Demonstration of a tool for automatic learning and re-use of knowledge in the activated sludge process

J. Comas*,** , I. Rodríguez-Roda*, M. Poch*, K.V. Gernaey**, C. Rosen** and U. Jeppsson**

*Laboratori d'Enginyeria Química i Ambiental (LEQUiA), Universitat de Girona, Campus Montilivi s/n, E-17071 Girona, Catalonia, Spain (E-mail: *quim,ignasi,manel@lequia.udg.es*)

**Department of Industrial Electrical Engineering and Automation (IEA), Lund University, P.O. Box 118, SE-22100 Lund, Sweden (E-mail: *krist.gernaey,christian.rosen,ulf.jeppsson@iea.lth.se*)

Abstract Wastewater treatment plant operators encounter complex operational problems related to the activated sludge process and usually respond to these by applying their own intuition and by taking advantage of what they have learnt from past experiences of similar problems. However, previous process experiences are not easy to integrate in numerical control, and new tools must be developed to enable re-use of plant operating experience. The aim of this paper is to investigate the usefulness of a case-based reasoning (CBR) approach to apply learning and re-use of knowledge gained during past incidents to confront actual complex problems through the IWA/COST Benchmark protocol. A case study shows that the proposed CBR system achieves a significant improvement of the benchmark plant performance when facing a high-flow event disturbance.

Keywords Activated sludge; case-based reasoning; IWA/COST benchmark; learning; reuse; supervision

Introduction

Wastewater treatment is a complex process. Although progress in control engineering and process sensors has enabled significant automatic control improvements of wastewater treatment plants (WWTP), generic solutions for plant-wide control are still lacking. Performance of WWTP control systems is especially poor when facing problematic situations of biological origin, such as foaming or sludge bulking. This is partly caused by the limited number of reliable sensors able to provide real-time process information. However, the main reason for poor performance of plant-wide control systems in these situations is the lack of basic knowledge about the complex interactions between the microorganisms' communities in the WWTP and their reaction to disturbances related to the influent composition or plant operating conditions.

Problems of biological origin appear frequently, and plant operators must take appropriate actions to deal with them. In practice, the operators try to manage these situations by integrating different types of information (e.g. derived from on-line and off-line data, and heuristic information), and by reasoning over specialised WWTP operation manuals or heuristic knowledge. In addition, WWTP operators also apply their intuition and process experience when dealing with operational problems. Indeed, the problem at hand is usually solved by taking advantage of lessons learnt from successes and failures in plant operation that resulted as the consequence of operational decisions taken in the past when reacting to similar problems.

The registration of lessons learnt from past experience is essential from the process management point of view. It enables re-use of knowledge whenever similar problems arise and when new operators with less experience are in control of the process. However, past process experience is not easily integrated in numerical control systems (e.g. PI or PID controllers), and thus new tools must be developed. Case-based reasoning (CBR)

is a promising tool for this purpose. CBR systems are intelligent/knowledge-based systems able to store past experiences efficiently and to retrieve them automatically whenever similar problems arise. Recently, CBR systems have been developed and successfully applied in several weak-theory domains where previous experience provides a good opportunity to deal with current problems, e.g. medical applications, architecture and urban planning, process control and modelling of biological systems. CBR has also been proposed as a support tool in the wastewater treatment field (R.-Roda et al., 2001b; Wiese et al., 2003; Poch et al., 2004). Martínez et al. (2006) recently proposed a CBR approach to guarantee a successful performance as a support tool even when facing problems with slow dynamics, such as solids separation problems. In these situations, it is essential to consider a flexible CBR approach, with undefined temporal duration of cases according to the real dynamics of the activated sludge process.

A CBR system is a simple way to accumulate facts, knowledge and even troubleshooting methods for complex problems, providing learning capabilities to WWTP management. Therefore, the integration of such a tool within a control system allows an improvement in the management of recurring problems to be achieved since it enables combining knowledge and experience re-use with simulation models for similar events.

The purpose of this paper is to demonstrate the capabilities of a CBR system for automatic learning and re-use of knowledge in activated sludge process supervision. These benefits are numerically illustrated by using the standardised IWA/COST simulation benchmark. The IWA/COST benchmark and the CBR approach proposed by the authors are firstly presented, emphasising the key steps in the definition of the case (or experience) and the CBR working cycle. Secondly, the methodology to use CBR within the IWA/COST simulation benchmark is outlined. Finally, a simulated case study illustrates the potential usefulness of this CBR approach in terms of providing support to plant operators when dealing with activated sludge process operational problems.

Materials and methods

IWA/COST simulation benchmark

The IWA/COST simulation benchmark plant was selected as the basis for illustrating the effects of considering a CBR system for WWTP management. The benchmark WWTP model allows biological nitrogen (N) removal control strategies to be evaluated via simulations and is described in detail by Copp (2002). The plant model is a predenitrification system consisting of two anoxic reactors, three aerated reactors and a secondary settler (Figure 1).

Biological processes are modelled using the ASM1 (Henze et al., 1987) and a one-dimensional settler model with the double-exponential settling velocity function (Takács et al., 1991) is used to model the settling process. Kinetic and stoichiometric model parameters, as well as basic operational conditions, are also provided within the benchmark description. The default control strategy consists of two control loops: one for dissolved oxygen in the last aerated tank, and one for the internal recirculation using a nitrate measurement in the second anoxic reactor. The default control strategy or any other
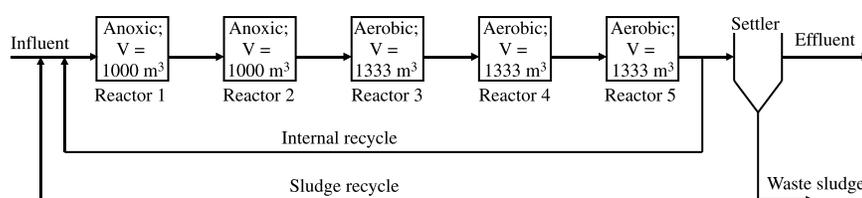

Figure 1 Lay-out of the IWA/COST benchmark plant

proposed control strategies are evaluated for three different predefined weather disturbance scenarios corresponding to dry weather, storm events and rainy days, respectively. Compared with dry weather, the rain event is characterised by a significant increase in the influent flow rate, combined with a decrease of the influent pollutant concentrations. The control strategy performance is evaluated by applying several performance criteria to the simulation output. These criteria include those defined in the original benchmark description (Copp, 2002), as well as the total operating cost index (TCI) proposed by Vanrolleghem and Gillot (2002).

All simulations were performed using the MATLAB (Mathworks, Inc.) implementation of the benchmark system. The simulations were run as specified by the benchmark protocol, i.e. perform a 150-day steady-state simulation to obtain adequate initial state values, simulate the plant with the dry weather scenario for 14 days, and then apply the dry, rain or storm influent file conditions for another 14 days. Only the last week of the simulation is used for plant performance evaluation.

CBR approach
Case-based reasoning can be defined as a knowledge-based technique that permits the use of past experiences to solve new problems that arise in a process. The basic idea behind the functionality of a CBR system is that it is usually easier to solve a similar problem the second time than the first time we face it, either because the previous solution is remembered and repeated, or because mistakes are recalled and subsequently tried to be avoided (Kolodner, 1993). The case definition and the implementation of the CBR working cycle are the two key steps required to make the CBR system operational.

Case definition. In this work, a case is a conceptualised piece of knowledge representing an activated sludge process experience. In our approach, a case could represent a rain event, as well as an episode with solids separation problems within the activated sludge process. A case is defined as a detailed description of the incident including an identifier, the problem diagnosis, the potential cause(s), the episode duration, a feature vector of some selected process characteristics, the applied control strategy and its evaluation, and the fundamental lesson learnt from the event.

CBR working cycle. The traditional CBR working cycle consists of a four-step process: retrieve, re-use, revise and retain (Aamodt and Plaza, 1994). A new problem is (if possible) solved by retrieving one or more previously experienced cases from the case library, re-using those experiences to produce a solution, revising that solution through simulation or test execution, and finally retaining the new experience by incorporating it into the existing case library for future use. The case library is a plant-specific database that stores the historical cases in an easily retrievable way. In order to improve the initial performance of the full-scale CBR system, the case library is typically initialised with some representative seed cases obtained from historical data through the application of data mining techniques (Comas *et al.*, 2001; R.-Roda *et al.*, 2001a). In this paper, the case library is initially empty. However, its size will increase dynamically over time as the CBR system and the benchmark plant are exposed to typical successive disturbances.

Further details of the CBR implementation for WWTP supervision can be found elsewhere (e.g. R.-Roda *et al.*, 2001b; Martínez *et al.*, 2006).

Methodology to use CBR within the IWA/COST benchmark
The proposed procedure followed to integrate a CBR approach within the framework of the IWA/COST simulation benchmark consists of six steps (see Figure 2).
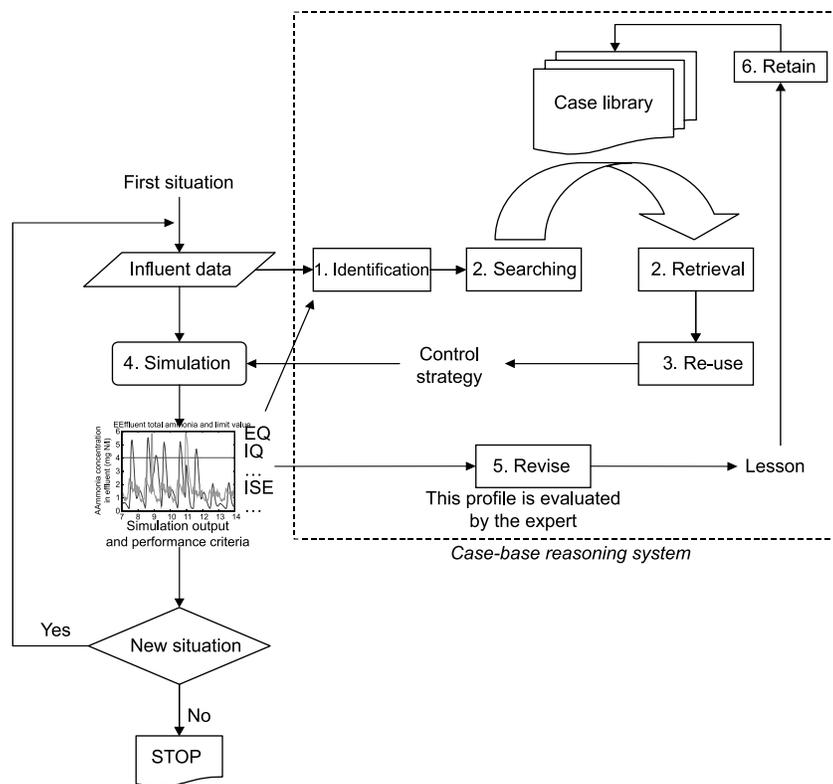
305

Figure 2 Procedure for a CBR approach for WWTP supervision, within the framework of the COST simulation benchmark

Step 1. Identification of the current disturbance/situation from the influent, effluent and process monitoring data. The identification of the current situation can be made by combining multivariate statistics and clustering techniques (Rosen and Yuan, 2000), by estimating similarities to the meta-cases or representatives of the possible operational status (Sànchez-Marrè *et al.*, 2000) or by diagnosing the status of the process by means of some heuristic rules (Martínez *et al.*, 2006).

Step 2. Search in the case library for cases resembling the current situation and retrieval of the most similar case. The technique used for case retrieval is based on a database query searching for at least partial matches between the current case and the historical cases. The similarity of a new case to the examples in the case library is computed by means of a distance function. The distance function combines all the partial-matchings through the relevant process features, into a full-dimensional partial-matching between the searched cases and the new case. Each feature has an importance value (weight) that is incorporated in the distance function (Sànchez-Marrè *et al.*, 2000).

Step 3. Re-use of the lesson retrieved from the historical case. Typically, the solutions have to be somewhat adapted to compensate for the remaining differences between the most similar historical cases and the new situation.

Step 4. Perform the simulation implementing the control strategy provided by the CBR system.

Step 5. Analysis of the simulation outputs, through examination of the effluent concentration profiles and calculation of the different performance criteria, combined with the reasoning based on the available process knowledge enable the control strategy applied to be revised. As a general procedure, this key task should be made by the WWTP

operators or process engineers because they are the ones with detailed knowledge about the plant. Since a general scheme is described here, this expert evaluation has been performed by the authors. The expert evaluation produces a piece of knowledge (lesson), incorporating experiential and/or heuristic knowledge, which can be derived from every encountered case.

Step 6. The last step of the CBR approach involves retaining the lesson learnt together with a detailed description of the current case into the case library. Useless experiences, i.e. cases that are not sufficiently different from all the examples that are already available in the case data base or non-used cases, are forgotten. This lazy learning gives the CBR system the capability to improve its reasoning ability and accuracy over time as new experiences are encountered and resolved.

The above iterative procedure is repeated every time the WWTP is confronted by a new situation. As an example to present the methodology, steps 1–6 will be performed twice in this paper.

Simulation results: a case study

Following the procedure illustrated in Figure 2, this section of the paper presents a simulated case study that illustrates the potential of the CBR approach, within the framework of the IWA/COST simulation benchmark, when dealing with an influent weather disturbance.

First rain event

A first problematic situation corresponding to a weather disturbance is influencing the benchmark plant. Specifically in the case presented, a long rain event has been selected because it results in a lower effluent quality and a higher cost than the dry weather conditions and the storm event. As this rain event is the first disturbance handled by the plant, and since the case library is initially empty, the CBR system cannot infer on an event that has never taken place before. Therefore a closed-loop simulation of the default benchmark control strategy using the rain influent file has been performed following the indications of the benchmark protocol. The simulation results for selected effluent pollutant concentrations and total suspended solids (TSS) of reactor 5 are shown in Figure 3.

After analysing the simulation results and performance criteria, and after evaluating the control strategy applied when faced with the influent rain disturbance, the experts conclude that the plant should be operated differently when facing such an event. Therefore, they agree that the RainEvent_ActionPlan_1 to improve plant performance during rain events should be:

1. The waste activated sludge (WAS) flow rate must be stopped during rain events to avoid excessive reduction of the biomass concentration; if rain events last for more than 2 d, waste 385 m^3/d until TSS in reactor 5 is lower than 4,500 g/m^3.
2. The return activated sludge (RAS) flow rate should be increased to 36,892 m^3/d (default value is 18,446 m^3/d, equal to the average influent flow rate during dry weather) to avoid sludge losses through the secondary effluent.

As a complement, the lesson learnt from cases can incorporate some actions that models cannot describe (i.e. protection of DO and nitrate sensors, increase the cleaning intensity of screens, check water level sensors, etc.). All this information together with a detailed description of the first rain event is retained in the case library as Case-RainEvent-1 (Table 1).

307
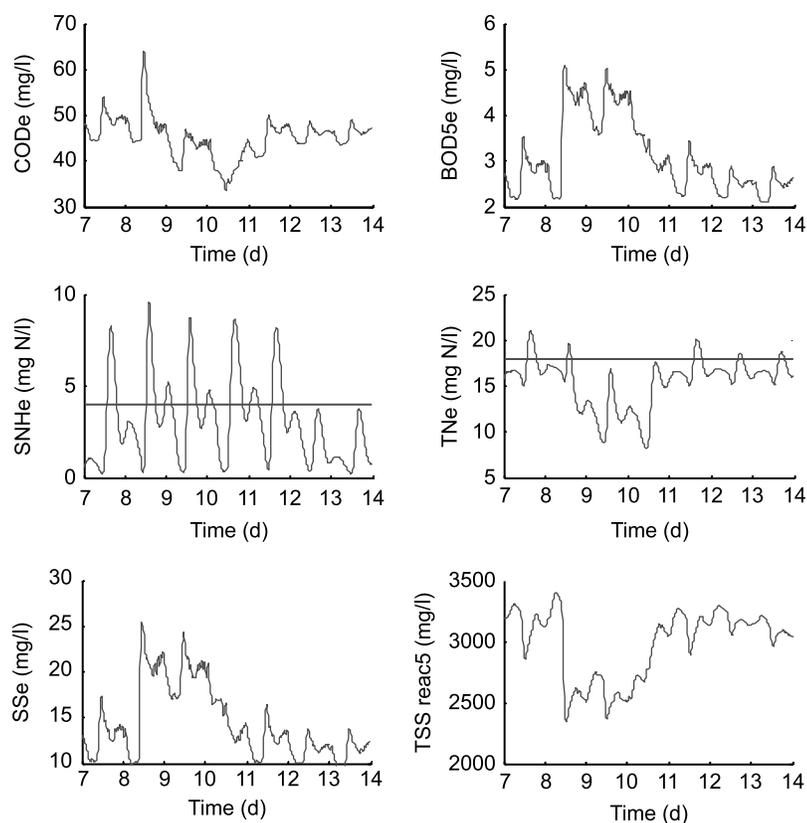
Figure 3 Effluent concentrations and TSS of reactor 5 for the default control strategy during rainy conditions

Subsequent rain events

A second rain event. When the simulated plant is confronted by a second rain event and identified by the diagnosis system, the CBR system is ready to search the case library, retrieve the most similar case by applying the similarity function and thus provide the experience of this previous situation. In this simulation study, since both rain events are the same, no adaptation of the retrieved control strategy (RainEvent_ActionPlan_1) is required. Otherwise, the proposed solutions are weighted according to the difference between the historical and the current case. The CBR system does not only present the fundamental lesson learnt from the previous experience to the operators, but also all the relevant information stored (Table 1). Then a closed-loop simulation of the benchmark plant exposed to this second rain incident is performed, but now applying the action plan provided by the CBR system while the disturbance lasts. The plant performance is also evaluated for the different control strategies tested in relation to the criteria defined by the benchmark (Table 2). The evaluation of the results presented in Table 2 shows that the new control strategy improves the overall effluent quality, and a lower number of violations of effluent total nitrogen and ammonia limit concentrations is obtained. Moreover, sludge production and energy required for pumping are reduced while energy consumption for aeration is increased. However, the operating cost index is significantly lower than in the default control strategy case. Thus it is illustrated that this second strategy is more effective than the original one when handling long rain events. The new rain event is also registered in the case library since we have a new experience that allows the CBR performance to be improved. Besides the control strategy applied, the

308

Table 1 A 41 h long rain event retained in the case library

Identifier	Case-RainEvent-1			
Diagnosis	Long rain event			
Episode duration	41 h			
Process characteristics (average values)	Inflow rate (m^3/d)	21,320	$COD_{effl.}$ (g/m^3)	44.9
	COD_{in} (g/m^3)	329.8	$TN_{effl.}$ (g N/m^3)	14.8
	TN_{in} (g N/m^3)	33.3	–	–
	–	–	TSS reac5 (g/m^3)	2562.4
Control strategy	DO setpoint: 2 g/m^3			
	WAS flow rate: 385 m^3/d			
	RAS flow rate: 18,446 m^3/d ($Q_{in,avg}$)			
	S_{NO} setpoint: 1.0 g/m^3			
Evaluation	Adequate C removal but incomplete N removal. Severe decrease of the biomass concentration during the rain event. Possible losses of solids in the effluent if too high sludge blanket level in secondary clarifier			
Lesson	In order to keep biomass in good condition and avoid solids losses in the effluent, during rain events it is recommended to:			
	Decrease the WAS flow rate to 0 m^3/d (if rain events last for more than 2 d, waste 385 m^3/d until TSS in reactor 5 is lower than 4,500 g/m^3)			
	Increase the RAS flow rate to twice the average dry weather Q_{in} (36,892 m^3/d)			
	Protect the DO and S_{NO} sensors, increase the cleaning intensity of screens and check water level sensors			

Case-RainEvent-2 incorporates a new lesson since the expert reasoning carried out during the revising step enables an even better solution for long rain events to be inferred. Since complete denitrification is not achieved, and especially because the effluent nitrate concentrations exceed 8 mg N-NO$_3$/L, the RainEvent_ActionPlan_2 suggests a RAS flow rate of 36,892 m^3/d as a new default value (both in dry and rainy periods) to prevent the occurrence of rising sludge in the secondary settler.

A third rain event. When the benchmark plant experiences another similar rain event, the CBR system searches again among similar cases contained in the case library. If Case-RainEvent-2 is retrieved as the most similar and no adaptation is required, the RainEvent_ActionPlan_2 will be re-used. Results for this proposed control strategy are shown in Figure 4. When applying this control strategy, significant effluent quality improvements are obtained, and the amount of sludge production for disposal is notably decreased since decay rates are increased due to higher biomass concentrations in the activated sludge tanks. The number of violations and the amount of time the plant is violating the effluent limits are reduced, particularly for the total nitrogen concentration. Energy for aeration is increased and again high energy for pumping is needed due to a

Table 2 Plant performances obtained using three different control strategies under rainy influent conditions

Performance criterion	Default cont. strategy	RainEvent_ActionPlan_1	RainEvent_ActionPlan_2
Op. cost index (€/year)	885,047	874,623	852,427
Effluent quality (kg/d)	9,038	8,959	8,652
Sludge production (kg/d)	2,358	2,230	2,049
Aeration energy (kWh/d)	7,170	7,262	7,252
Pumping energy (kWh/d)	1,927	1,881	2,051
No. of TN viol.[a] -% t viol.[b]	5–11.3	4–7.7	1–1.5
No. of S_{NH} viol.[a] – % t viol.[b]	8–26.8	5–14.7	5–7.1
No. of TSS viol.[a] – % t viol.[b]	0–0	0–0	0–0

[a]Number of violations of limits for TN, S_{NH} and TSS, respectively
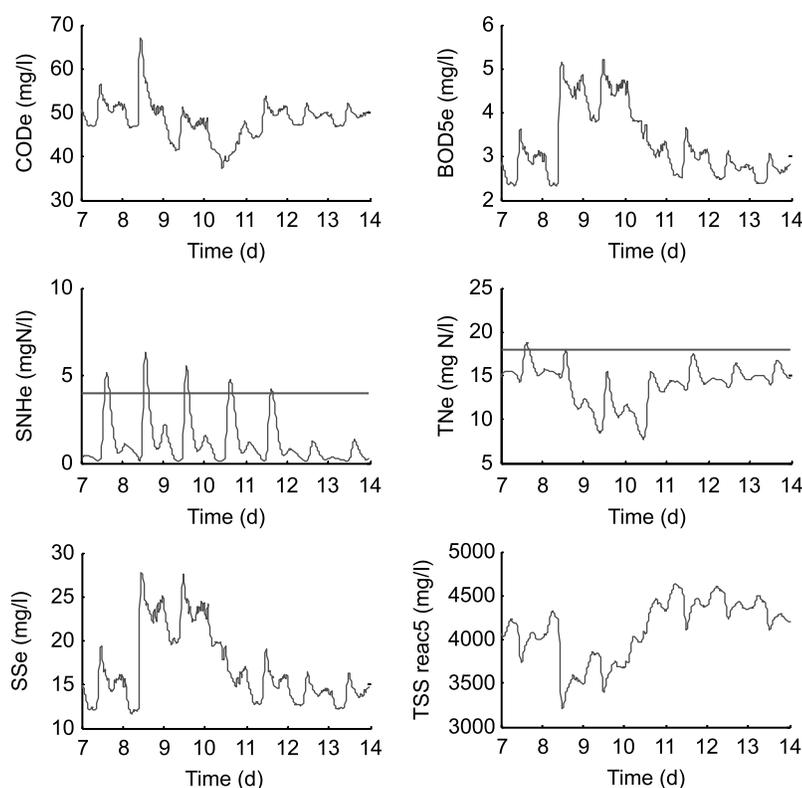[b]Percentage of time that the plant is violating the effluent limits

Figure 4 Effluent concentrations and TSS of reactor 5 for the RainEvent_ActionPlan_2 control strategy during rain conditions

higher RAS flow rate. Overall a lower cost is attained than by the other two control strategies, which reflects the higher effectiveness of the last simulated strategy (Table 2).

To be more realistic, the RainEvent_ActionPlan_1 and RainEvent_ActionPlan_2 were finally tested by using rain events different to the one handled by the basic control strategy. These new rain events were obtained by adding noise to the influent flow rate of the benchmark rain influent file. Evaluation of results show that these two *rain* strategies still make it possible to save about 28,000 €/year and 61,000 €/year, respectively, compared to the default control strategy.

However, in spite of these results, caution is required for stating that a CBR system will always improve real WWTP supervision since the simulation benchmark was not developed as a tool for evaluating plant-wide control approaches, which is the case of a CBR system, but for assessing the performance of low-level controllers. Besides, so far the CBR application for WWTP supervision was not compared to any other supervisory control. These limitations reveal the need for a platform for testing plant-wide control strategies (Jeppsson *et al.*, 2006).

Conclusions

In this paper, a CBR approach is presented for automatic learning and re-use of knowledge derived from past experiences in the activated sludge process. The potential and benefits of using such a system to cope with operational problems have been shown using the IWA/COST simulation benchmark. Integration of the CBR system for WWTP supervision within the benchmark protocol was illustrated. A case study has concluded that, if a CBR system is considered, the benchmark plant performance improves considerably

the second time a rain event occurs because knowledge learnt from earlier experiences can be re-used. It has also been shown that the knowledge retained by a CBR system is plant-specific and dynamic and that its performance increases over time. Further research includes evaluation of the CBR approach within a benchmark platform for plant-wide control and when confronted by other perturbations, such as toxic influent events or sludge settling problems.

References

Aamodt, A. and Plaza, E. (1994). Case-based reasoning: fundamental issues, methodological variations and system approaches. *AI Comm.*, **7**, 39–59.

Comas, J., Dzeroski, S., Gibert, K., R-Roda, I. and Sànchez-Marrè, M. (2001). Knowledge discovery by means of inductive methods in wastewater treatment plant data. *AI Comm.*, **14**, 45–62.

Copp, J.B. (ed.) (2002). *The COST simulation benchmark: description and simulator manual*. ISBN 92-894-1658-0, Office for Official Publications of the European Communities, Luxembourg.

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

Jeppsson, U., Rosen, C., Alex, J., Copp, J., Gernaey, K.V., Pons, M.-N. and Vanrolleghem, P. (2006). Towards a benchmark simulation model for plant-wide control strategy performance evaluation of WWTPs. *Wat. Sci. Tech.*, **53**(1), 287–295.

Kolodner, J. (1993). *Case-Based Reasoning*, ISBN 1-55860-237-2, Morgan Kaufmann, San Mateo, USA.

Martínez, M., Sànchez-Marrè, M., Comas, J. and Rodríguez-Roda, I. (2006). Case-based reasoning, a promising tool to face solids separation problems in the activated sludge process. *Wat. Sci. Tech.*, **53**(1), 209–216.

Poch, M., Comas, J., Rodríguez-Roda, I., Sànchez-Marrè, M. and Cortés, U. (2004). Designing and building real environmental decision support systems. *Env. Mod. Soft.*, **19**(9), 857–873.

Rosen, C. and Yuan, Z. (2000). Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering. *Wat. Sci. Tech.*, **43**(7), 147–156.

R.-Roda, I., Comas, J., Poch, M., Sànchez-Marrè, M. and Cortés, U. (2001a). Automatic knowledge acquisition from complex processes for the development of knowledge-based systems. *Ind. Eng. Chem. Res.*, **40**, 3353–3360.

R.-Roda, I., Sànchez-Marrè, M., Comas, J., Cortés, U. and Poch, M. (2001b). Development of a case-based system for the supervision of an activated sludge process. *Environ. Tech.*, **22**(4), 477–486.

Sànchez-Marrè, M., R.-Roda, I., Comas, J., Cortés, U. and Poch, M. (2000). Using meta-cases to improve accuracy in hierarchical case retrieval. *Comput. Sistemas*, **4**, 53–63.

Takács, I., Patry, G.G. and Nolasco, D. (1991). A dynamic model of the clarification thickening process. *Wat. Res.*, **25**, 1263–1271.

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

Wiese, J., Schmitt, A., Stahl, A., Hansen, J. and Schmitt, T.G. (2003). Experience management for wastewater treatment. GI-Workshopwoche Lehren-Lernen-Wissen-Adaptivität, Proc. Workshop on Knowledge and Experience Management, FGWM03, Karlsruhe, Germany, October 6–8, 2003.