Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Qualitative trend analysis for process monitoring and supervision based on likelihood optimization: state-of-the-art and current limitations

Kris Villez *

**Spike, Process Engineering, Eawag, Überlandstrasse 133, P.O. Box 611, 8600 Dübendorf, Switzerland, (e-mail: kris.villez@eawag.ch)*

Abstract: In this study, two recently developed methods for qualitative trend analysis are applied and compared on the basis of two different data sets. One of the methods is globally optimal in the maximum likelihood sense but is computationally expensive. This method is based on shape constrained spline function. The second method is based on kernel regression and a Hidden Markov Model. This is more efficient but cannot be guaranteed to be optimal. Nevertheless, both methods deliver satisfying results with respect to the estimation of the location of inflection points as well as the corresponding tangent slopes. In contrast, only the globally optimal method appears useful to identify time series which do not satisfy a presupposed shape.

Keywords: Qualitative trends analysis, Process monitoring, Hidden Markov Model, Oxygen Uptake Rate, Membrane reactor operation

1. INTRODUCTION

Qualitative trend analysis (QTA) constitutes a relatively small area of the literature on signal processing in which so called qualitative characteristics of a signal are discovered and identified. Historically, the sign of the first and second derivatives have received most attention (e.g., Bakshi and Setphanopoulos, 1994, Rengaswamy and Venkatasubramanian, 1995, Dash et al., 2004, Maurya et al., 2005, Villez et al., 2013a). In this case, the aim is to identify segments of time series in which the signal is characterized as isotonic or antitonic and/or convex or concave. Such a characterisation results in a qualitative representation (QR) of the signal. Such a QR consists of episodes, i.e., time segments in which the sign of one or more derivatives does not change. Each of such episodes is uniquely characterized by (1) a start time, (2) an end time and (3) a primitive. The primitives correspond to a unique combination of signs for the signal derivatives. In this study, the sign of the signal value and the second derivative only are relevant. The relevant primitives are shown in Table 1.

Table 1. Primitives as used in this work according to the signs of the signal value and second derivatives. The marker "?" signifies that the sign of the signal is unknown or unspecified.

		Sign of the second derivative	
		-	+
Sign of the signal	-	N^-	P^-
	?	N	P
	+	N^+	P^+

The use of QRs as time series descriptors is motivated foremost by the ideas that (1) such characterization allows automated process supervision by incorporating expert knowledge, which is often expressed in rough and qualitative terms rather than precise and quantitative terms and (2) a qualitative description of a phenomenon can be extrapolated further than a detailed quantitative description of the same. The latter capacity can be extremely useful if one wants to identify rare events online for which limited repeated instances are available. Because of this feature of QRs, they have been used mostly for process fault diagnosis.

Unfortunately, simple differentiation of a signal does not allow such characterization because of noisy features in the time series. Whereas the identification of a QR appears trivial to the human eye, it is certainly not straightforward to construct reliable algorithms to do the same in an automated, computer-based fashion. Several attempts have been made to do so however. The currently available toolset includes methods based on wavelet analysis (Bakshi and Setphanopoulos, 1994, Villez et al., 2013a), neural networks (Rengaswamy and Venkatasubramanian, 1995), piece-wise polynomials (Dash et al., 2004, Maurya et al., 2004), spline functions (Villez et al., 2013b), and kernel regression (Villez and Rengaswamy, 2013). This is in part because several methods are based on (1) a locally optimal, greedy optimization method or (2) heuristics, making comparison or selection of methods on theoretical grounds practically impossible. A rather common feature of these methods is that they make use of a quantitative feature generation step, e.g., to compute wavelet coefficients, polynomial coefficients or neural network outputs, which are then further abstracted into a qualitative descriptors (see Fig. 1). As such, we refer to these methods as feature extraction based methods. Even though the feature extraction techniques themselves are often approximating the original signal optimally because of their universal approximating properties, the further abstraction into qualitative descriptors usually has no statistical basis at all. As such, statistical inference regarding the found QRs is found impossible. Furthermore, the results in Villez et al. (2013a) suggest that many of the methods may not work as well as suggested in the original work or that they require excessive efforts in tuning for a particular application.

The above observations have led to improved methods which may also permit statistical inference. In Villez et al. (2013b), a new method is proposed in which the start and end points of the episodes in a QR are optimized in a globally optimal fashion. In this case, a spline function is fitted optimally while being constrained to exhibit a particular shape. The method is hence referred to as the Shape Constrained Splines (SCS) method. This shape is identified as a sequence of primitives, called a qualitative sequence (QS). The start and end points of the episodes, a.k.a. transitions, are unknown a priori and optimized simultaneously with the coefficients of the spline function. Selection of the best QS is possible by enumerating all possible QS candidates and optimizing the corresponding transitions for each. Importantly, the underlying model can be simulated, meaning that the shape constrained function can be used to the generate estimates for the underlying noise-free signal. It is this generative model property (e.g., Bischop, 1999) which leads to a maximum likelihood estimation of the QR, conditional to a number of QSs, at the cost of possibly lengthy diagnosis procedures as it is based on the branch-and-bound algorithm. A sketch of the method, stressing generative properties, is given in Fig. 1.

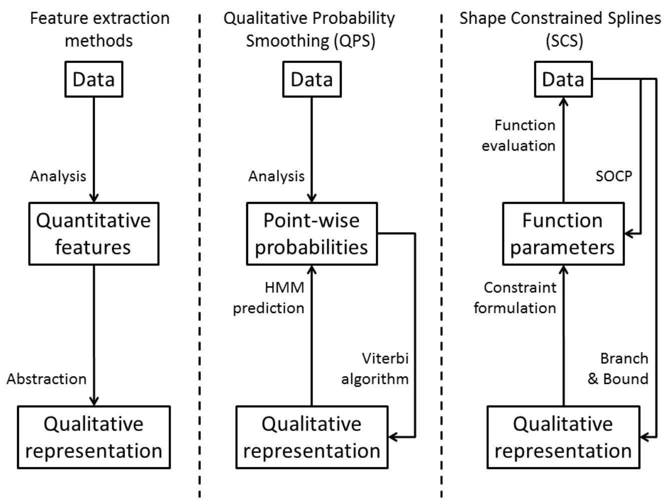

Fig. 1. Conceptual comparison of historical feature-based methods and newly developed Qualitative Probability Smoothing (QPS) and Shape Constrained Splines (SCS) methods for qualitative trends analysis (QTA).

To counter the observed computational complexity, a faster method has been developed in Villez and Rengaswamy (2013). In this case, a compromise is found between the feature extraction methods and the desire for an optimized likelihood. Indeed, point-wise probabilities for the shape of the underlying noise-free signals are derived based on kernel regression. These point-wise probabilities are then smoothed by means of a Hidden Markov Model and the Viterbi path finding algorithm. This method is further referred to as the Qualitative Probability Smoothing (QPS) method. In this case, the method exhibits the generative property to the extent that the point-wise probabilities can be generated by a Hidden Markov Model (HMM). As a result a globally defined objective function, i.e. defined over the whole time series, is still available for optimization. However, the QPS method is expected to suffer from (1) the approximation of the point-wise probabilities and (2) ignoring of correlation and auto-correlation within the Viterbi algorithm. The simulation study in Villez and Rengaswamy (2013) suggests that the computational benefits may outweigh the unguaranteed optimality of the method.

In this contribution, we apply both the QPS and SCS method, in this order, to two real-life data sets. It is the first time these methods are compared directly. The two case studies are first explained in the next section, after which the two existing methods are explained in a nutshell. This is followed by the results section. Further interpretation to the results is given in the Discussion and Conclusions sections.

2. CASE STUDIES

2.1 Case study 1

The first case study concerns the oxygen measurement profiles of a Sequencing Batch Reactor (SBR) process. In the aerobic stage of this 6-hour cycle process, a bang-bang controller ensures that the oxygen concentration remains within preset upper and lower limits. The oxygen measurement is available at a 1-minute time interval. The resulting profiles exhibit alternating upward and downward trends corresponding to aerated and non-aerated phases. The downward trends are used to estimate the Oxygen Uptake Rate (OUR). Initial attempts have been aimed at implementing a local linear regression approach as advocated in standard text books (e.g., Eaton et al., 2005). Unfortunately, several practical issues arise with such method. First, the assumption that a sufficiently long linear segment of data exists within the downward trend is not always met. Secondly, even if such a linear segment exists, then the identification of the linear segment is affected by subjective choices such as the criterion for linearity and the linear segment identification procedure, which is usually an ad hoc and suboptimal procedure. In this contribution, we discard the notion of linearity and instead identify the tangent slope at the downward inflection point as the oxygen uptake rate. It is assumed that this approach minimizes the effects of auto-correlated and delayed actuator and sensor responses. It is noted that this assumption remains to be validated.

The shape of the oxygen profile over each aeration cycle consisting of an aeration “on” and “off” phase is constrained to be $P^+N^+P^+$, i.e. positive and convex-concave-convex (see Table 1). This leads to a number of advantages such as (1) the use of all available data rather than a subjective and suboptimal selection and (2) implicit accounting for nonlinear effects on the oxygen time series such as (a) the effect of remaining air bubbles when the aeration is shut off and (b) the response time of the oxygen sensor, both of which severely reduce the time length of the linear segment –if even present- in the first aeration on/off cycles. The initial results as obtained for 11 cycles are presented in this abstract.

2.2 Case study 2

The second study concerns the full-scale side stream Membrane Bio-Reactor (MBR) "De drie Ambachten" (Terneuzen, The Netherlands). The analysed data consist of the Trans Membrane Pressure (TMP) measurements recorded at one-second time intervals. A single filtration cycle lasts 455 seconds. In each cycle, the TMP typically drops first because of backwashing, then increases fast when the normal flow is reinstated. This increase typically continues until the end of the cycle but the slope reduces, leading to a concave profile. As such, an NPN sequence (see Table 1) is typical. The locations of the inflection points implied by such shape are identified as well as the tangent slopes in these points. Whereas case study 1 is aimed at robust quantification of the OUR, this case study is set up as a data mining exercise.

3. METHODS

The recently developed generative methods for Qualitative Trend Analysis (QTA) were both used in this work to obtain (1) the location of inflection points, (2) the corresponding tangent slopes, and (3) measures of goodness of fit. The firstly applied method is the one presented in Villez and Rengaswamy (2013). This method is faster than the second because its objective function can be optimized by means of a dynamic programming method, which computational demand is linear in the number of data points. The second method is the SCS method (Villez et al., 2013b) which is more complex in terms of computational load, due to the fact that the objective function is nonlinear in the optimized parameters. However, it is currently hypothesized that the results of this method of more reliable in general. In addition, the objective function can be interpreted as a likelihood function for the data, which is not the case for the first method.

3.1 Method 1 – Qualitative Probability Smoothing (QPS)

This method consists of a two-step procedure. The first one is a feature generating step and aims to provide point-wise probabilities for the shape of the underlying data-generating function. This is done by means of kernel regression. The tri-cube kernel was used with a window support of 7 data points, which is considered relatively narrow. The point-wise probabilities for the shapes are then smoothed by means of the Viterbi algorithm (Forney, 1973). This algorithm is a specific instance of dynamic programming to find the maximum likelihood sequence of discrete states as modelled by an HMM (Rabiner, 1989). Indeed, the method is based on the association between HMM states with each of the primitives in a QS and a transition probability between each ordered pair of such states. As soon as the optimal sequence is found, the method is used to report (1) the transition times, i.e. the points in time at which a transition from one state to another is most likely, (2) the smoothed function value (for maxima and minima) or its derivatives (in inflection points), and (3) the optimized objective function value. Although this method is based on well-known principles and methods, the optimized objective function may not reflect the true objective function because (a) the correlation between derivatives of different orders is ignored and (b) the analysis of point-wise probabilities in an additive manner ignores the fact that the computed derivatives can exhibit strong auto-correlation. A strong advantage is that the computational time for both kernel regression and the Viterbi algorithm is linear in the number of data points.

3.2 Method 2 – Shape Constrained Splines (SCS)

The second method in this work is based on shape constrained spline functions (Papp, 2011). Such shape constraints allow controlling the signs of any derivative in any argument interval of a spline functions to be positive, negative or zero. By extension, it allows to fit a function to a time series which satisfies the shape as defined by a predetermined QR. This means that the transition times are fixed a priori in this case. Fortunately, this shape constrained spline fitting method has been adapted to permit the automated identification of the transition times given a predetermined QS. A drawback of the method is that the objective function is non-convex in the transition times and requires a branch-and-bound algorithm for deterministic global optimization (Villez et al., 2013b). This leads to a lengthy optimization procedure which may prevent online application. The transition times as found by the QPS method are used as initial guesses for contraction as in Faria and Bagajewicz (2011) so to reduce the initial search space and, thereby, the computational load. At this point in time, the effectiveness of this procedure has not been evaluated yet.

4. RESULTS

4.1 Case study 1

Fig. 2 shows the oxygen measurements in the first alternating cycles between aerated and non-aerated phases. In addition, the fitted shape constrained splines as identified by the SCS method are plotted. As one can see, the obtained fit is very good and the identified inflection point locations make sense.

In the top panel of Fig. 3, the oxygen measurements of the whole first SBR cycle are shown as well as the locations of the identified inflection points with the SCS method. The bottom panel shows the (absolute) slopes of the tangent in the identified inflection point for both methods and for both the rise (aeration) and fall (no aeration) phases. While the slopes of the falling trends match each other very well, it appears that the rising slope estimates are consistently lower for the QPS method. This may indicate that the kernel support window, even if it is narrow already, may be too high. The slopes of the falling trends are local estimates of the OUR and match typical expectations in the sense that the first estimates show a relatively high OUR which decrease relatively quickly at sample 2650, after which the OUR levels off. This transition is generally known as the switch from exogenous to endogenous respiration conditions.

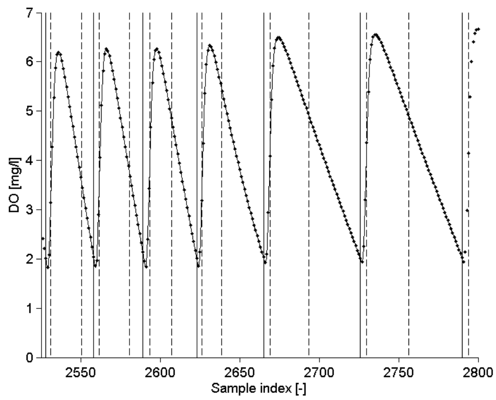

Fig. 2. First aeration cycles of the first SBR cycle. Dots represent the data. Full lines indicate the start and shutdown of the aeration. Dashed lines indicate the identified inflection points.

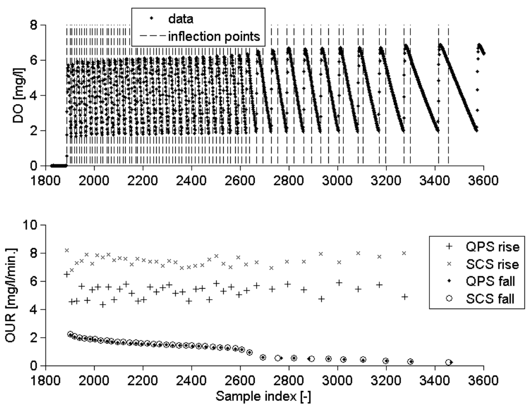

Fig. 3. First full SBR cycle. Top – Data and identified location of inflection points based on the SCS method. Bottom – Tangent slopes as function of identified inflection point location for both methods.

The shape identification procedure was repeated over all 11 cycles. The QPS method optimizes a path likelihood, shown in the top panel of Fig. 4, whereas the SCS method optimizes an SSR which is shown in the bottom panel. In case of the latter, a statistical limit was computed based on the assumption that SSR is distributed as a χ^2 distribution. This is based on an estimate of measurement noise variance and the effective degrees of freedom of a smoothing spline function (Hastie et al., 2009). Because the bang-bang controller adapts the aerated and non-aerated cycles on-line, the length of each analysed time series segment is not constant. Because of this, the degrees of freedom change for the χ^2 distribution and, therefore, its statistical limit changes as well. Note that this computed limit does not take the shape constraints into account and is therefore only approximately correct. Nevertheless, this approach results in a clearly anomalous aeration cycle around sample 18000 (Fig. 4), as the SSR is much higher than the computed limit. Visual inspection of the data (not shown) indicated that this is because the measurement profile does not conform to the expected $P^+N^+P^+$ profile, due to handling of the sensor during cleaning. All other data appears normal (not shown), in line with the SSR based statistic in Fig. 4. The likelihood function as optimized with the first method (Fig. 4, top panel) does not permit such clear distinction. Further notes on this are made in the discussion section that follows.

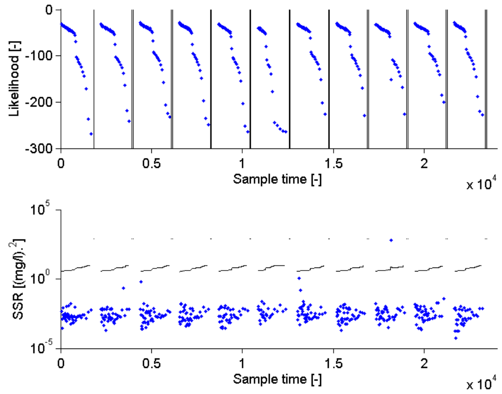

Fig. 4. All SBR cycles. Top – Likelihood obtained with the Qualitative Probability Smoothing (QPS) method. Bottom – Sum of Squared Residuals (SSR) and statistical limit as obtained with the Shape Constrained Splines (SCS) method.

4.2 Case study 2

The inflection points as identified for the NPN shape constrained profiles for both methods are shown against each other in Fig. 5. As can be seen, the methods do deliver similar but not the exact same results. Interestingly, there seems to be a very good correlation between the two methods.

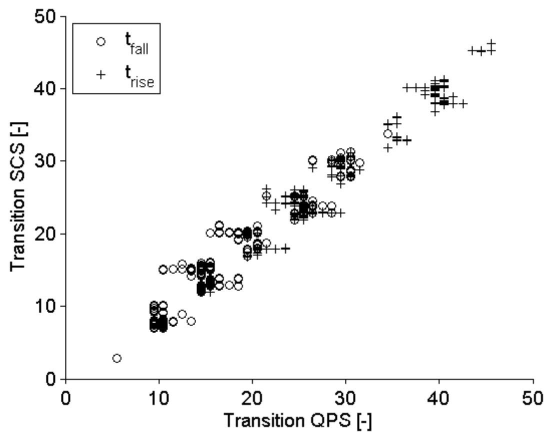

Fig. 5. Scatterplot of the identified time location of inflection points for both the falling and rising trend in the TMP profiles for both methods. The locations identified by both methods resemble each other very well.

The corresponding slopes of the tangents are shown in Fig. 6. Here, substantial differences can be seen as the QPS method generally delivers a lower estimate of the absolute slope than the SCS method. Also in this case, the kernel support window may have been too large for the QPS method. Despite this difference, both methods result in similar trends over the range of cycles, indicating that both are useful to find changes in the tangent slopes. In this case, a slow increase in the (absolute) slopes can be seen until about cycle 800 after which the slopes return to their initial levels. It is unclear at this time what the significance of this change is.

Fig. 7 shows the SSR values as a function of the cycle index. Here, the statistical limit is constant as the cycle length is constant. While none of the statistic values rise above this limit, the large distance between the limit and the obtained values suggests that the limit may be too conservative, i.e., it is possible that anomalous cycles are not detected. This is believed to be due to the fact that the effects of the shape constraints on the degrees of freedom for the spline function fitting are not accounted for.

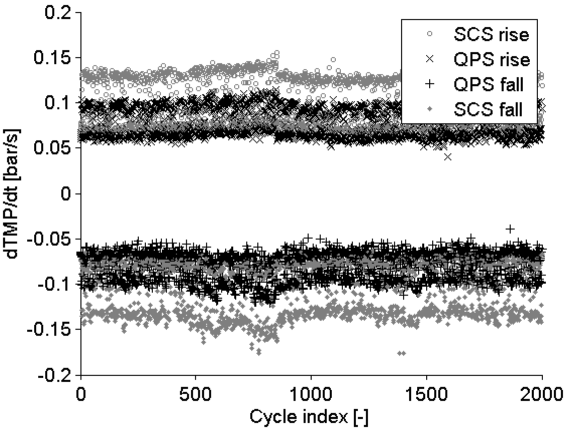

Fig. 6. The identified slopes of the tangents in each of the falling and rising inflection points as a function of cycle index. The QPS method generally results in a lower absolute value for the slope than the SCS method.

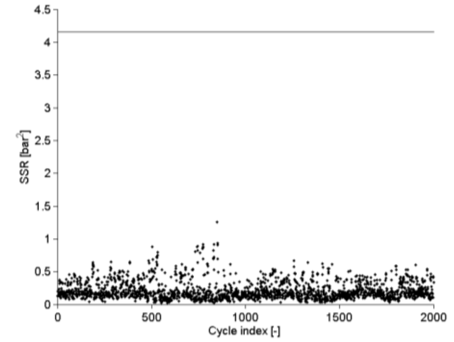

Fig. 7. The SSR as a function of cycle index.

5. DISCUSSION

Two recently developed methods for qualitative trend analysis were tested on two real-life case studies. In both case studies, a predetermined shape was imposed onto time series of a cyclic nature. Each time, two methods were used to obtain estimates of the time location of inflection points, the corresponding tangents and the optimized objective function. The methods delivered comparable results with respect to the time location of inflection points and tangent slopes. This shows that the first method, although only approximately optimal, allows a reliable analysis despite its computational efficiency. However, the presence of anomalous data, which affects the tangent and inflection point estimation, cannot be detected reliably on the basis of the optimized objective function. The current hypothesis is that the changing time series lengths affect the computed likelihood so strongly that any other effect, e.g., due to mismatch of imposed and true shape, is dominated by it. This remains to be investigated.

In contrast to the historical feature extraction based methods, the recently developed methods offer the advantage of a clearly defined objective function defined for the whole data series. The optimization of a globally defined objective function appears to lead to reliable results in terms of inflection point identification as well as tangent slopes, even though the objective functions have a different basis. Still, it should be noted that the applied methods require the shape to be known as a qualitative sequence whereas the original feature extraction methods do not. A unique explanation for the reliable performance is therefore not established yet. Further research will be directed at finding methods that identify both the qualitative sequence and the transitions defining the qualitative representation, instead of only the transitions. It is hypothesized that better performance can also be expected from such generative methods compared to the historical feature extraction methods.

Despite the effort in finding globally optimal methods for qualitative trend analysis, a number of challenges remain to be addressed. For instance, the methods have been applied in a maximum likelihood setting by (1) using an ad hoc defined prior for alternative shape constraints depending on subjective beliefs and the problem setting (Villez et al., 2012, Villez et al., 2013a). In addition, (2) method 1 does not define a likelihood for the data directly but rather on derived features for which the distribution is assumed independent which is only approximately true. Furthermore, (3) effects of the shape constraints on the prior distribution of the spline function coefficients have been ignored when using method 2. The fact that using a prior distribution in a Bayesian setting is possible is a fortunate side-effect of focusing on generative methods. However, for a proper Bayesian analysis it is required that (a) better conditional likelihoods can be defined for method 1, (b) reasonable prior likelihoods can be defined for the coefficients shape constrained spline functions, and (c) effective sampling strategies become available for second order cone constrained function fitting problems. To address (a), the approach taken in Gorinevsky (2008) for smoothing of an isotonic process by means of Moving Horizon Estimation can possibly be extended to allow for shape constraints in multiple derivatives as well as multiple, sequential episodes with different shapes. In a number of special cases of method 2, one can reduce the original second order cone programming problem reduces to a linearly constrained quadratic problem for fixed locations of the transitions. A 4^{th} order (cubic) spline function fitting problem with a quadratic objective function and constraints on the second derivative only is an example. In such case, prior likelihoods can be adopted as in the work of Romeijn and Van De Schoot (2008), which is specifically oriented at finding proper information criteria for model selection when the hypothesized alternative models define a parameter set rather than a specific parameter value. Combined with the sampling strategy of Rodriguez-Yam et al. (2004), a full Bayesian approach may within reach. On the other hand, it is unclear how one could sample from the shape constrained spline coefficient distribution in the more general case involving second order cone constraints (or, by extension, semi-definite cone constraints). This appears especially challenging because the question whether a given spline coefficient vector belongs to the feasible set can –so far- only be answered by solving an optimization problem. This means that Monte Carlo sampling strategies based on such evaluation suffer from (1) the time needed for optimization and (2) the lack of structural information about the problem which can be used to optimize sampling strategies.

Most of the fault diagnosis applications of the qualitative trend analysis deal with time series data. As such, it is not surprising that a number of methods are based on time series analysis methods such as wavelet analysis (Bakshi and Stephanopoulos, 1994, Villez et al., 2013a). In contrast, the generative methods as well as others are based on static models which do not explicitly adopt the notion of causality as is done in classic methods for state estimation and model identification. It may be of benefit to account for this as it may be considered implausible that the underlying noise-free time series truly behaves in a piece-wise polynomial fashion. It is unclear at this time whether a linear time-invariant causal model could lend itself to a constrained state estimation formulation of QTA, possibly inspired by Gorinevsky (2008). Importantly, such possibility would lend itself to the online application of the resulting methods, which has not been attempted yet with the newly developed generative methods.

5. CONCLUSIONS

Two methods for Qualitative Trend Analysis have been applied to two real-life data sets. In both case, both the suboptimal QPS method and the optimal SCS method deliver similar results for the identified qualitative representation of the data. This suggests that the approximation error of the QPS method may be minimal. In contrast, a good lack-of-fit statistic has not been obtained yet for the QPS method. The obtained results further support the general applicability of QTA methods for computer-based tasks unrelated to fault diagnosis. Despite the promise of these methods, several challenges remain open, including the need for a Bayesian inference framework as well as the adoption and extension of the methods for on-line applications.

ACKNOWLEDGMENTS

The author wants to thank Dr. Jan-Willem Mulder at Evides (Terneuzen, The Netherlands) and collaborators at Biomath, Ghent University (Ghent, Belgium) for sharing the MBR data and Ing. Jonathan M. Habermacher and Dr. Nicolas Derlon at Eawag (Zürich, Switzerland) for sharing the SBR cycle data.

REFERENCES

Bakshi, B. R., & Stephanopoulos, G. (1994). Representation of process trends—III. Multiscale extraction of trends from process data. Computers & chemical engineering, 18(4), 267-302.

Bishop, C. M. (1999). Bayesian PCA. Advances in neural information processing systems, 382-388.

Dash, S., Maurya, M. R., Venkatasubramanian, V., & Rengaswamy, R. (2004). A novel interval-halving framework for automated identification of process trends. AIChE journal, 50(1), 149-162.

Faria, D. C., & Bagajewicz, M. J. (2011). Novel bound contraction procedure for global optimization of bilinear MINLP problems with applications to water management problems. Computers & chemical engineering, 35(3), 446-455.

Forney Jr, G. D. (1973). The viterbi algorithm. Proceedings of the IEEE, 61(3), 268-278.

Eaton, A. D., Clesceri, L. S., Greenberg, A. E., & Franson, M. A. H. (1995). Standard methods for the examination of water and wastewater. APHA

Gorinevsky, D. (2008, June). Efficient filtering using monotonic walk model. In American Control Conference, 2008 (pp. 2816-2821). IEEE.Romeijn, Jan-Willem, and Rens Van De Schoot. "A philosophical analysis of bayesian model selection for inequality constrained models." Null, alternative and informative hypotheses, Springer, New York (2008): 329-357.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., & Tibshirani, R. (2009). The elements of statistical learning (Vol. 2, No. 1). New York: Springer.

Maurya, M. R., Rengaswamy, R., & Venkatasubramanian, V. (2005). Fault diagnosis by qualitative trend analysis of the principal components. Chemical Engineering Research and Design, 83(9), 1122-1132.

Papp, D. (2011). Optimization models for shape-constrained function estimation problems involving nonnegative polynomials and their restrictions (Ph.D. thesis). Rutgers University.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286.

Rengaswamy, R., & Venkatasubramanian, V. (1995). A syntactic pattern-recognition approach for process monitoring and fault diagnosis. Engineering Applications of Artificial Intelligence, 8(1), 35-51.

Rodriguez-Yam, G., Davis, R. A., & Scharf, L. L. (2004). Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression. Unpublished Manuscript.

Villez, K., Rieger, L., Keser, B., Venkatasubramanian, V. (2012). Probabilistic qualitative analysis for fault detection and identification of an on-line phosphate analyzer. International Journal of Advances in Engineering Sciences and Applied Mathematics, 4, 67-77.

Villez, K., Rosén, C., Anctil, F., Duchesne, C., Vanrolleghem, P. A. (2013a). Qualitative Representation of Trends (QRT): Extended method for identification of consecutive inflection points. Computers and Chemical Engineering, 48, 187-199.

Villez, K., Venkatasubramanian, V., Rengaswamy, R. (2013b). Generalized Shape Constrained Spline Fitting for Qualitative Analysis of Trends. Computers & Chemical Engineering, 58, 116-134.

Villez, K., & Rengaswamy, R. (2013). A generative approach to qualitative trend analysis for batch process fault diagnosis. In Proceedings of the European Control Conference (pp. 1958-1963).