Qualitative representation of trends: an alternative approach to process diagnosis and control

K. Villez, C. Rosén, F. Anctil, C. Duchesne and P. A. Vanrolleghem

ABSTRACT

The potential for qualitative representation of trends in the context of process diagnosis and control is evaluated in this paper. The technique for qualitative description of the data series is relatively new to the field of process monitoring and diagnosis and is based on the cubic spline wavelet decomposition of the data. It is shown that the assessed qualitative description of trends can be coupled easily with existing process knowledge and does not demand the user to understand the underlying technique in detail, in contrast to, for instance, multivariate techniques in Statistical Process Control. The assessed links can be integrated straightforwardly into the framework of supervisory control systems by means of look-up tables, expert systems or case-based reasoning frameworks. This in turn allows the design of a supervisory control system leading to fully automated control actions. The technique is illustrated by an application to a pilot-scale SBR.

Key words | B-splines, nutrient removal, qualitative representation of trends (QRT), SBR, wavelet analysis

K. Villez
P. A. Vanrolleghem
BIOMATH, Ghent University,
Coupure Links 653, B-9000 Gent,
Belgium
E-mail: *Kris.Villez@biomath.ugent.be*;
Peter.Vanrolleghem@gci.ulaval.ca

C. Rosén
IEA, Lund University, LTH,
Box 118, SE-221 00, Lund,
Sweden
E-mail: *Christian.Rosen@iea.lth.se*

F. Anctil
P. A. Vanrolleghem
gEAU: Groupe de recherche en génie des eaux,
Dépt. génie civil, Pavillon Pouliot,
Université Laval,
Québec, QC, G1K 7P4
Canada
E-mail: *Francois.Anctil@gci.ulaval.ca*;
Peter.Vanrolleghem@gci.ulaval.ca

C. Duchesne
LOOP: Laboratoire d'observation et d'optimisation des procédés, Département de génie chimique, Pavillon Adrien-Pouliot,
Université Laval,
Québec G1K 7P4
Canada
E-mail: *Carl.Duchesne@gch.ulaval.ca*

INTRODUCTION

Wavelet analysis has become a popular tool in the past decade for joint analysis of characteristics of time series in both the time and frequency domains. Applications in the context of statistical process control relate to different topics such as fault detection (e.g. Luo *et al.* 1998; Rosén & Lennox 2001; Aradhye *et al.* 2003) and data reconciliation (e.g. Tona *et al.* 2005). For a complete overview of applications of wavelet-based methods for process monitoring we refer to Ganesan *et al.* (2004). This paper focuses on the application of a wavelet-based method for qualitative description of time series, originally devised by Bakshi & Stephanopoulos (1994) and improved by Villez *et al.* (2007). Alternatively to the method applied in this paper, Flehmig *et al.* (1998) and Akbaryan & Bishnoi (2000) present other wavelet-based methods for qualitative interpretation of data. Typical applications of these methods aim at process monitoring and diagnosis (Akbaryan & Bishnoi 2001; Rubio *et al.* 2004; Flehmig & Marquardt 2006) or process data mining (Stephanopoulos *et al.* 1997). Other ways to obtain qualitative descriptions of trends are based on piece-wise polynomial fitting (Dash *et al.* 2004; Charbonnier *et al.* 2005), PCA-based (principal component analysis) clustering (Wang & Li 1999) or neural networks (Rengaswamy & Venkatsubramanian 1995). The latter approach is applied in Maurya *et al.* (2005) to obtain qualitative presentations of principal scores. Qualitative representations of series are also used for model structure discrimination, as in Vanrolleghem & Van Daele (1994) and Schaich *et al.* (2001).

doi: 10.2166/wst.2008.141

The methods reviewed so far are inductive in nature. The deductive research field twin, typically referred to as Qualitative Physics (QP), is older. Among the early and most important works in the field of qualitative modelling and reasoning we count Forbus (1984) and Kuipers (1986). An extensive overview of the early research in this field is given by Bourseau *et al.* (1995). While the interpretation of wavelet spectra has become common for time series analysis and has been introduced into the context of process monitoring, diagnosis and control, the use of qualitative methods for such purposes in the context of waste water treatment processes has not been evaluated extensively. Therefore, the potential of the method to control an SBR for nutrient removal is evaluated in this paper. Given the ultimate aim of the project to create an integrated system for monitoring, diagnosis and control for the SBR system, attention is given to (1) the direct usefulness of the technique for monitoring and diagnosis and (2) the ability to couple the monitoring/diagnosis outcomes with a supervisory controller.

MATERIALS AND METHODS

Data

The data set used in this paper consists of 100 complete batches from a pilot-scale SBR setup, collected at the end of 2006 (Nov. 22–Dec. 20). The SBR under study has a working volume of 64 L and is fed with synthetic sewage resembling domestic wastewater (Insel *et al.* 2006). The length of one cycle is 6 hours, which consists of a 60 min. anaerobic phase, a 35 min. anaerobic phase, a 130 min. aerobic phase, an 80 min. anoxic phase, a 30 min. final aerobic phase and a 60 min. settling/draw phase. A total of 34 L influent is added in the course of one cycle, of which 25.5 L during the first 25 min. of the anaerobic phase and 8.5 L in the first 10 min. of the anoxic phase. This operation delivers a hydraulic residence time (HRT) of 12 hours, which is typical for nutrient removal plants. The on-line pH signal is sampled at a 2 second interval. In this study, the pH trajectories in the first aerobic phase are analysed. See Figure 1 for two typical trajectories.

Qualitative representation of trends

The method used in this paper aims at the presentation of a time series as a set of contiguous periods in which the series exhibits a constant sign of the first and second derivatives, called triangular episodes. A time series can then be presented by a sequence of characters, i.e. a *word*, resulting in a *dictionary* when sets of time series are analysed. The sequences often confirm mental models stated by human operators, which allows the addition of a *meaning* to each word. Presence of noise prevents that the qualitative presentation of a trend can be obtained by direct computation of (discrete) derivatives. The method therefore consists of three steps: (1) wavelet decomposition into band-pass signals, (2) qualitative trend identification at

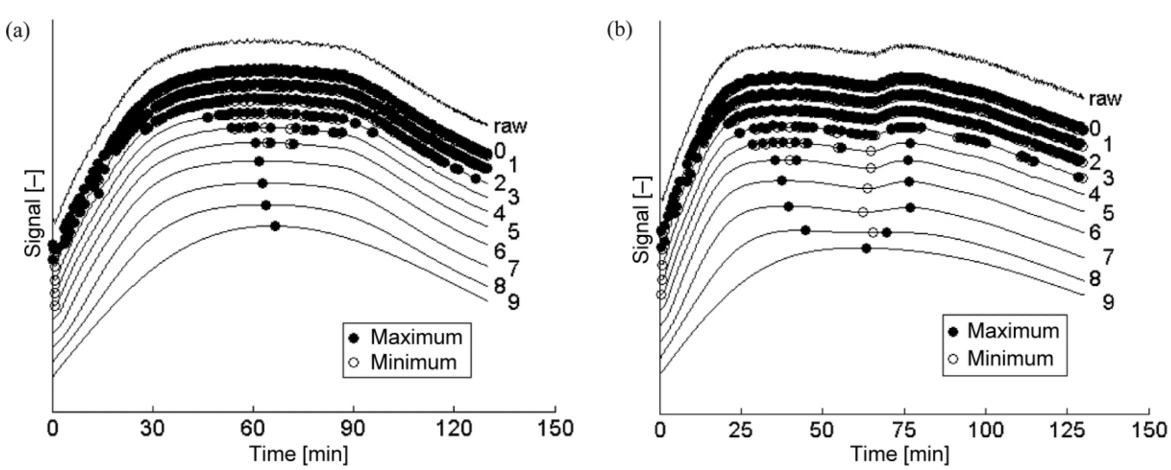

Figure 1 | Effect of quadratic spline filtering on apparent extrema. Original ('raw') and scaled signals (p = '0'…'9') are rescaled for reasons of clarity. (a) batch 5. (b) batch 7.

each wavelet scale (frequency band) and (3) assessment of relevant qualitative features. Each of these steps is described below.

Step1: wavelet decomposition. The continuous wavelet transform is used to decompose the signal. The (finite) time series, x_n, is convoluted with a daughter wavelet, $\psi(j,s)$, being a dilated version of the mother wavelet, ψ_0, for a given set of scales, s, and all time instants, k (1 to N):

$$W(k,s) = \sum_{j=0}^{N-1} x_j \cdot \psi(j,s)^* \left[\frac{(j-k)\cdot\delta t}{s} \right] \quad (1)$$

where * indicates the complex conjugate and δt the sampling period. For the purpose of qualitative representation of trends, the cubic spline wavelet is chosen (Bakshi & Stephanopoulos 1994). The method is based in the mother wavelet as defined by Mallat & Zhong (1992). The applicable daughter wavelet in Equation (1) is then obtained by scaling, translation and normalisation of the mother wavelet. Torrence & Compo (1998) deliver practical details and an efficient computation strategy of the described convolution in the Fourier domain. Of practical interest are the studied scales, noted as:

$$s_p = s_0 \cdot 2^{p \cdot \delta p}, \quad p = 0, 1, \ldots P \quad (2)$$

In our study, s_0, P and δp were set to 2, 9 and 1 respectively so that the studied scales ranged from 2 to 2^{11} times the measuring interval (i.e. periods between 4 seconds and approx. 68 minutes) with intervals of 1 octave. As the quadratic spline (band-pass) wavelet is the derivative of the cubic spline (low-pass) wavelet, the zero-crossings and extrema in the detail (band-passed) signals indicate the extrema and inflection points, respectively, in the corresponding scaled (low-passed) signals. At the same time, the quadratic spline wavelet filter does not add extrema or inflection points to an analysed signal (see Figure 1). As a result, the construction of a filter bank of cubic spline band pass filters and their corresponding quadratic spline low pass filters allows a straightforward assessment of the qualitative representation of the low-passed signals at each scale (see next paragraph). The latter property and the efficiency of the constructed filter bank lead to an accurate assessment of qualitative, semi-quantitative and quantitative features of signals at different frequency scales in a computationally efficient manner.

Step 2: triangular presentation at each scale. Qualitative representations of the low-passed signals are constructed at each scale. The identified extrema and inflection points now define the boundaries of the maximal time windows in the series in which the qualitative behaviour (sign of 1st and 2nd derivative) remains the same, called triangular episodes (Cheung & Stephanopoulos 1990). Seven types of qualitative behaviour are possible (see Table 1). Episodes with zero second derivatives (triangular primitives E, F and G) are not common in practice since filtering often distorts the form of linear parts of a signal in such a way that the second derivative in the filtered signal is non-zero. As a result, the resulting words are typically generated by a 2-letter alphabet (U/D) in case the first order behaviour is of interest only or a 4-letter alphabet (A/B/C/D) in case the first and second order behaviour are both of interest. In this study, only the first order behaviour was studied.

Step 3: assessment of relevant qualitative features. In Figure 1b one can observe that at scale 9 a single maximum results from (excessive) filtering while the representation at scale 8 is more appropriate. In order to assess the true final representation, the qualitative representations of a series are jointly analysed by application of Witkin's stability criterion. For a detailed explanation and example we refer to Bakshi & Stephanopoulus (1994).

RESULTS AND DISCUSSION

Qualitative representation of pH trajectories

A qualitative representation of the pH trajectories in the first aerobic phase of the system under study was obtained by means of the method described above. In Figure 2, each horizontal bar corresponds to the qualitative representation

Table 1 | Overview of primitives for characterisation of signals in terms of 1st and 2nd order behaviour (U: upward, D: downward)

Derivative:	1st	+	−	−	+	+	−	0
	2nd	+	−	−	+	0	0	0
Primitive:	Monotonic	U	D	D	U	U	D	G
	Triangular	A	B	C	D	E	F	G

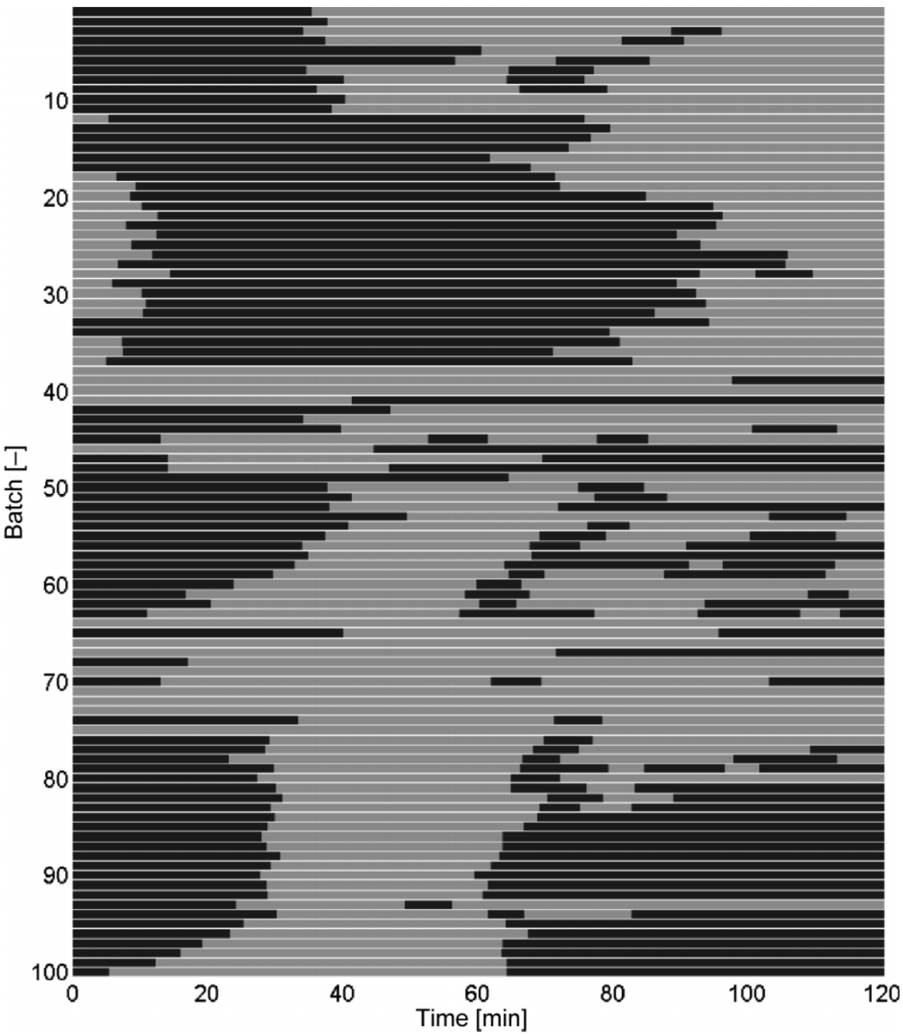

Figure 2 | Qualitative representations of 100 pH trajectories in the 130 minute aerobic phase. Dark shading indicates upward trends, light shading indicates downward trends.

of a single trajectory. Note that only the monotonic primitives (upward/downward) were used for this study. For example, the analysed trajectories of batches 13–17 are represented as UD (upward/downward) sequences. Batches 7–9 exhibit an UDUD sequence. By simple listing of the qualitative representations for all batches in the study, a so-called dictionary is automatically generated, in which a meaning is yet to be assigned to each entry. In this case, a 10-word dictionary results. In Table 2, the numbers of batches (cluster size) for each observed type of qualitative behaviour (cluster label) are given. Interestingly, the four most populated clusters (40% of the assessed behaviours) represent 70% of the batches. In addition, it is observed that the corresponding qualitative behaviours are relatively simple in nature (all sequences exhibit four characters at most). A major part of the batches thus corresponds to a limited set of relatively simple qualitative behaviours.

Diagnosis

Let us now try to provide diagnostic information to each of the clusters (i.e. adding a meaning to the entries in the dictionary). According to the operators, an UD presentation (e.g. batches 17–19) corresponds to a high load situation with an incomplete aerobic phase (incomplete nitrification, i.e. the pH did not stop decreasing). An UDUD presentation (e.g. batches 7–9) is related to a completed aerobic phase (complete nitrification) under high load conditions

Table 2 | Qualitative representations, corresponding diagnostics and occurrence

Pattern		Diagnostic information		Occurrence (%)
Observation	Dictionary	Load	Completion	
–	U	1 (high)	2 (incomplete)	–
UD	UD	1	2	16
UDU	UDU	1	1 (complete)	20
UDUD	UDUD	1	1	16
UDUDU		1	0 (unknown)	8
UDUDUD	UDUDU…	1	0	6
UDUDUDU		1	0	2
D	D	2	2	9
DU	DU	2	1	4
DUD	DUD	2	1	18
DUDUD	DUDU…	2	0	1

(i.e. after nitrification is completed the pH starts to rise because of CO_2 stripping). Interestingly, a high load is diagnosed by the operators if the pH trajectory starts with an upward trend (U), while downward trends (D) at the start of the aerobic phase were related to low load conditions. It is thus possible to diagnose the system under study on the basis of qualitative representations of the pH trajectories. In Table 2, the diagnosis given by the operators is given for each observed representation together with the frequency at which the observed representations occurred. This table functions as the dictionary of the qualitative representations of pH trajectories. As discussed above, an accurate diagnosis with respect to the load is possible for all observed behaviours. For some qualitative representations, no unambiguous diagnostics concerning the completion of the biological processes could be assessed. This was either due to the operator not being familiar with the observed pattern or due to the fact that different diagnoses (complete and incomplete) were possible within the set of batches with the same qualitative behaviour. Still, a complete diagnosis was possible for 83% of the batches.

Adding diagnostics for unobserved situations

The framework of qualitative representation of trends allows one to also consider *imaginative representations of trajectories* and the assessment of corresponding diagnostics and control actions, even if data of such sequences is not available. Such an injection of knowledge into this data-driven methodology reveals that the coupling of deductive and inductive methods is practically feasible. In this study, relatively simple sequences, such as 'U' and 'DUDU' were not observed within the one month of SBR monitoring (see Table 2). However, the operators of the studied system are able to complete the diagnosis dictionary with unencountered sequences without the necessity of factual observations of these. Table 2 gives the completed dictionary for the intended diagnosis and control system. Note that complex sequences (starting with 'UDUDU' or 'DUDU') were grouped into a single entry in the table so that an entry exists in the table for all possible patterns ever to appear.

The proposed methodology for diagnosis of a batch can be combined with the selection of an appropriate control action. This combination leads to a generic diagnosis and control system based on the assessment of qualitative representation of trends (Figure 3). The raw signal is processed to obtain the qualitative representation as proposed. The resulting qualitative representation is then looked up in the developed dictionary (as defined in Table 2) which relates the qualitative representation to a table in which the entries are defined as a possible combination of premises and a corresponding (set of) control action(s). While this appears feasible for simple control problems, more complex problems may require the use of rule-based systems or case-based reasoning. To start up and update any of these inferencing systems, operators can assess the relations between qualitative representations and diagnostics in a straightforward manner as qualitative presentations of trends are often concurring with their mental models in many cases. It is especially interesting that (1) operators do

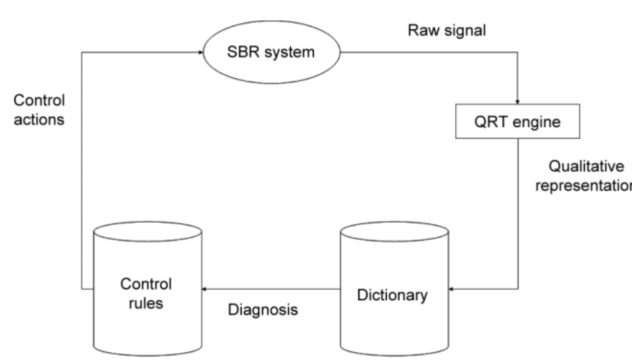

Figure 3 | Overview of the diagnosis and control system.

Table 3 | Diagnostics, associated control actions and occurrence

Diagnostic information			
Load	**Completion**	**Proposed automatic action**	**Occurrence (%)**
1 (high)	0 (unknown)	No change, call operator	16
1	1 (complete)	Reduce air supply	36
1	2 (incomplete)	Increase air supply	16
2 (low)	0/1	Increase load, equal air supply	23
2	2	Increase load, increase air supply	9

not need explicit knowledge on the mathematical details of the applied technique to make such a control system work and (2) the implementation of the technique does not require explicit process insight. This stands in contrast with the use of statistical models in process monitoring and diagnosis, e.g. PCA. Given the complex relation between the outcomes of such models (scores, statistics, cluster membership) and the original data, interpretation of the model outcomes and diagnostics is often difficult and requires a good understanding of the modelling technique and the process. Hence, qualitative presentations of trends offer a straightforward way to avoid such difficulties.

Control

Given the assessed diagnostics over the 1-month historical data set, the operators of the studied system were asked for an appropriate action to be taken in order to optimise operation in terms of effluent quality and plant economy while safeguarding acceptable operation. In Table 3, these are presented together with the % of batches for which they would be taken. As can be seen, the operators would increase the load to the system under low load conditions, regardless of the explicit assessment of completion (i.e. all ammonia is oxidised). This is not so surprising since none of the low load observed conditions corresponded to the diagnosis of an incomplete process. For the high load conditions a more refined set of actions was suggested by the operators. In case the biological processes are finished, operators would reduce the air supply in the next cycle (by reducing the aerobic phase length), while they would increase the air supply (by extending the aerobic phase length) if the biological processes are incomplete in the past aerobic phase. In case the load is high and no accurate assessment of the process completion is available, the operators suggested not changing the operation without further analysis. As a result, an automated adjustment of the load and aerobic phase length is possible for 84% of the batches.

Put otherwise, only 16% of the batches need to be diagnosed by means of a more detailed investigation. We note here that the location of the identified characteristics (e.g. extrema) in time were not included as criteria for diagnosis in this preliminary study.

From raw data to supervisory control: complete procedure

In Table 4 the complete procedure by which the supervisory controller can be established is given together with the expected interaction with the operator in each step. Quite interesting for implementation of such a supervisory controller is that no interaction is required in the complex step involving the qualitative representation of trends. In other words, the

Table 4 | From historical data to supervisory control: procedure

Step	Description	Operator interaction
1	Input of data	Optional
2	Generate qualitative representation of trends	No
3	Generate/update dictionary	No
4	Add diagnostics to (new) entries	Yes
5	Complete dictionary with unobserved entries	No
6	Add diagnostics to unobserved entries	Yes
7	Link control actions with diagnostics	Yes

end user does not need to understand the mathematical and computational aspects of the underlying technique.

CONCLUSIONS

In this paper, the usefulness of a technique for qualitative representations for diagnosis and control of an SBR for nutrient removal was evaluated. Even if the trends of only one variable (pH) and only one reaction phase of the SBR cycle were studied, it could be shown that for a major part of the batches an accurate diagnosis was possible on the basis of the presented methodology. Control actions were associated with all possible diagnoses. As every part of the running system can be automated, a closed-loop diagnosis and control system for the SBR plant under study is possible. A real-time implementation of such a system will be aimed for to fully validate the control performance. It is noted here that the proposed control loop is a preliminary result and may require further improvements on the basis of temporal information regarding the identified episodes (start time, end time, time length) in the inferencing steps, or, by addition of qualitative features in multiple sensor trajectories, may further improve the performance of the intended control system.

Importantly, the well-understood behaviour of the pH variable in aerobic conditions made straightforward development of the controller possible. Future studies may focus on or include sensor data and phases for which the understanding is less complete to evaluate whether qualitative representation of trends allows (1) the retrieval of new knowledge about the biological system and (2) the assessment of diagnosis and control strategies with minimal process knowledge available.

While the qualitative representation of trends is essentially an inductive method, the qualitative nature of its results can be coupled easily with deductive approaches to diagnosis (e.g. expert systems, case-based reasoning). Behaviours imagined by experts but not part of the data set may take part in the premises of certain rules in an expert system or may define artificial cases in a case-based reasoning system. The possibility to diagnose future faults that show little similarities with faults in historical data sets, and the straightforward link between the outcome of the technique and existing process knowledge, are considered major strengths in comparison with quantitative methods, e.g. PCA, which generally do not provide this opportunity. However, a formal treatment of possible interaction between the deductive and inductive framework of respectively Qualitative Physics (QP) and qualitative representation of trends is lacking in literature up to today.

ACKNOWLEDGEMENTS

This work was supported by the Institute for Encouragement of Innovation by means of Science and Technology in Flanders (IWT). Peter Vanrolleghem holds the Canada Research Chair in Water Quality Modelling.

REFERENCES

Akbaryan, F. & Bishnoi, P. R. 2000 Smooth representation of trends by a wavelet-based technique. *Comput. Chem. Eng.* **24**, 1913–1943.

Akbaryan, F. & Bishnoi, P. R. 2001 Fault diagnosis of multivariate systems using pattern recognition and multisensor data analysis technique. *Comput. Chem. Eng.* **25**, 1313–1339.

Aradhye, H. B., Bakshi, B. R., Strauss, R. A. & Davis, J. F. 2003 Multiscale SPC using wavelets: theoretical analysis and properties. *AIChE J.* **49**, 939–958.

Bakshi, B. R. & Stephanopoulos, G. 1994 Representation of process trends - part III. Multiscale extraction of trends from process data. *Comput. Chem. Eng.* **18**, 267–302.

Bourseau, P., Bousson, K., Dague, P., Dormoy, J. L., Evrard, J. M., Guerrin, F., Leyval, L., Lhomme, O., Lucas, B., Missier, A., Montmain, J., Piera, N., Rakoto Ravalontsalama, N., Steyer, J. P., Tomasena, M., Trave Massuyes, L., Vescovi, M., Xanthakis, S. & Yannou, B. 1995 Qualitative reasoning: a survey of techniques and applications. *AI Commun.* **8**, 119–192.

Charbonnier, S., Garcia-Beltan, C., Cadet, C. & Gentil, S. 2005 Trends extraction and analysis for complex system monitoring and decision support. *Eng. Appl. Artif. Intell.* **18**, 21–36.

Cheung, J. T.-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. 2004 A novel interval-halving framework for automated identification of process trends. *AIChE J.* **50**, 149–162.

Farge, M. 1992 Wavelet transforms and their applications to turbulence. *Annu. Rev. Fluid Mech.* **24**, 395–457.

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

Flehmig, F., Watzdorf, R. V. & Marquardt, W. 1998 Identification of trends in process measurements using the wavelet transform. *Comput. Chem. Eng.* **22**(suppl.), S491–S496.

Ganesan, R., Das, T. K. & Venkataraman, V. 2004 Wavelet-based multiscale statistical process monitoring: a literature review. *IIE Trans.* **36**, 787–806.

Insel, G., Sin, G., Lee, D. S. & Vanrolleghem, P. A. 2006 A calibration methodology and model-based systems analysis for SBR's removing nutrients under limited aeration conditions. *J. Chem. Technol. Biotechnol.* **81**(4), 679–687.

Kuipers, B. 1986 Qualitative simulation. *Artif. Intell.* **29**, 289–338.

Luo, R., Misra, M., Qin, S. J., Barton, R. & Himmelblau, D. M. 1998 Sensor fault detection via multiscale analysis and nonparametric statistical reference. *Ind. Eng. Chem. Res.* **37**, 1024–1032.

Mallat, S. & Zhong, S. 1992 Characterization of signals from multiscale edges. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 710–721.

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

Rengaswamy, R. & Venkatasubramanian, V. 1995 A syntactic pattern-recognition approach for process monitoring and fault diagnosis. *Eng. Appl. Artif. Intell.* **8**, 35–51.

Rosén, C. & Lennox, J. A. 2001 Multivariate and multiscale monitoring of wastewater treatment operation. *Water Res.* **35**(14), 3402–3410.

Rubio, M., Colomer, J., Ruiz, M. L., Colprim, J. & Mélendez, J. 2004 Qualitative trends for situations assessment in SBR wastewater treatment process. In: *Proceedings of the 4th ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence (BESAI)*, August 2004, Valencia, Spain.

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

Stephanopoulos, G., Locher, G., Duff, M. J., Kamimura, R. & Stephanopoulos, G. 1997 Fermentation database mining by pattern recognition. *Biotechnol. Bioeng.* **53**, 443–452.

Tona, R. V., Benqlilou, C., Espuña, A. & Puigjaner, L. 2005 Dynamic data reconciliation based on wavelet trend analysis. *Ind. Eng. Chem. Res.* **44**, 4324–4335.

Torrence, C. & Compo, G. P. 1998 A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78.

Vanrolleghem, P. & Van Daele, M. 1994 Optimal experimental design for structure characterization of biodegradation models: on-line implementation in a respirographic bionsensor. *Water Sci. Technol.* **30**(4), 243–253.

Villez, K., Rosén, C., Anctil, F., Duschesne, C. & Vanrolleghem, P. A. 2007 Qualitative representation of trends: an improved method for multiscale extraction of trends from process data. *In preparation*.

Wang, X. Z. & Li, R. F. 1999 Combining conceptual clustering and principal component analysis for state space based process monitoring. *Ind. Eng. Chem. Res.* **38**, 4345–4358.