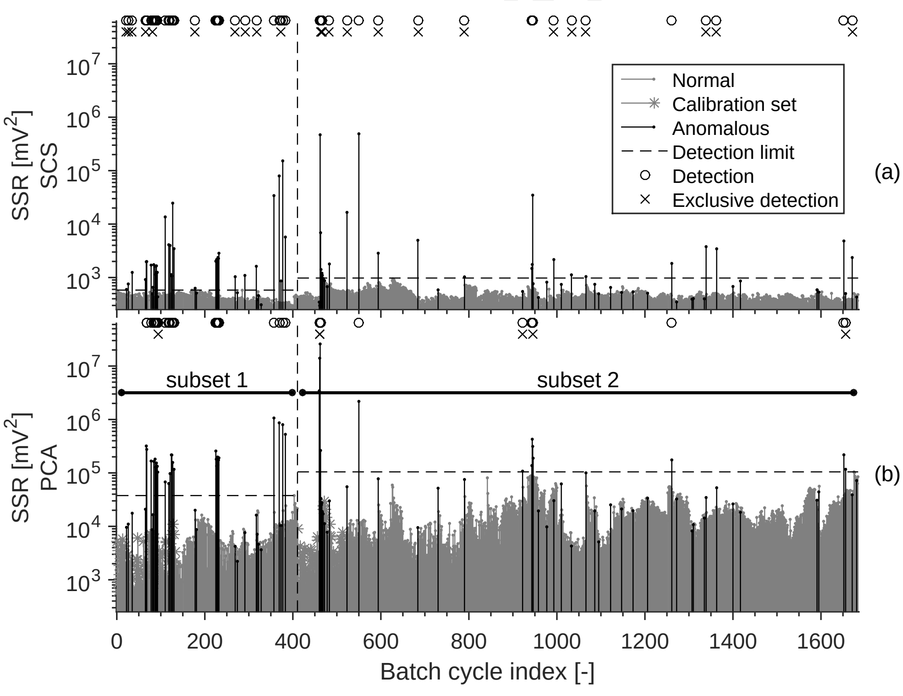


Figure 9: Top: shape-constrained spline fitting - Bottom: principal component analysis. The detection limits are the lowest ones giving no false alarms in each of the two considered periods.

Both PCA models are again evaluated first by setting the UCL to the highest SSR_{PCA} obtained for the normal cycles. This is done separately for subsets 1 and 2. In subset 1, 30 out of 46 cycles are now detected (TPR: 65%). In subset 2, 13 out of 50 abnormal batches are positively detected (TPR: 26%). Constructing two separate PCA models thus improves the model performance dramatically. For a fair comparison with the SCS method, the SCS method is now evaluated by setting different UCLs for each subset. However, the original SSR_{SCS} values are used (Fig. 8, top panel). The resulting UCLs are 581 mV^2 (subset 1) and 977 mV^2 (subset 2). This leads to a positive detection of 39 out of 46 batches in subset 1 (TPR: 84%) and 23 out of 50 abnormal batches in subset 2 (TPR: 46%). The SCS and PCA methods both detect 29 abnormal batches in subset 1 and eight (8) in subset 2. The SCS method identifies ten (10) abnormal time series not identified by the PCA method in subset 1 and 15 in subset 2. The PCA method leads to the exclusive detection of one abnormal time series in subset 1 and five (5) in subset 2. The time series exclusively identified by the SCS or PCA method and those identified by both methods are shown separately for each subset in the *Supplementary Materials* (Fig. S.9-S.14).

3.2.4. Receiver-Operator-Characteristic

The above paragraphs permitted a comparative analysis by discussing detection results which were all obtained with a zero FPR. However, the corresponding UCL is an unlikely choice as a typical approach is to trade off false positives against false negatives. Where this trade-off lies, is seldom known exactly as it involves an assessment of the frequencies of normal and abnormal conditions and the associated costs and benefits, all of which are hard to assess, as abnormal conditions tend to be rare and diverse in nature. In order to compare anomaly detection methods without specifying the trade-off, the ROC can be used as described above. The ROC is computed for the cases studied so far, *i.e.* for both methods and for the complete data set (global), subset 1, and subset 2. Batch cycles included in the PCA calibration sets are excluded from this evaluation.

All ROCs are shown in Fig. 10. The left-hand side of the graph shows the TPRs corresponding to an FPR of 0% which were discussed above. As the UCL is decreased, both TPR and FPR increase. The black diagonal line is the expected performance for a random classifier. A good anomaly detection method should deliver high TPRs and low FPRs. It follows that the first PCA model applied over the whole data set leads to a TPR lower than the FPR for FPRs from 32.2% to 83%. In contrast, the SCS method applied to the whole data set leads to better than random performance results for every possible choice for the FPR. At conventional choices of 1%, 5%, and 10% for the FPR, the TPRs are 64.6%, 72.9%, and 77.1%. The SCS method delivers combinations of TPR and FPR values which are simultaneously better than any combination obtained with PCA, except for FPRs from 94.7% to 96.8%, where the TPR is equal to 97.9% with SCS and 99.0% with PCA.

The use of separate PCA models for subsets 1 and 2 leads to universally improved ROCs compared to the single PCA model approach. The effect is most dramatic for subset 1 and delivers an ROC which is similar to that for the SCS method (SCS_{global}). At 1%, 5%, and 10% FPRs, the TPRs are 65.2%, 69.6%, and 76.1%. For subset 2, the corresponding FPRs are 28%, 42%, and 70%. By choosing a separate UCL for each subset in the case of the SCS method, a high performance is obtained for subset 1. For FPRs of 1%, 5%, and 10%, TPRs of 84.8%, 84.8%, and 89.1% are obtained. For subset 2, the ROC shows a decreased performance compared to the original ROC for the whole data set. The TPRs are 56%, 66%, and 70% for FPRs of 1%, 5%, and 10%. For subset 1, the ROC for SCS completely dominates the ROC for the PCA model, meaning that the SCS method is universally better than PCA. No matter which FPR is chosen, SCS delivers the highest TPR. This is also the case for subset 2, except for FPRs from 72.7% to 99.3% where SCS delivers TPRs from 94% to 98% and PCA leads to TPRs from 96% to 100%.

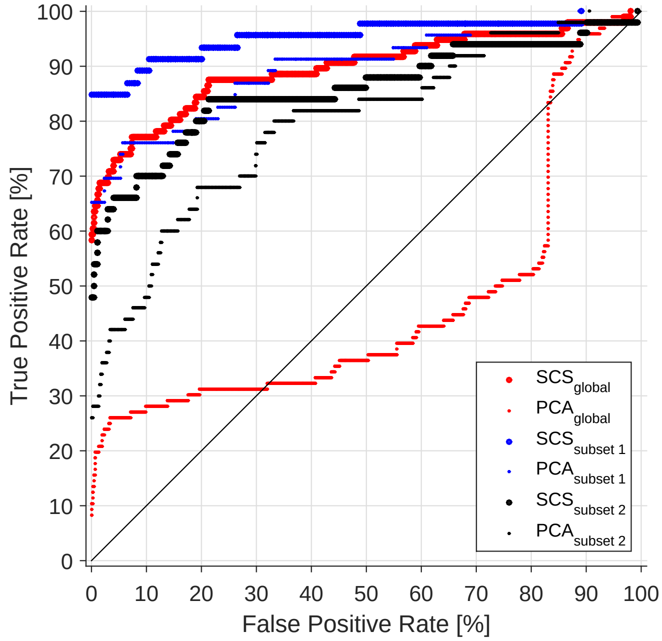


Figure 10: Receiver-Operator Characteristic computed for the complete data set and the two identified subsets. In all cases, the SCS method is preferable over the PCA method at almost all choices for the false positive rate.

4. Discussion

In this work, a modified method based on shape-constrained splines (SCS) is presented and evaluated as a tool for anomaly detection. The results demonstrate that it is feasible to fit spline functions with shape constraints implying discontinuous trends by means of a globally optimal deterministic optimization algorithm. Comparative analysis indicates an almost universally better performance of the SCS method over the more conventional PCA method. This is in part due to a greater flexibility of the fitted function as well as the nonlinear nature of the SCS model. In addition, the SCS method requires a minimal amount of prior information about the process and does not depend on a large representative data set for calibration. The next paragraphs describe a number of limitations of this study and an analysis of the SCS method.

4.1. Limitations of This Study

The following limitations of this study are recognized:

Experimental Laboratory-Scale Data. This study is the first in a recent series of studies on QTA co-authored by the first author in which experimental data are analyzed. While the use of experimental data demonstrates the applicability of the deployed SCS method and its preferred performance, the reported benefits in detection performance cannot be guaranteed for every process and for full-scale processes which may be subject to larger operational variability. However, given the earlier work which included benchmarking tests with simulated data, it is our opinion that the method shows great potential as an intuitive tool for anomaly detection in many systems.

Single Instrument and Single Reactor System. This study focuses on the analysis of data obtained with a single ORP sensor installed in the same reactor. Given the demonstrated robustness of the SCS method, it is also considered very valuable to test whether the underlying SCS model remains appropriate for other reactor units and for different instruments (spatial variability). It is hypothesized that this is indeed the case. However, the studied data do not allow this to be demonstrated.

Anomaly Detection Limited to the Lack-of-Fit Statistic. A deliberate choice was made to restrict anomaly detection to the use of sum-of-squared-residuals (SSR). One argument in favor of this decision is that *(i)* a lack-of-fit statistic is exactly the right measure for assessing the degree to which a new data sample matches the applied model. Other measures, *e.g.* Hotelling's T^2 statistic as applied to PCA, are applicable when it is useful to identify extreme data samples which do, however, fit the identified model, *i.e.*the correlation structure, rather well. A second argument is that *(ii)* the SCS method does not come with features, such as the principal scores given by PCA, for which a theoretical distribution can be easily proposed. Although semi-parametric methods, such as kernel density estimation, could be used to describe the normal behavior of the spline function coefficients and/or the identified transitions, this adds a level of complexity which could impede evaluation by the SCS method. Thirdly and lastly, *(iii)* the statistical description of these features would require an extensive calibration data set, which is not a requirement for the SCS method in its current form. Nevertheless, further improvement of the detection performance should be expected if such data are available, representative, and easy to model. Note that the possibility to use the transitions as process indicators has been explored before (Villez et al., 2008).

Incomplete Data Set. The analyzed time series consist of only a fraction of the available data for the studied SBR process. The inclusion of more data was prevented by a number of factors. First of all, the shape of the ORP time series consisting of the first 513 data points is roughly the same for all normal batch cycles with two distinct transitions. Analysis of longer time series results in a rather diverse set of normal QSs. This leads to a dramatic increase in computational time as the required computations increase exponentially with the number of transitions and linearly with the number of alternative QSs which have to be checked against. Although this is shown to be feasible in (Villez et al., 2013), a simpler approach was taken here given the first-time application of the SCS method to a reasonably large experimental data set. Secondly, the SCS method as proposed here does not support the joint analysis of multivariate time series and therefore prevents the inclusion of data originating from other instruments. Thirdly, it was observed that the analysis of the time series covering the complete SBR cycles on the basis of an interpolating cubic spline function led to insufficient memory availability on both desktop and laptop machines tested for this purpose. It remains an open question as to how to deal effectively with *(i)* complex and diverse QSs, *(ii)* multivariate time series, and *(iii)* long time series in the SCS framework.

4.2. Strengths, Weaknesses, Opportunities, and Threats of the SCS method

In the following paragraphs, a detailed assessment of the proposed modified SCS method is given in terms of its strengths, weaknesses, opportunities, and threats (SWOT).

4.2.1. Strengths

Intuitiveness. One of the major advantages of QTA methods like the one presented here, is their intuitive interpretation. In the presented work, the anomaly detection method is essentially based on the computer-based recognition of a known pattern which is tied to normal operating conditions. Importantly, this pattern *(i)* is described in a coarse-grained fashion, and *(ii)* can be established easily by a process operator or by visual inspection of a few normal cycles. In contrast, conventional methods, especially unsupervised ones such as PCA, rely on representative calibration data sets of considerable size and require considerable expertise in statistical process control methods for proper model identification. It may be argued that the SCS method is less of an art and thus not as sensitive to subjective judgments common to the application of unsupervised latent variable models (*e.g.* PCA). Note that this kind of reasoning is similar to the dynamic model identification philosophy in (Shaich et al., 2001).

Optimality. In contrast to alternative QTA techniques, the SCS method solves the pattern recognition problem by means of a deterministic global optimization scheme. This allows avoiding the challenges associated with greedy and stochastic optimization methods such as *(i)* obtaining locally optimal solutions and *(ii)* the need for tuning to increase chances of finding the global optimum.

Statistical Framework. The pattern recognition problem is cast as a maximum likelihood estimation problem. This means that the resulting estimates for the spline function parameters and transitions are consistent estimators as long as the spline basis and the shape constraints are consistent with the true data generating process (*i.e.* the true model is included in the feasible model set).

Furthermore, anomaly detection is based on a lack-of-fit statistic which is similar to those for existing anomaly detection methods, such as PCA. The proposed method can thus be easily integrated in existing statistical process control schemes.

Support of Discontinuities. The proposed extension of the SCS method now allows explicit accounting of discontinuous trends in QTA. Although previously identified methods also allow for such discontinuities (Dash et al., 2004; Charbonnier & Gentil, 2007), these do not automatically lead to the guaranteed continuity of derivatives in function arguments different from the identified transitions.

Robust Data Model. As borne out by the comparative analysis between the anomaly detection performance of the SCS and PCA models, it is apparent that the SCS model represents relationships between the collected data which remain true throughout the data collection period. In contrast, the PCA model requires recalibration following sensor maintenance to recover reasonable detection performance levels. This inherent robustness stems from the fact that the SCS model, with as many spline coefficients as there are data points, is extremely flexible and can thus track both incipient and abrupt numerical changes in the collected data series while clearly rejecting anomalous data not fitting the qualitatively described expectations. The conventionally identified PCA model with two principal components exhibits far fewer degrees of freedom and lacks the flexibility of the SCS model. The SCS data model thus provides a robust approach to anomaly detection. Other use cases which may benefit from this property, such as data reconciliation and missing data estimation, remain to be evaluated.

4.2.2. Weaknesses

Computational Effort. The global optimality of the branch-and-bound search algorithm requires a large computational effort. In the worst case, this effort increases exponentially with the number of optimized transitions. The identification is thus practically limited to a small number of transitions given the computational capacity for a typical wastewater treatment plant.

End-of-batch use only. Due to the computational requirements of the presented method, specifically the optimization of the transitions, the current use of the SCS model is restricted exclusively to end-of-batch use. Exceptions to this are possible however *(i)* when the qualitative sequence consists of a single primitive only (there are no transitions to be optimized) or *(ii)* when the transitions are known exactly.

Shape-based detection only. As discussed above, the proposed method does detect anomalous shifts in time of the identified transitions. Indeed, the SSR statistic computed with the SCS data model only evaluates departure from the assumed profile shape, not its location in time. Still, the changes necessary to enable process monitoring on the basis of the identified transitions are likely limited.

Univariate method. The proposed method is limited to the analysis of univariate data series, similarly to previous studies (*e.g.* Villez et al., 2012, 2013; Villez, 2015) and in contrast to alternative methods based on wavelet and piece-wise polynomial fits (Maurya et al., 2005; Flehmig & Marquardt, 2006). However, a multivariate SCS method is currently being developed.

4.2.3. Opportunities

Knot placement. The extension of the SCS method allows the location of knots to be optimized. In principle the provided bounds also make this possible when no shape constraints are enforced. The branch-and-bound algorithm can thus be used to optimize knot placements. However, alternative methods (*e.g.* Beliakov, 2004) are likely to be more efficient for this purpose. It is unclear whether existing methods for globally optimal knot placement are useful in the context of shape-constrained spline fitting. This can potentially lead to improved bounds for the optimization problem and remains open for exploration.

4.2.4. Threats

Supervised nature. A major drawback of the proposed SCS method is that the targeted QS needs to be specified a priori. The method could easily be extended to allow for multiple permissible sequences (as in Villez et al., 2013; Villez, 2015) but a completely unsupervised application based on shape-constrained function fitting is not yet considered feasible.

5. Conclusions

In this study, shape-constrained spline (SCS) function fitting is proposed as a method for qualitative trend analysis (QTA). For the first time, a QTA method is proposed which deals with discontinuous trends in an explicit way. Furthermore, the resulting QTA method is the first of its kind to produce a confirmatory data model which is useful for fault detection, as opposed to fault diagnosis. This means that the SCS model can be used in a similar way to any other data model, including principal component analysis (PCA), as is demonstrated with this work. The application to a data set obtained from an experimental wastewater treatment pilot plant further indicates that the SCS method leads to tangible improvements in fault detection performance compared to the classic PCA model. However, an even more important benefit is that the SCS method is fairly simple on a conceptual level and easy to explain to process operators and experts, in contrast to more conventional tools such as PCA. This is true despite the relatively complex mathematical forms required to provide a globally optimal deterministic function-fitting algorithm.

Acknowledgments

All results were obtained by the joint use of Matlab (The MathWorks Inc., 2014), the Functional Data Analysis software package by (Ramsay & Silverman, 2002), and the Mosek optimization software (MOSEK ApS, 2012). The authors thank Dr. Nicolas Derlon for his valuable assistance to this study.

References

Alizadeh, F., & Goldfarb, D. (2003). Second-order cone programming. *Math. Program., Ser. B*, *95*, 3–51.

Beliakov, G. (2004). Least squares splines with free knots: global optimization approach. *Applied Mathematics and Computation*, *149(3)*, 783–798.

Boyd, S., & Vandenberghe, L. (2009). *Convex optimization*. Cambridge University Press.

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

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

Cheung, J.-Y., & Stephanopoulos, G. (1990). Representation of process trends – Part I. a formal representation framework. *Comput. Chem. Eng.*, *14*, 495–510.

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

Dürrenmatt, D. J., & Gujer, W. (2012). Data-driven modeling approaches to support wastewater treatment plant operation. *Environmental Modelling & Software*, *30*, 47–56.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 861–874.

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

Habermacher, J., Benetti, A. D., Derlon, N., & Morgenroth., E. (2015). The effect of different aeration conditions in activated sludge - side-stream system on sludge production, sludge degradation rates, active biomass and extracellular polymeric substances. *Water Research*, *85*, 46–56.

Jackson, J., & Mudholkar, G. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, *21*, 341–349.

Joliffe, I. (2002). *Principal component analysis*. (2nd ed.). Springer, New York, USA.

Kresta, J., MacGregor, J., & Marlin, T. (1991). Multivariate statistical monitoring of process operating performance. *Can. J. Chem. Eng.*, *69*, 35–47.

Kuipers, B. (2001). Encyclopedia of physical science and technology. chapter Qualitative Simulation. (pp. 287–300). Academic Press.

Kuipers, B. J. (1994). *Qualitative Reasoning: Modeling and simulation with incomplete knowledge*. Cambridge, MA, USA: MIT Press.

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. *Ind. Eng. Chem. Res.*, *42(20)*, 4789–4810.

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

Maurya, M. R., Rengaswamy, R., & Venkatasubramanian, V. (2007). Fault diagnosis using dynamic trend analysis: A review and recent developments. *Engineering Applications of Artificial Intelligence*, *20*, 133–146.

Montgomery, D. (2005). *Introduction to Statistical Quality Control*. (5th ed.). John Wiley & Sons, Inc., Hoboken, NJ, USA.

MOSEK ApS (2012). *MOSEK Optimization Software for MATLAB, Version 6.0*. Available from http://www.mosek.com/.

Nopens, I., Villez, K., Rieger, L., & Vanrolleghem, P. A. (2007). Monitoring the water cycle - state of the art and future needs. (pp. 33–36). IWA Publishing, London, UK.

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

Ramsay, J. O., & Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies.*. Springer-Verlag.

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

Rieger, L., Takács, I., Villez, K., Siegrist, H., Lessard, P., Vanrolleghem, P. A., & Comeau, Y. (2010). Data reconciliation for wastewater treatment plant simulation studies – planning for high-quality data and typical sources of errors. *Water Environment Research*, *82*, 426–433.

Schuermans, M., Markovsky, I., Wentzell, P., & Van Huffel, S. (2005). On the equivalence between total least squares and maximum likelihood PCA. *Anal. Chim. Acta*, *544*, 254–267.

Shaich, D., Becker, R., & King, R. (2001). Qualitative modelling for automatic identification of mathematic models of chemical reaction systems. *Control Eng. Practice*, *9*, 1373–1381.

Spindler, A., & Vanrolleghem, P. A. (2012). Dynamic mass balancing for wastewater treatment data quality control using CUSUM charts. *Water Science and Technology*, *65*, 2148–2153.

The MathWorks Inc. (2014). *MATLAB Release 2014b*. Natick, Massachusetts.

Thomann, M. (2008). Quality evaluation methods for wastewater treatment plant data. *Wat. Sci. Technol.*, *10*, 1601–1609.

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

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., & Yin, K. (2003b). A review of process fault detection and diagnosis - Part III: Process history based methods. *Computers & Chemical Engineering*, *27*, 327–346.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. N. (2003c). A review of process fault detection and diagnosis - Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, *27*, 293–311.

Villez, K. (2015). Qualitative path estimation: A fast and reliable algorithm for qualitative trend analysis. *AIChE Journal*, *61*, 1535–1546.

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

Villez, K., Rosén, C., Anctil, F., Duchesne, C., & Vanrolleghem, P. A. (2008). Qualitative representation of trends: an alternative approach to process diagnosis and control. *Wat. Sci. Technol.*, *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. *Comp. Chem. Eng.*, *48*, 187–199.

816 # Appendices

817 ## A. Acronyms and Symbols

Table A.1: List of acronyms

Acronym	Full expression
FPR	False positive rate
ORP	Oxidation-reduction potential
PC	Principal component
PCA	Principal component analysis
QP	Quadratic program
QPE	Qualitative path estimation
QR	Qualitative representation
QS	Qualitative sequence
QTA	Qualitative trend analysis
ROC	Receiver-operator-characteristic
SBR	Sequencing batch reactor
SCS	Shape-constrained splines
SPE	Squared prediction error
SSR	Sum of squared residuals
SStR	Side-stream reactor
SWOT	Strengths, weaknesses, opportunities, and threats
TPR	True positive rate
UCL	Upper control limit

Table A.2: Typography

Style	Description
x, x_i, $X_{i,j}$	Scalar
$\boldsymbol{x}$, $\boldsymbol{X}_{\cdot,j}$	Column vector
$\boldsymbol{X}$	Matrix
$\hat{x}$, $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{X}}$	Estimate

Table A.3: Symbol definitions

Symbol	Description	Class
a	Index of nodes generated during optimization	Integer
b	Interval boundary for enforced shape constraints	Continuous
c	Highest continuous derivative	Integer
d	Index for transitions associated with a discontinuity	Integer
e	Index of primitives in QS and episodes in QR	Integer
$f()$	Piece-wise polynomial function	Function
$f^j()$	j^{th} derivative of $f()$	Function
$g()$	Objective function	Function
$h()$	Indicator function	Function
i	Index for data pairs	Integer
j	Index for derivatives	Integer
k	Index for internal knots	Integer
m	Count (number of samples)	Integer
n	Count (number of episodes, knots, samples)	Integer
p	Power for lack-of-fit objective criterion	Integer
q_j	Power for smoothness objective criterion	Integer
r	Degree of the spline function	Integer
s	Sign value	Integer
t	Score	Continuous
v	Integrand	Continuous
x	Function argument	Continuous
y	Measurement	Continuous
$\boldsymbol{P}$	Loading vector matrix	Continuous
$\boldsymbol{R}$	Residuals matrix	Continuous
$\boldsymbol{S}$	Sign matrix	Continuous
$\boldsymbol{T}$	Score matrix	Continuous
$\boldsymbol{Y}$	Data matrix	Continuous

α	Argument for hypothesized spline knot	Continuous
β	Function parameter (polynomial coefficient)	Continuous
δ	Transitions associated with a discontinuity	Continuous
κ	Internal spline knot location	Continuous
λ	Penalty factor	Continuous
θ	Transition parameter	Continuous
Θ	Set for transition vectors	Continuous
Ω	Set for parameter vectors	Continuous

B. Bounding Procedures for Shape Constrained Splines with Discontinuities

The branch-and-bound algorithm as applied in this study proceeds in the same fashion as in Villez et al. (2013) to compute optimal values for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$. This requires the provision of bounds to the objective function over subsets of Θ. The root set to initialize the branch-and-bound algorithm, Θ_0, is defined as follows:

$$\boldsymbol{\theta} \in \Theta_0 \quad \Leftrightarrow \quad \forall t \in \{1, 2, \ldots, n_t\} \; : \; x_1 \leq \theta_t \leq x_n \tag{B.1}$$

This root set, Θ_0, is essentially an n_t-dimensional box (*i.e.* hyper-rectangle or orthotope). Just as in Villez et al. (2013), every branching step splits a parameter set (a.k.a. parent node) into two constitutive and contiguous parameter sets (a.k.a. leaf nodes). Each newly generated node in the solution tree (indexed with a: Θ_a, $a \in \mathbb{N}^+$) can be described in the following form:

$$\boldsymbol{\theta} \in \Theta_a \quad \Leftrightarrow \quad \forall t \in \{1, 2, \ldots, n_t\} \; : \; \theta_t^L \leq \theta_t \leq \theta_t^U \tag{B.2}$$

In our implementation of the algorithm, each parent node is halved along its longest dimension during branching. This means that the set is halved along the dimension (t) with the lowest range ($\theta_t^U - \theta_t^L$). It follows that each set in the solution tree is a subset of the root of the solution tree:

$$\forall a \in \mathbb{N}^0 \quad : \quad \Theta_a \subseteq \Theta_0 \tag{B.3}$$

The bound computations in the branch-and-bound optimization scheme are based on semi-definite programming which permits the optimal solution for $\hat{\boldsymbol{\beta}}$ to be computed given a feasible solution for $\hat{\boldsymbol{\theta}}$. Indeed, shape-constrained spline function fitting with given transitions is a convex optimization problem which can be solved effectively by interior-point solvers (Alizadeh & Goldfarb, 2003). The lower bound solution as defined below is based on sufficient relaxations to prove the bounds, as will be shown below. However, these relaxations are not known to be strictly necessary. In addition, it is unlikely that the gap between the upper and lower bounds can be driven to absolute zero (unlike the case in Villez et al., 2013). Thus, finite computation times are only guaranteed when the branch-and-bound algorithm is allowed to terminate after the gap between the bounds and/or the size of the solutions sets (nodes) have reached critical numerical tolerances. These tolerances can be set to arbitrarily small strictly positive values. Upon termination, optimal values for $\hat{\boldsymbol{\beta}}$ are given as those values associated with the best known upper bound solution for $\hat{\boldsymbol{\theta}}$. Further notes on the optimization algorithm can be found in Villez et al. (2013). The following paragraphs discuss the bounding procedures.

In order to prove the applied bounds of the objective function, the following definitions are required. The applied set of spline knots consists of a set of n_{fix} fixed knots, ($\boldsymbol{\kappa}_{fix}$) and a set of n_{var} variable knots ($\boldsymbol{\kappa}_{var}$). The latter set of knots corresponds to those transitions in the considered Qualitative Sequence (QS) which imply a discontinuity in the fitted function and/or one or more of its derivatives. The number of variable knot locations, n_{var}, is thus equal to the number of transitions in the QS implying a discontinuity, n_d. The complete set of transitions is now described mathematically as follows:

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \dots & \theta_j & \dots & \theta_{n_t} \end{bmatrix}^T \tag{B.4}$$

It follows for this set that:

$$\boldsymbol{\kappa}_{var} = \boldsymbol{\delta} \subseteq \boldsymbol{\theta} \tag{B.5}$$

The set of fixed knots and variable knots is mutually exclusive so that:

$$\boldsymbol{\kappa} = \boldsymbol{\kappa}_{fix} \cup \boldsymbol{\kappa}_{var} = \boldsymbol{\kappa}_{fix} \cup \boldsymbol{\delta} = \begin{bmatrix} \kappa_1 & \kappa_2 & \dots & \kappa_k & \dots & \kappa_{n_k} \end{bmatrix}^T \tag{B.6}$$

$$\boldsymbol{\kappa}_{fix} \cap \boldsymbol{\kappa}_{var} = \boldsymbol{\kappa}_{fix} \cap \boldsymbol{\delta} = \varnothing \tag{B.7}$$

The proof of the bounding procedures is easier to follow when the fitted spline function is formulated explicitly as a piece-wise polynomial function. This means that a piece-wise polynomial basis is used and the function is parametrized by its polynomial coefficients and not –as in a more conventional approach– by its spline function coefficients. It also means that continuity of the function and its existing derivatives in the knots is only achieved by formulating continuity constraints explicitly as equality constraints in the mathematical problem formulation. To this end, a degree of continuity, c_k, is associated with each knot in $\boldsymbol{\kappa}$ and specifies the highest derivative which is continuous in the considered knot. The degree of the highest derivative which remains continuous for the k^{th} variable knot location is given as $c_{var,k}$. Without any discontinuities this integer is equal to $r-1$ for all knots. Indeed, the r^{th} derivative of any spline function is piece-wise linear with discontinuities in the knot locations. This problem formulation, while apparently inefficient, allows the variable location of knots to be accounted for.

The above specifications lead to the following mathematical formulation of the shape-constrained spline fitting problem:

$$\min_{\boldsymbol{\beta},\boldsymbol{\theta}} g(\boldsymbol{\beta},\boldsymbol{\theta}) \quad = \quad \sum_i |y_i - f(\boldsymbol{\beta}, x_i)|^p + \sum_{j=0}^{j=r} \lambda_j \int_{x_1}^{x_n} \left| f^j(\boldsymbol{\beta}, v) \right|^{q_j} \, dv \quad \text{(B.8)}$$

subject to:

$$\begin{aligned}
&\forall k \in \{1, 2, \ldots, n_k\}, \\
&\forall j \in \{0, 1, \ldots, c_k\} \quad : \quad \lim_{v \to \kappa_k^-} f^j(\boldsymbol{\beta}, v) = \lim_{v \to \kappa_k^+} f^j(\boldsymbol{\beta}, v)
\end{aligned} \tag{B.9}$$

$$\begin{aligned}
&\forall e \in \{1, 2, \ldots, n_e\}, \\
&\forall j \in \{0, 1, \ldots, r\} \quad : \\
&b_e^L \leq v \leq b_e^U \quad \Rightarrow \quad \begin{cases} f^j(\boldsymbol{\beta}, v) \leq 0, & \textit{if } s_{e,j} = -1 \\ f^j(\boldsymbol{\beta}, v) = 0, & \textit{if } s_{e,j} = 0 \\ f^j(\boldsymbol{\beta}, v) \geq 0, & \textit{if } s_{e,j} = +1 \end{cases}
\end{aligned} \tag{B.10}$$

$$\begin{aligned}
\boldsymbol{b}^L &= \begin{bmatrix} b_1^L & b_2^L & \ldots & b_{n_e-1}^L & b_{n_e}^L \end{bmatrix}^T \\
&= \begin{bmatrix} x_1 & \theta_1 & \ldots & \theta_{n_t-1} & \theta_{n_t} \end{bmatrix}^T
\end{aligned} \tag{B.11}$$

$$\begin{aligned}
\boldsymbol{b}^U &= \begin{bmatrix} b_1^U & b_2^U & \ldots & b_{n_e-1}^L & b_{n_e}^U \end{bmatrix}^T \\
&= \begin{bmatrix} \theta_1 & \theta_2 & \ldots & \theta_{n_t} & x_n \end{bmatrix}^T
\end{aligned} \tag{B.12}$$

$$\forall t \in \{1, 2, \ldots, n_t - 1\} \quad : \quad \theta_t \leq \theta_{t+1} \tag{B.13}$$

$$\forall t \in \{1, 2, \ldots, n_t\} \quad : \quad \theta_t^L \leq \theta_t \leq \theta_t^U \tag{B.14}$$

with definitions as in Table A.3.

To enable bounding of the objective function for the optimization problem described by Eq. B.8–B.14 over any set, Θ_a, two cases must be considered. These are:

1. No feasible solution for $\boldsymbol{\theta}$ exists in Θ_a
2. A feasible solution for $\boldsymbol{\theta}$ exists in Θ_a

Two procedures are available to determine whether a feasible parameter set exists. One consists of applying formal methods such as solving feasibility problems (Boyd & Vandenberghe, 2009). A more conventional and intuitive procedure consists by finding a parameter set which minimizes the following quadratic objective function over the set Θ_a:

$$\min_{\boldsymbol{\theta}} \sum_{t=1}^{n_t} \left(\theta_t - \theta_t^L\right)^2 + \left(\theta_t - \theta_t^U\right)^2 \tag{B.15}$$

subject to (B.13)-(B.14). If no solution can be found to this quadratic program (QP), it follows that no feasible solution exists. If a solution has been found, it can applied to compute an upper bound to the shape-constrained spline fitting problem as will be shown below. In this case, the values obtained for $\boldsymbol{\theta}$ are further referred to as $\hat{\boldsymbol{\theta}}^{QP}$. This optimization strategy is followed because of its intuitiveness and convenience.

B.1. Case 1: Infeasible Problem

In the first case, both lower and upper bounds to the objective function value are set to infinity:

$$g^L = g^U = +\infty \tag{B.16}$$

The proof is rather trivial. Indeed, if no feasible solution can be found for $\boldsymbol{\theta}$, then there is no solution with any objective function value lower than $+\infty$. This automatically also defines the upper bound at the same infinitely large value.

B.2. Case 2: Feasible Problem

In the second case, a feasible solution is given by $\hat{\boldsymbol{\theta}}^{QP}$. Bounds to the objective function can then be computed as described below.

B.2.1. Upper Bound to the Objective Function

Procedure. In this case, it remains relatively trivial to evaluate an upper bound. The problem in Eq. B.8–B.14 is solved after replacement of $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}^{QP}$. This leads to the following convex optimization problem in $\boldsymbol{\beta}$:

$$\min_{\boldsymbol{\beta}} g(\boldsymbol{\beta}) \quad = \quad \sum_i |y_i - f(\boldsymbol{\beta}, x_i)|^p + \sum_{j=0}^{j=r} \lambda_j \int_{x_1}^{x_n} \left| f^j(\boldsymbol{\beta}, v) \right|^{q_j} \, dv \tag{B.17}$$

subject to:

$$\forall j \in \{0,1,\ldots,c_k\}, \quad \forall k \in \{1,2,\ldots,n_k\} \quad : \quad \lim_{v\to\kappa_k^-} f^j(\boldsymbol{\beta},v) = \lim_{v\to\kappa_k^+} f^j(\boldsymbol{\beta},v) \tag{B.18}$$

$$\forall e \in \{1,2,\ldots,n_e\}, \quad \forall j \in \{0,1,\ldots,r\} \quad :$$

$$b_e^L \leq v \leq b_e^U \quad \Rightarrow \quad \begin{cases} f^j(\boldsymbol{\beta},v) \leq 0, & \textit{if } s_{e,j} = -1 \\ f^j(\boldsymbol{\beta},v) = 0, & \textit{if } s_{e,j} = 0 \\ f^j(\boldsymbol{\beta},v) \geq 0, & \textit{if } s_{e,j} = +1 \end{cases} \tag{B.19}$$

$$\boldsymbol{b}^L = \begin{bmatrix} b_1^L & b_2^L & \ldots & b_{n_e}^L \end{bmatrix}^T = \begin{bmatrix} x_1 & \hat{\theta}_1^{QP} & \ldots & \hat{\theta}_{n_t}^{QP} \end{bmatrix}^T \tag{B.20}$$

$$\boldsymbol{b}^U = \begin{bmatrix} b_1^U & b_2^U & \ldots & b_{n_e}^U \end{bmatrix}^T = \begin{bmatrix} \hat{\theta}_1^{QP} & \ldots & \hat{\theta}_{n_t}^{QP} & x_n \end{bmatrix}^T \tag{B.21}$$

This optimization completes the computation of an upper bound (g^U).

Proof. The objective function value for the computed solution is indeed an upper bound since the existence of the associated solution proves that at least one solution has an objective function value equal to or lower than g^U.

B.2.2. Lower Bound to the Objective Function

Procedure. In the following, three sufficient relaxations leading to a provable lower bound are discussed.

Relaxation 1. To describe the computation of the lower bound, the relaxations used in Villez et al. (2013) for the continuous case are also applied here. Specifically, this means that the upper bound problem is solved, except that equations Eq. B.20–B.21 are replaced with the following equations:

$$\begin{aligned} \boldsymbol{b}^L &= \begin{bmatrix} b_1^L & b_2^L & \ldots & b_{n_e-1}^L & b_{n_e}^L \end{bmatrix}^T \\ &= \begin{bmatrix} x_1 & \theta_1^U & \ldots & \theta_{n_t-1}^U & \theta_{n_t}^U \end{bmatrix}^T \end{aligned} \tag{B.22}$$

$$\begin{aligned} \boldsymbol{b}^U &= \begin{bmatrix} b_1^U & b_2^U & \ldots & b_{n_e-1}^U & b_{n_e}^U \end{bmatrix}^T \\ &= \begin{bmatrix} \theta_1^L & \theta_2^L & \ldots & \theta_{n_t}^L & x_n \end{bmatrix}^T \end{aligned} \tag{B.23}$$

In words, the lower (upper) bounds for the intervals over which shape constraints are implemented are replaced by the upper (lower) bounds for the transition arguments. This relaxation reduces the argument intervals over which the shape constraints are enforced. Importantly, it follows that the shape constraints applied for the lower bound are also applied when solving with any feasible set of values for $\boldsymbol{\theta}$ within the considered set:

$$\forall a \in \mathbb{N}^0, \forall e \in \{1, 2, \ldots, n_e\}, \forall \boldsymbol{\theta}^{QP} \in \Theta_a : \left[b_e^L, b_e^U\right]^L \subseteq \left[b_e^L, b_e^U\right]^U \tag{B.24}$$

The shape constraints in the modified lower bounding problem are always included in the original problem for any choice of $\boldsymbol{\theta}$ within the considered solution set. The obtained objective function following this relaxation, denoted here as g_1^L, is thus guaranteed to be lower than or equal to the computed upper bound, g^U:

$$g_1^L \leq g^U \tag{B.25}$$

As long as no discontinuities are implied by the qualitative sequence, this relaxation is sufficient to obtain a provable lower bound (Villez et al., 2013). In the more general case where some of the transitions imply the presence of a knot with multiplicity, this is not a sufficient relaxation to obtain a provable lower bound. However, two further relaxations are sufficient to achieve this. These are explained below.

Relaxation 2. The second relaxation consists of adding knots with multiplicity to the set of knots implemented for the upper bound solution. For any transition, with index d, the corresponding index of the highest continuous derivative is denoted as $c_{var,d}$. The solution for the transition argument obtained for the upper bound is written as $\hat{\delta}_d^{QP}$. Whereas $c_{var,d}$ knots are placed in $\hat{\delta}_d^{QP}$ for the upper bound, $r + 1$ ($r + 1 \geq c_{var,d}$) knots are now placed in the same argument. This means that the spline function and all its derivatives are discontinuous in this location. As a result, the piece-wise polynomial function fitting is now completely separable since the data and polynomial coefficients on the left (right) hand side, of each discontinuity argument, $\hat{\boldsymbol{\theta}}_d^{QP}$, have no influence on the polynomial coefficients for the right (left) hand side. This also implies additional degrees of freedom for the piece-wise polynomial function. In general, the number of applied constraints is either reduced or remains the same while the objective function itself remains unchanged. The resulting objective function, referred to as g_2^L, is then lower than or equal to the previously defined objective function value:

$$g_2^L \leq g_1^L \leq g^U \tag{B.26}$$

Relaxation 3. A third relaxation is required to obtain a provable lower bound. It consists of minimizing the following modified objective function:

$$\begin{aligned} \min_{\boldsymbol{\beta}} g(\boldsymbol{\beta}) &= \sum_i h(x_i) \cdot |y_i - f(\boldsymbol{\beta}, x_i)|^p \\ &+ \sum_{j=0}^{j=r} \lambda_j \int_{x_1}^{x_n} h(v) \cdot \left| f^j(\boldsymbol{\beta}, v) \right|^{q_j} dv \end{aligned} \tag{B.27}$$

with:

$$h(x) = \begin{cases} 0, & \text{if } \exists d : \theta_d^L \leq x \leq \theta_d^U \\ 1, & \text{otherwise} \end{cases} \tag{B.28}$$

In words, the residuals corresponding to data points lying within an interval defining the potential location of any transition implying knot multiplicity are not accounted for in the objective function. In addition, the integrals to compute smoothness penalty functions are only integrated over intervals which are guaranteed not to contain a transition implying knot multiplicity for any feasible solution. The resulting objective function value, g_3^L, is naturally lower than or equal to all previously defined objective function values:

$$g_3^L \leq g_2^L \leq g_1^L \leq g^U \tag{B.29}$$

The combined relaxations discussed above are sufficient to obtain a provable lower bound. This is proven in the following paragraphs. The underlying principle of the proof is that adding a discontinuity in any additional feasible location cannot lower the computed objective function value below values given by g_3^L (maximal relaxation).

Proof. To prove the lower bound, consider first that the objective function can be rewritten as a sum of terms associated with three contiguous and non-overlapping intervals of the function domain by using the bounds for the argument location of a discontinuity, δ_d^L and δ_d^U, as interval boundaries:

$$\begin{aligned}
\min_{\boldsymbol{\beta}} \quad g(\boldsymbol{\beta}) = & \sum_{i:x_i \in \left[x_1, \delta_d^L\right]} h(x_i) \cdot |y_i - f(\boldsymbol{\beta}, x_i)|^p \\
+ \quad & \sum_{i:x_i \in \left[\delta_d^L, \delta_d^U\right]} h(x_i) \cdot |y_i - f(\boldsymbol{\beta}, x_i)|^p + \sum_{i:x_i \in \left[\delta_d^U, x_n\right]} h(x_i) \cdot |y_i - f(\boldsymbol{\beta}, x_i)|^p \\
+ \quad & \sum_{j=0}^{j=r} \lambda_j \int_{x_1}^{\delta_d^L} h(v) \cdot \left| f^j(\boldsymbol{\beta}, v) \right|^{q_j} \, dv + \sum_{j=0}^{j=r} \lambda_j \int_{\delta_d^L}^{\delta_d^U} h(v) \cdot \left| f^j(\boldsymbol{\beta}, v) \right|^{q_j} \, dv \\
+ \quad & \sum_{j=0}^{j=r} \lambda_j \int_{\delta_d^U}^{x_n} h(v) \cdot \left| f^j(\boldsymbol{\beta}, v) \right|^{q_j} \, dv
\end{aligned} \tag{B.30}$$

Upon explicit evaluation of $h(x_i)$ and $h(v)$ (Eq. B.28) it can be observed that the 2^{nd} and 5^{th} terms are equal to zero, so that the optimization problem above is equivalent to the following:

$$\begin{aligned}
\min_{\boldsymbol{\beta}} \quad g(\boldsymbol{\beta}) = & \sum_{i:x_i \in \left[x_1, \delta_d^L\right]} h(x_i) \cdot |y_i - f(\boldsymbol{\beta}, x_i)|^p \\
+ \quad & \sum_{i:x_i \in \left[\delta_d^U, x_n\right]} h(x_i) \cdot |y_i - f(\boldsymbol{\beta}, x_i)|^p + \sum_{j=0}^{j=r} \lambda_j \int_{x_1}^{\delta_d^L} h(v) \cdot \left| f^j(\boldsymbol{\beta}, v) \right|^{q_j} \, dv \\
+ \quad & \sum_{j=0}^{j=r} \lambda_j \int_{\delta_d^U}^{x_n} h(v) \cdot \left| f^j(\boldsymbol{\beta}, v) \right|^{q_j} \, dv
\end{aligned} \tag{B.31}$$

It is now easy to verify that the objective function above cannot be reduced further by adding any knot within the interval $[\delta_d^L, \delta_d^U]$. Indeed, the fitted function as above includes two piece-wise polynomial segments, defined over the intervals $[\delta_d^L, \hat{\delta}_d^{QP}]$ and $[\hat{\delta}_d^{QP}, \delta_d^U]$. The polynomial coefficients for the considered interval are tied to those defined for intervals left and right of the considered interval by means of continuity constraints. In the argument $\hat{\delta}_d^{QP}$ no continuity constraints are applied. Now consider that a knot is added in the interval $[\delta_{L,d}, \hat{\delta}_d^{QP}]$ in the argument α. In this case, the original left-side polynomial is split into two new piece-wise polynomials with continuity constraints for the function value and the derivatives up to the r^{th} derivative. The last (r^{th}) derivative is discontinuous. Another way to interpret this is that the original polynomial with $r+1$ coefficients is now replaced by two polynomials involving $2 \cdot (r+1)$ coefficients. This results in the net addition of a single degree of freedom. Importantly however, this added degree of freedom cannot be exploited to reduce the objective function value.

To see this, it should be noted that the coefficients of the polynomial terms up to degree $r-1$ over $[\alpha, \hat{\delta}_d^{QP}]$ can be computed from the coefficients of the polynomial over $[\theta_d^L, \alpha]$ thanks to existing continuity constraints. The coefficient for the polynomial term of r^{th} degree for the interval $[\alpha, \hat{\delta}_d^{QP}]$ remains to be chosen. Interestingly, this value can be chosen freely since the objective function is not influenced by the value of this coefficient. Any further addition of a knot in the same or other location will lead to the same effect. In summary, the further addition of any number of knots within the interval $[\delta_d^L, \hat{\delta}_d^{QP}]$ adds degrees of freedom to the fitted function which cannot be used to improve the objective function because it is not sensitive to the added parameters. Furthermore, the set of applied constraints does not change when adding such additional knots. Similarly, the addition of any number of knots within the interval $[\hat{\delta}_d^{QP}, \delta_d^U]$ cannot be used to reduce the computed objective function value. It follows from the above that the addition of any number of knots with multiplicity $r+1$ in the interval $[\hat{\delta}_d^L, \delta_d^U]$ will result in the same objective function with value g_3^L. This is true for all discontinuous transitions ($d = 1 \ldots n_d$).

The last paragraph proves that any value obtained for g_3^L for any feasible choice for $\boldsymbol{\theta}^{QP}$, g_3^L is equal to the lowest attainable value for this function for any number of added knots within the permitted intervals. In other words, the optimization problem cannot be relaxed further by adding any number of additional knots in the permitted intervals. Importantly, this lowest attainable value is also obtained by applying the (unknown) globally optimal values for $\boldsymbol{\theta}$, here referred to as $\boldsymbol{\theta}^*$. This is written mathematically as:

$$g_3^L = g_3^{L*} \leq g^{U*} \leq g^U \tag{B.32}$$

In words, this means that the lower bound objective function g_3^L evaluated for any feasible solution is equal to the same lower bound evaluated for the globally optimal solution, g_3^{L*}, which, in turn, is lower than or equal to the objective function for the global optimum of the original problem. This concludes the proof of the lower bound.