Towards a common benchmark for long-term process control and monitoring performance evaluation

C. Rosen*, U. Jeppsson* and P.A. Vanrolleghem**

* Department of Industrial Electrical Engineering and Automation, Lund University, Box 118, SE-221 00 Lund, Sweden (E-mail: *christian.rosen@iea.lth.se, ulf.jeppsson@iea.lth.se*)

** BIOMATH, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure Links 653, B-9000 Gent, Belgium (E-mail: *peter.vanrolleghem@ugent.be*)

Abstract The COST/IWA benchmark simulation model has been available for seven years. Its primary purpose has been to create a platform for control strategy benchmarking of biological wastewater treatment processes. The fact that the benchmark has resulted in more than 100 publications, not only in Europe but also worldwide, demonstrates the interest for such a tool in the research community. In this paper, an extension of the benchmark simulation model no. 1 (BSM1) is proposed. It aims at facilitating evaluation of two closely related operational tasks: long-term control strategy performance and process monitoring performance. The motivation for the extension is that these two tasks typically act on longer time scales. The extension proposed here consists of 1) prolonging the evaluation period to one year (including influent files), 2) specifying time varying process parameters and 3) including sensor and actuator failures. The prolonged evaluation period is necessary to obtain a relevant and realistic assessment of the effects of such disturbances. Also, a prolonged evaluation period allows for a number of long-term control actions/handles that cannot be evaluated in a realistic fashion in the one week BSM1 evaluation period. In the paper, models for influent file design, parameter changes and sensor failures, initialization procedure and evaluation criteria are discussed. Important remaining topics, for which consensus is required, are identified. The potential of a long-term benchmark is illustrated with an example of process monitoring algorithm benchmarking.

Keywords Benchmark; fault detection and isolation; modelling; monitoring; simulation; wastewater treatment

Introduction

The use of a benchmark for assessment of process performance, control system evaluation, monitoring method efficiency, etc. is well established within chemical engineering and research. An example of this is the Kodak Tennessee Eastman Process (Downs and Vogel, 1993). The success of the COST/IWA benchmark (Copp, 2002; Spanjers *et al*., 1998) for control strategy development and evaluation clearly indicates the usefulness of such a tool for the wastewater research community. More than 100 reports on work related to the benchmark have of today been published.

The benchmark simulation model no. 1 (BSM1) definition, in short, consists of the model, control system, benchmarking procedure and evaluation criteria. The model is a five reactor activated sludge plant configuration with a (non-reactive) settler, utilizing the Activated Sludge Model no. 1 (ASM1) for modelling of the biological reactions and a ten-layer Takács model describing the settler. Parameter values and files characterizing the influent wastewater are also provided. The control system is user-defined but only specified control handles and sensors can be used by researchers to design control strategies. The benchmarking procedure includes implementation, initialization and evaluation of control system performance using a predefined one-week evaluation period. The evaluation is carried out according to a number of specified criteria (e.g. effluent quality, sludge production and

Water Science and Technology **Vol 50 No 11 pp 41–49** © IWA Publishing 2004

aeration energy). The definition of BSM1 has been and is an ongoing task and is not yet finalized. Remaining tasks in the definition are, for instance, an all-embracing criterion for evaluation of system performance (although there have been some proposals, e.g. Vanrolleghem and Gillot (2002)) and description of realistic sensor behaviour (including noise, time lags, etc.). These final issues have a potential to be resolved within a near future.

Although a very potent tool, the basic BSM1 does not allow for long-term evaluation. In the definition, the evaluation period is limited to one week. However, many of the control actions at a wastewater treatment plant have an effect on the process in longer time scales: sludge age control, equalization basin control and sludge storage to mention a few. Also, a short-term evaluation does not allow for realistic equipment (sensor/actuator) modelling, including failures, drift and maintenance, as these typically appear on longer time scales than one week.

Closely linked to process control is process monitoring. By process monitoring we mean tracking measurement variables to detect process deviations, failures and faults. Process monitoring also involves isolation of variables contributing to the deviations, facilitating further analysis of the problems. Monitoring of wastewater treatment operation has in the last decade become an intensive area of research and many different methods have been proposed (Rosen *et al*., 2003). Unfortunately, so far there has been no objective way to compare the success of methods for wastewater treatment monitoring, since researchers generally have used real data specific for a certain plant and, thus, not generally available to others. Although some researchers have started using the BSM1 for monitoring testing (e.g. Yoo *et al*., 2002), it is not a feasible alternative for benchmarking monitoring systems on a realistic set of disturbances and failures, typically observed at treatment plants.

In this paper, an extension of the BSM1 as discussed above is proposed. It is important to stress that this is a proposal and it does not provide a final definition. Instead it should be seen as a discussion starter within the research community, so that consensus can be sought for a long-term evaluation tool, i.e. the long-term benchmark simulation model no. 1 (BSM1_LT). In this paper, we will reflect upon some aspects of a future definition of the BSM1_LT model, benchmarking procedure and evaluation criteria.

BSM1_LT model definition

In this section, we indicate issues that need to be addressed and propose some possible directions for the final definition of the BSM1_LT. It is our opinion that it should be defined so that BSM1 is a subset of BSM1_LT, i.e. the BSM1_LT is an extension of BSM1 and at least one week of the longer evaluation period is consistent with the one of BSM1. Although a more thorough discussion will follow, we would like to start by emphasizing the differences between BSM1_LT and BSM1: 1) The evaluation period is extended significantly; 2) Temperature is added to the influent file; 3) The process parameters are time varying due to the fact that temperature is taken into account; 4) Sensor, actuator and process faults are included; 5) Control strategies with a long-term horizon are allowed; 6) The BSM1_LT has an additional application to process monitoring.

Evaluation time period

It is known that seasonal effects have a significant impact on the operation of a treatment plant. Also, typical failures, be it equipment or process failures, do not occur more than a few times per year. To include such phenomena in the assessment of a control or monitoring strategy, the required evaluation period is at least one year. It is realistic to assume that during a year of operation, even less probable faults and disturbances will occur and affect the operation significantly. The start and stop of the evaluation period is preferably set to summer, beginning of July and end of June, respectively, since the relatively favourable summer conditions would minimize the risk that effects of a proposed control strategy are pushed forward, beyond the end of the evaluation period.

Temperature trajectory

Since the evaluation period spans one year, temperature will be a determining factor. A common trajectory for the temperature over a year is more or less sinusoidal with its maximum value at the start of August and its minimum value at the start of February (in the southern hemisphere this is reversed). A more complex description of the temperature may be necessary to emulate the changes in temperature due to precipitation or snow melting. Such a description remains to be decided. As a starting point, however, we believe that a sinusoidal as described would suffice as an approximation of the temperature trajectory. The temperature is modelled as $T = 15 + 5 \cdot \cos(2\pi/365(t - 28))$°C, where t is the time in days and the shift is 28 days (4 weeks).

Temperature dependency

The values of the temperature dependent kinetic parameters in the ASM1 model are, consequently, varying during the evaluation period. In Henze *et al*. (1987), kinetic parameter values are given for 10 and 20°C, respectively, and intermediate values can be calculated according to an Arrhenius function. The BSM1 parameter values are defined at 15°C, but rounded to one or two decimals. Since it is desirable that BSM1 and BSM1_LT have exactly the same parameter values at 15°C, the Arrhenius function should be based on values at 10 and 15°C, using the BSM1 values for 15°C. At 20°C, this gives slightly different values than those of Henze *et al*. (1987).

It should be noted that the saturation concentration for dissolved oxygen is temperature dependent. This has an impact on the mass transfer rate of oxygen, since it is modeled as $K_La(S_{O,sat}–S_O)$. K_La is also temperature dependent. In BSM1, and also in the proposed BSM1_LT, the oxygen mass transfer rate is expressed as K_La. This means that no temperature compensation is required for K_La. However, as soon as K_La is to be expressed in terms of energy (an important evaluation criterion), the temperature dependency is crucial. Currently, the aeration-energy relationship is defined at 15°C only.

Influent file design

The design of the influent file is based on the existing influent files of BSM1. This means that the one-year file contains dry weather periods (exactly the same as in BSM1), rain events and storm events. However, the influent file needs to be modified to incorporate a temperature variable.

Rain events. A rain event is defined by an increased flow rate and pollutant flux, i.e. the mass transport of pollutants, which remains the same as during dry weather. It is assumed that the flow caused by the rain is added to dry weather flow but that it carries no pollutants. Thus, the additional flow will dilute the dry weather flow. The time of the day for the start of the rain is a random variable, with a uniform distribution, i.e. any time of the day is just as likely to have a rain event start. The same is true for the end time of the rain. The intensity of the rain may vary over time but the details remain to be specified.

Storm events. A storm event is defined by an increase in flow rate and an initially increased pollutant flux, to mimic sewer flush out. The intensity of the storm and the time since the last flush out determine the increase in pollutant flux. A storm event may be combined with a rain event to model flush phenomena in conjunction to rain. The changes in pollutant flux are defined in BSM1 and the same definition can be used here. Also, for the storm event the start and stop time should be random.

Sensor models

It is well known that sensors have dynamic properties, they are afflicted by noise, drift is commonly occurring and that sensors need to be calibrated and maintained in a reoccurring fashion. To emulate such sensor behaviour and test the control or monitoring scheme for its robustness and/or sensitivity to this is crucial. Realistic sensor model behaviour requires the dynamic properties and disturbance sources to be represented. This has been discussed in Rieger *et al*. (2003) and here we have adopted their approach to sensor modelling. This includes modelling of noise, time response, drift, signal saturation and, if not a continuous sensor, the measuring interval. Rieger *et al*. (2003) also include a procedure for modelling calibration and maintenance of the sensors. This is an important aspect for both control and monitoring. A control or monitoring system must handle periods of no or non-representative data during sensor calibration/maintenance.

Sensor and actuator failures

Sensor and actuator failures occur in all industrial environments. It is therefore important that the impact of failures on the behaviour of a control system as well as a monitoring algorithm can be evaluated. Sensor and actuator failures can be modelled in different ways, but this aspect is not included in the work of Rieger *et al*. (2003). A straightforward way is to assume that a failure occurs according to a Poisson distribution, with a specified failure rate. A failing sensor may result in an erroneous signal, no signal at all, change in noise level or excessive drift. An actuator failure may manifest itself as, for example, a complete or partial loss of capacity. The time for repair may of course vary. Rather than including all factors that may influence the time for repair, the duration of a malfunction has been assumed to be random as well. The time for repair can be described by a lognormal or exponential distribution with a minimum repair time. This results in a high probability for repair times close to the minimum repair time and a decreasing probability for longer repair times.

Process disturbances

Process disturbances are defined as disturbances acting only through the model parameters, e.g. maximum specific growth rates, settling parameters, etc. Thus, process disturbances are emulated using time varying process parameters. For instance, temporary nitrification inhibition can be imposed on the simulation model by reducing the maximum specific growth rate for autotrophs. Variation of the sludge settling properties due to influent changes, seasonal effects or sludge age alternation is another typical process disturbance. Varying the parameters of the Takács settling velocity function can simulate this behaviour. The particular type of process disturbances to be simulated is yet to be defined by the scientific community.

Additional control handles

The long-term perspective of the BSM1_LT allows for other control handles than those of the BSM1. Equalization tanks and sludge storage tanks have been already discussed within the benchmark community and could be implemented as a result of the proposed extension. Wastage flow rate is another control handle that, even though it is available within the BSM1 definition, does not have a stable effect within a week. However, an evaluation period of a year creates possibilities for, for instance, sludge retention time (SRT) control.

BSM1_LT benchmarking procedure

The benchmarking procedure includes implementation of the control or monitoring system, initialization and evaluation. We will in this section discuss what the procedures for long-term control and process-monitoring benchmarking could look like and describe how the

BSM1_LT can be used for objective evaluation of different control or monitoring strategies. But before this is addressed, there may be a need to clarify on the definitions of control and monitoring benchmarking.

Definition of process control and monitoring

The concept of process control is quite clear. Process control is the task of (generally) automatically controlling a process, utilizing measurements, so that one or several control objectives are reached, despite disturbances. As opposed to what is meant by the term process control, it is not always clear what process monitoring means. The term process monitoring, as used here, refers to the procedure of detecting abnormal process behaviour and to isolate the measurement variables most probable to contribute to the deviation. Process monitoring is, thus, a way to obtain information and knowledge of the current process state.

Implementation

The benchmark user carries out the implementation of a control or monitoring system. Issues such as choice of sensors, control handles, sampling time, etc. are resolved by the user utilizing what is available and allowed within the benchmark definition. An interesting issue is whether the data for plant influent, equipment failures, process disturbances, noise, etc. need to be determined so that all benchmark users implement exactly the same data or whether the user only specifies the statistical characteristics of the data. This has been a topic for some debate within the benchmark community. One possible solution is that one set of complete "standard" data is provided for ease of comparison between researchers. If other sets of data are used, it is imperative that the implementation is well documented to allow for repeatability. An alternative is to choose a statistical approach to the evaluation, which implies statistical analysis of numerous simulations. This approach is somewhat unrealistic today due to computational limitations.

Initialization

In the BSM1, the initialization consists of simulating the benchmark, with the control strategy that is to be evaluated, for 100 days with constant input, followed by three weeks of dry weather dynamic input data. This is done to obtain a quasi steady state and to allow for fair comparison between various strategies. The initialization period must be prolonged for the BSM1_LT. The reason for this is that control strategies acting in long time scales, need a long initialization phase. We believe that an initialization period of six months is a reasonable choice, as most control strategies would most likely settle and reach a quasi steady state within six months. Moreover, since many monitoring techniques are based on statistical data on which models are identified or algorithms trained, there is a need for historical process data. It is, thus, beneficial to include not only varying process conditions but also rain and storm events, faults and process disturbances in the initialization phase for the identification/training of representative monitoring models. Consequently, the initialization period ranges from beginning of January to end of June but must be different from the last 6 months of the evaluation period.

BSM1_LT evaluation criteria

To assess the performance of a control or monitoring strategy, evaluation criteria are necessary. These criteria aim at condensing the simulation output to a few indices or key variables that can be said to represent the system, controller or monitoring performance to allow for easy comparison of results. The criteria definition for control system performance is already part of the BSM1. These criteria will also be applicable to BSM1_LT. However, for the monitoring performance evaluation, new criteria must be defined. Some ideas for the development of criteria for monitoring system performance will be discussed in this section.

Control

In BSM1, the system performance is evaluated according to an effluent quality index (a weighted sum of effluent TSS, COD, BOD, TKN and nitrate), effluent violations and operational costs. Also, the controller performance is assessed in terms of controlled variable performance and manipulated variable performance (Copp, 2002). We can see no reason for changing this definition when using BSM1_LT for control benchmarking.

Monitoring

A difficulty when defining criteria for process monitoring performance evaluation is that it implies that all possible disturbances also need to be defined. Some are straightforward to define, such as equipment failures (sensors and actuators). Others are more complicated. For instance, is a high flow rate (e.g. a rain event) a disturbance or is it just an inevitable fact? In some situations, information on high flow rate may be crucial whereas in other situations it is considered to be normal. Thus, the criteria are also dependent on the aim of the monitoring procedure and the definition of "normal". This implies that assessment of the monitoring system performance can only address the disturbance types included in the user's definition of disturbances. In some sense, this corresponds to the controller assessment criteria, controlled variable performance and manipulated variable performance. When the disturbance types are defined, the monitoring performance can be evaluated in terms of detection and isolation performance.

Detection performance. When evaluating the monitoring performance from a detection perspective, at least four possible criteria can be set up: 1) The number of detections of true faults. This means that the monitoring system indicates a fault when a true fault occurs. The ratio between detections and the number of true faults should be maximized, i.e. close to one; 2) The faults that are not detected. The system indicates no detection when there is in fact a true fault. The ratio between missed faults and number of true faults should be minimized; 3) The false alarms. A fault is detected when there is no true fault. The ratio between false alarms and the total number of detections should be minimized; 4) The average time delay for detections. This is the time between an occurrence of an actual fault and its detection. Criteria 1 and 2 are complementary and only one of the two must be used.

Isolation performance. It is not trivial to objectively define isolation success, as it is dependent on the monitoring method. The definition of an objective isolation performance index is somewhat premature at this stage and is left as an open issue for further research.

The given criteria for monitoring performance form a relatively blunt tool for evaluation. However, they should be regarded as a starting point for the discussion within the scientific community.

Example of monitoring performance evaluation

This example illustrates how a future, full-fledged BSM1_LT could be used for benchmarking a monitoring system. Instead of focusing on monitoring theory and different algorithms we would like to emphasize the procedure for objective evaluation of the monitoring performance. In the example, we evaluate the impact of different types of data pre-treatment methods on the monitoring result. Thus, the same monitoring method is applied to the same original data set but with four different methods for data pre-treatment.

The model is simulated for a year (although only four months are shown here) with a varying temperature. The monitoring system samples every half hour (0.5 h^{-1}) and the resulting data contains the measurement signals listed in Table 1. In a realization of what has been discussed in this paper, a number of faults have been imposed on the system. Using the models for equipment failures previously discussed, the nitrate sensor in tank 2 fails twice during the shown evaluation period, at days 53–58 and 107–109. The return sludge flow pump fails once at day 32–33. In addition to this, four rain events are imposed (Figure 1). The aim of the monitoring model is to detect equipment failures as well as rain events. However, since one of the rain events is prolonged in time (46 days) we want the model to adapt to the high flow rate situation. Hence, a recursive PCA algorithm that adapts to the data is used (e.g. Rosen *et al.*, 2003). The updating criterion is that the model is only updated when the squared prediction error (SPE) is within its control limit.

Three different pre-treatment methods are benchmarked: 1) The first involves replacing the samples of the calibration events of the B0 type sensors (see Rieger *et al.*, 2003) with estimates. In the simulation, all B0 sensors were modelled with calibration events once a week to give the data a realistic behaviour. 2) The second method includes

Table 1 Available measurements. The sensor types refer to those of Rieger *et al.* (2003)

Measurement	Sensor type	Measurement	Sensor type
influent flow rate	ideal	control signal DO-control	ideal
influent NH_4 concentration	B0	SS concentration, reactor 5	ideal
NO_3 concentration, reactor 2	B0	effluent nitrate concentration	B0
control signal Q_{ret}-control	ideal	effluent NH_4 concentration	B0
DO concentration, reactor 5	ideal		

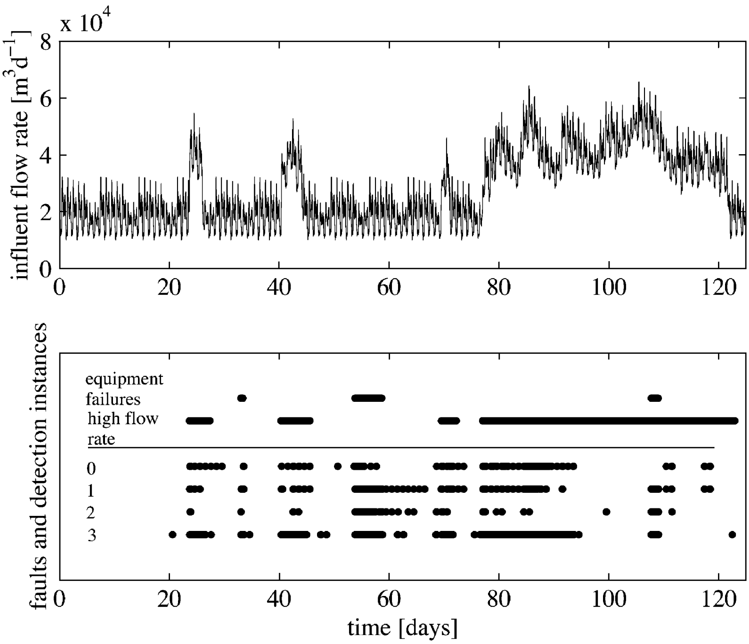

Figure 1 The influent flow rate during the first four months of a one-year evaluation period (top). The instances of the imposed faults (equipment failure and high flow rate) and detection for the four different data sets (bottom). 0: no data pre-treatment, 1–3: pre-treatment methods 1–3

Table 2 The resulting indices of the benchmarking of monitoring using different data pre-treatment methods

Treatment method	Correct detections/ true faults	Missed detections/ true faults	False alarms/ total detections	Average detection time (h)
0: no pre-treatment	5/7	2/7	13/18	1.6
1: calibration instances removed	7/7	0/7	13/20	1.5
2: pre-treatment 1 + calculated fluxes	7/7 (6/7)*	0/7 (1/7)*	9/16 (10/16)*	8.7 (1.5)*
3: pre-treatment 2 + deviations from dry weather diurnal pattern	7/7	0/7	9/16	1.1

* One detection was very late (during the third fault event). The values within brackets are the result if this instance is classified as a missed detection

the result of the first pre-treatment and, in addition, some fluxes, i.e. concentrations times flow rates, are calculated from the raw data. 3) The third method includes pre-treatments 1 and 2 plus the calculated signal deviation from the average dry weather diurnal pattern. The monitoring result using the three data pre-treatment methods as well as the no pre-treatment is shown in Figure 1 (bottom). The criterion for detection of a fault is that the SPE exceeds the control limit in two out of three consecutive samples.

A detection event is defined as one or several consecutive samples indicating faults. Furthermore, as long as there is a true fault, only the first event is counted and the rest are disregarded. Looking at Figure 1 (bottom), it is evident that the pre-treatment of data has an impact on the monitoring results. Qualitatively it is possible to state that the more advanced schemes result in fewer false alarms and more distinct detections. However, to be able to compare the results in a more quantitative fashion, the indices discussed before are calculated. The result is shown in Table 2. It is noteworthy that regardless of the pre-treatment, the algorithm adapts to the long rain event and in all but one case (0) the NO_3 sensor fault at days 107–109 is detected.

Concluding remarks

This paper is a proposal for a future definition of a long-term benchmark simulation model, based on the original COST/IWA benchmark simulation model. The new definition is efficient as it serves two purposes: long-term control and monitoring benchmarking. To do so, the benchmark definition should include a significantly longer evaluation period, temperature dependencies, realistic sensor and actuator models as well as equipment and process faults. Also, a new prolonged procedure for initialization is required. Criteria for control system evaluation are inherited from the original benchmark but a new definition of monitoring criteria must be included.

Computational aspects

It could be argued that the prolonged evaluation period is too long and that the dynamic simulation time would be cumbersome. However, using Moore's law, the computational speed during the last seven years has increased by a factor 25. This means that the definition of a one-year evaluation period requires less than twice the simulation time compared to the BSM1 defined in 1997. The same reasoning can be used to justify the future use of the statistical approach to evaluation, mentioned earlier, where a number of simulation runs are statistically evaluated.

What remains to be done?

Basically all the BSM1_LT components remain to be determined and defined. There are, however, a few topics that we think will need additional attention in order to reach consensus:

- Apart from the fact that the influent file in terms of rain and storm events remains to be defined, the implementation of equipment failures and process disturbances requires further work. Should they be strictly specified or should only their statistical properties be given? Should there be a "standard case" that allows for head-to-head comparison?
- The initialization period, here proposed to be six months, must be defined. Is six months an adequate choice? What should it contain in terms of faults and disturbances?
- An all-embracing criterion for control remains to be defined. However, this is also on the BSM1 agenda and the same criterion should be pursued. The monitoring criteria given here need further attention and especially how the quality of the isolation part of the monitoring task can be assessed, is a topic for further investigation.

A motivation for this work, not mentioned so far, is that the definition of a prolonged evaluation (and initialization) period is also necessary for the upcoming definition of the BSM2. The BSM2 is a "within fence" benchmark simulation model currently in preparation within the benchmark community. BSM2 includes all processes at a plant, particularly the sludge train.

References

Copp, J. (ed., 2002). *The COST Simulation Benchmark – Description and Simulator Manual*. Office for Official Publications of the European Communities, Luxembourg, ISBN 92-894-1658-0.

Downs, J.J. and Vogel, E.F. (1993). A plant-wide industrial-process control problem. *Comput. Chem. Eng.*, **17**(3), 245–255.

Henze, H., Grady, C.P.L. Jr., Gujer, W., Marais, G.v.R. and Matsuo, T. (1987). *Activated Sludge Model No. 1. IAWQ Scientific and Technical Report No. 1*, IAWQ, London, UK.

Rosen, C., Röttorp, J. and Jeppsson, U. (2003). Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation. *Wat. Sci. Tech.*, **47**(2), 171–179.

Rieger, L., Alex, J., Winkler, S., Boehler, M., Thomann, M. and Siegrist, H. (2003). Progress in sensor technology – progress in process control? Part I: Sensor property investigation and classification. *Wat. Sci. Tech.*, **47**(2), 103–112.

Spanjers, H., Vanrolleghem, P.A., Nguyen, K., Vanhooren, H. and Patry, G.G. (1998). Towards a simulation-benchmark for evaluating respirometry-based control strategies. *Wat. Sci. Tech.*, **37**(12), 219–226.

Vanrolleghem, P.A. and Gillot, S. (2002). Robustness and economic measures as control benchmark criteria. *Wat. Sci. Tech.*, **45**(4–5), 117–126.

Yoo, C.K., Choi, S.W. and Lee, I. (2002). Disturbance detection and isolation in the activated sludge process. *Wat. Sci. Tech.*, **45**(4–5), 217–226.