# DAI-DEPUR: An Environmental Decision Support System for control and supervision of Municipal Waste Water Treatment Plants

**U. Cortés**[1] and **I. Rodríguez-Roda**[2] and **M. Sànchez-Marrè**[1] and **J. Comas**[2] and **C. Cortés**[2] and **M. Poch**[2]

**Abstract.** This paper introduces DAI-Depur, an Environmental Decision Support System (EDSS). This systems supports the supervision of Municipal Waste Water Treatment Plants (WWTP). The paper highlights the succesful application of AI techniques to a complex environmental problem as the wastewater treatment that has both social and economic impact as well as in the environment. DAI-Depur is successfully performing real-time support to the operation of the Granollers facility, in Catalonia, since September 1999.

## 1 GENERAL SPECIFICATIONS

The increasing degradation of the environment has forced the society to consider changes in human behaviour for ensuring the essential conditions for the life in the Earth. This consideration has encouraged research and a great effort has been placed on understanding, preventing and correcting environmental degradation. In this sense, the treatment of water and wastewater has become one of the most important environmental issues. Wastewater treatment is fundamental to keep the water natural resources (rivers, lakes and seas) in as high quality as possible. Not only for this environmental reason, but also due to the more and more restrictive social regulations, the correct management of wastewater treatment facilities has become very important during the last 20 years.

Environmental systems possess several inherent characteristics which make their understanding and control difficult: they evolve over time, involve processes which take place in a 3-dimensional space, are complex, involve interactions between physical-chemical and biological processes, are stochastic, and, very often, are periodic in time [6]. The complexity and the multi-faceted nature of many environmental problems, suggests that their suitable management cannot be based on a single technique. Municipal Waste Water Treatment Plants (WWTP) are clear examples of complex processes which meet all these distinctive features of environmental systems.

Though an enormous amount of research effort has gone trying to improve control and supervision of WWTP, its correct management is still far from beings successfully solved by the research community. Classical control methods based on mathematical models, have been successfully used to improve and optimize WWTP operation. However, these classical control methods show some limitations when trying to control the activated sludge process of WWTP, mainly when the plant is not working under the ideal (normal) state. There are many characteristics of the process that difficult the success of classical methods, *e.g.* the inflow is variable (both in quantity and in quality); not only there is a living catalyst (the microorganisms) but also a population that varies over time both in quantity and in the relative number of species; the knowledge of the process is scarce; there are few and unreliable on-line analyzers; and most of the data related to the process is subjective and can not be numerically quantified.

Although progress in control engineering, computer technology, and process sensors has enabled automatic control improvement, integrated operation of WWTP is still far from being solved. The number of measured variables in a WWTP is increasing and the need and possibility to control the process is becoming greater. With this increasing instrumentation there is certainly more information available, but it must be reminded that data rich is not the same as information rich [13]. It is not an easy task for operators and process engineers to acquire, to integrate and to understand all this day-to-day increasing amount of information. The solution could arrive with the development of knowledge-based decision support systems that handle the particular characteristics of the process, using this available but incomplete information to guarantee the quality of the discharged water. Although knowledge-based systems came into picture in the 1980s, some authors suggest that they never succeeded for two reasons: they were too complex, and the available knowledge could not be captured in reliable models and advisory systems [9]. A support system cannot only be based on mathematical modelling, but must also take advantage of heuristic knowledge from literature and experts, while including specific experiences accumulated through years of experience in the facility itself. There have been some approaches to improve WWTP operation using single knowledge-based techniques, such as [1, 11]. However, an effective decision support system to supervise the operation of the process should be described as a hierarchical multi-level structure that integrates different concurrent modules, overcoming the limitations in the use of each single technique, and providing higher accuracy, reliability and usefulness [4].

A reasonable proposal should link advanced and robust control algorithms to some knowledge-based techniques, allocating the detailed engineering to numerical computations, while delegating the logical analysis and reasoning to supervisory intelligent systems [15].

This paper describes the implementation of an Environmental Decision Support System (EDSS) to supervise and control the operation of a real WWTP. This EDSS integrates advanced control algorithms

[1] Technical University of Catalonia. c/Jordi Girona 1 & 3. 08034 Barcelona, CAT. email: ia@lsi.upc.es
[2] University of Girona. LEQUIA. Campus Montilivi s/n. 17071 Girona. CAT. email: ignasi@lequia.udg.es

with knowledge based and cased based systems in an hybrid architecture. The EDSS is performing real-time support to process operation since September 1999. The development of this architecure has been previously described in [10].

### 1.1 Plan of the paper

Section §2 describes the implementation of DAI-Depur the proposed EDSS into a real WWTP. Results of the first four-month validation period are also shown and discussed in §3. These results are compared with those of year 2001. The facility selected to implement the EDSS is located in Granollers, in the Besòs river basin (Spain). The water treatment line encompasses preliminary, primary and secondary treatment to remove the organic matter, the suspended solids and, under some conditions, the nitrogen contained in the raw water of about 130,000 equivalent-inhabitants. The sludge treatment line encompasses thickening, anaerobic digestion and dewatering. The raw influent comes from a combined sewer.

In §4 we give the final conclusions and talk about the key steps which are necessary to transfer the system to another facility and the economic and social impact of this process.

## 2 DAI-Depur

DAI-Depur is structured into five separated levels: data gathering, diagnosis, decision support, planning and action providing an agent based architecture with additional modularity and independence (a key factor to guarantee the re-design and transferability of the system to another facility). See Figure 1 shows the multi-layer architecture of DAI-Depur, connecting the user (*e.g.* the head of the plant) with the WWTP. The different tasks of DAI-Depur are performed in a seven-step cycle: data gathering and update, diagnosis, supervision, prediction, communication, actuation, and evaluation phase. The cycle is routinely executed once a day, although it can also be started manually at any time, and it is fired up whenever any alarm symptom is fulfilled.

### 2.1 First level

The first level of the EDSS encompasses the tasks involved in data gathering and registration into the real-time database. The data acquisition system, composed by three C++ communication bridges, gathers all kind of data collected in the Granollers WWTP. On-line data, previously acquired by means of an SCADA system (Supervisory Control And Data Acquisition), include digital and analogical signals from sensors and equipment (e.g. flow rate, pH, pump status, etc.). Off-line data include both numerical and qualitative information, provided from analytical determinations in the laboratory, operator observations, and the result of activated sludge examination (e.g. chemical oxygen demand, suspended solids, food to micro-organism ratio, protozoa biodiversity, presence of bubbles in the secondary settler, etc.). Table 1 shows the amount and frequency of the data gathered by the EDSS in the Granollers WWTP.

**Table 1.** Data gathered by the EDSS in the Granollers WWTP.

| Data source | Quantity | Frecuency |
|---|---|---|
| Sensors and equipments | 1039 | 15s |
| Analytical measurements | 110 | Daily |
| Activated sludge observation | 43 | Weekly |

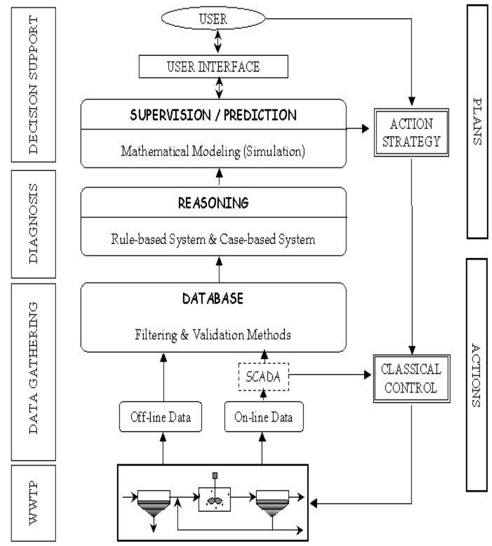


**Figure 1.** DAI-Depur Five-level Architecture

Original raw data are often defective, requiring a number of preprocessing procedures before being registered into the database in an understandable and interpretable way. The first step is the data validation, which includes filtering the correctness of the values (outliers), the noise, and the missing values. Then, all these filtered values are integrated into an homogeneous unit and time-scale. Since WWTP are dynamic systems, the optimal monitoring of the process state must also include information about the evolution of the main variables. Thus, the database calculates and analysis the derivatives of some specific variables to detect sudden deviations, trends, and periodicities. Moreover, some on-line data coming from sensors are reformatted in order to obtain different time scale averaged values and some daily accumulated values.

### 2.2 Second level

The second level of the EDSS includes the reasoning modules to infer the state of the process, in order to reach a reasonable proposal of actuation to support the whole plant supervision. The result of this layer is based on the combination between the heuristic knowledge of the ES and analogous experiences retrieved by the CBR. Both modules deal with qualitative information, and can interact with classical methods based on modelling and computing Expert System. The ES was codified into G2 [5]. The Knowledge Base of the ES reflects the problem-solving strategies of the process. The knowledge, codified by means of heuristic rules, was acquired from literature, from experts, and automatically from the database of the facility [12]. The structure of the Knowledge Base is modular using meta-rules and decision rules as the most appropriately knowledge representation scheme and reasoning strategy to handle sub-goals. Each module has a specific task and consists of different sets of rules, methods and/or procedures.

Once the ES is launched, it receives any requested data from the database. Then, the Data Abstraction module carries out a qualitative abstraction of the whole quantitative data as for example:

$$IF\ Suspended\_Solids - Effluent\ \ > 35\ g/m^3$$

$$THEN\ Suspended\_Solids - Effluent\ is\ high.$$

The boundaries among qualitative modalities like [*low, normal, high*] and [*increasing, maintaining, decreasing*] were first established by means of a statistical study of each variable. Afterwards, these ranges were submitted to judgement to the head of the plant, who finally adjusted them. Based on this qualitative abstraction, the Meta-Diagnosis module determines which tree and diagnosis paths (i.e. heuristic rules or procedures of the Knowledge Base) must be explored to infer the situation and to suggest an actuation strategy. Nowadays, the object oriented programming of the EDSS knowledge base in the Granollers WWTP consists of 440 rules, 2534 procedures, 179 text arrays, 218 quantity arrays, 1232 variables, and 71 objects. The Knowledge Base is structured into three main sub-modules: Fault Detection, Operational Problems and Transition States. The structure of the Knowledge Base is depicted in figure 2.

- **Fault Detection** module includes all the knowledge related to 8 operational faults due to mechanical equipment or electrical failures: e.g. damaged or clogged pumps and pipes, electrical fault detection, air system failure, sludge removal system break, etc.
- **Operational Problems** module includes the knowledge to diagnose 7 primary treatment and 17 secondary treatment problems, e.g. old sludge, storm, overloading, filamentous bulking, low pretreatment efficiency, foaming, etc. These problems are also divided into biological and non-biological nature depending on the cause of the dysfunction.
- **Transition States** module contains all the knowledge necessary to cope with transient states that can evolve in some of the undesirable problems contained in the previous modules.

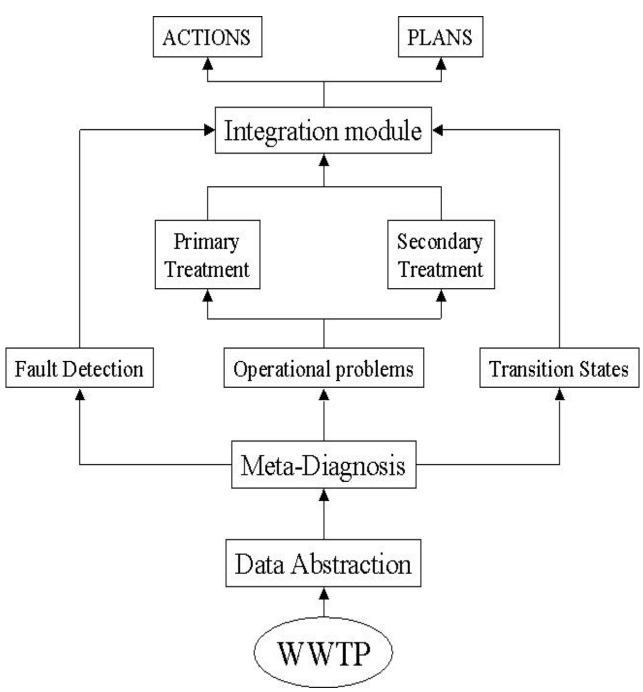


**Figure 2.** Knowledge Base

Whenever a *situation* is identified, the diagnosis task of the ES is reinforced with the detection of the cause of the problem. If the right cause of the problem is determined, a specific *actuation* is recommended (see§2.5). If the cause is not successfully determined, a non-specific actuation must be proposed to soften the effects of the trouble without tackling the real cause.

### *2.2.1 Case Based Reasoning*

In our approach each case is the codified description of a 24-hour period state or experience within the facility in a storable, easily retrievable way. Each experience is described through the most relevant measurements carried out in the process. The set of specific cases is stored in a structured fashion in a hierarchical case library. The 16 variables selected for our application include water and sludge flow rates, organic matter and nutrient concentration in different locations of the plant, biomass characteristics as settleability, biodiversity, or predominant species, and physical observations as presence of foam in the aeration tank. This selection was done according to the criteria of experts. Our CBR proposal is based on a working cycle [11] that consists of the following steps: i) gathering and processing data from the process to define the current case, ii) searching the case library and retrieving the case that best fits the current one, iii) adapting the solution if the retrieved case does not perfectly match the current case, iv) applying the adapted solution to the process, v) evaluating its consequences, and vi) learning details about the new experience.

To optimize retrieval and learning phases, the case library was organized in a hierarchical discretized tree [10, 14]. Among the whole historical days stored in the data base of the plant, twenty-five days representing typical situations in the Granollers WWTP were selected as initial cases to seed the case library (e.g. `storm, filamentous bulking, overloading, rising, foaming, and normal situation`). After the 4-month validation period, more than 100 new relevant experiences were learned and stored in the case library (see §3). The results for year 2001 are shown in §3.1

## 2.3 Third level

The third level of the EDSS establishes a supervisory and predictive task over the WWTP. The supervision task entails the gathering and combination of the conclusions from both knowledge-based techniques (ES and CBR modules), in order to identify whether there is a problem or not. The final diagnosis, together with the suggested actuation strategy, is sent to the user through the message-board of the computer interface. The expert evaluates the suggestions, checking its validity and deciding which is the best strategy to deal with the situation. A mechanistic model of the treatment process of the Granollers facility was developed using GPS-X, a commercial multipurpose modelling environment for the simulation of large-scale WWTP [8].

The biological reactor was modelled as four Continuous Stirred Tank Reactors, while primary and secondary settlers were modelled as one-dimensional tank with 10 layer of solids flux without biological reaction. The calibration procedure is still being carried out to adjust the kinetic, stoichiometric and settling parameters of the ASM1 model [7] and the settler model [16]. The standard values for these parameters given by the GPS-X were used in the first simulations except for those more relevant, which were changed manually. The model enables the EDSS to simulate several off-line scenarios with different operational conditions, changes in the influent characteristics (underloading, overloading, storms, etc), and alternative actuation strategies proposed by the EDSS. In spite of these capabilities, these kind of models show some limitations when dealing with problematic situations of biological origin (filamentous bulking, foaming, rising, etc.) as well as with situations for which it has not been calibrated. In this sense, the utilisation of soft-computing techniques to build a non-mechanistic model to simulate the behaviour of the plant at any situation is also being studied to be included as a new module into the EDSS [2]. According to the predictive results, the head of the plant performs a final validation and decides which kind of control should be carried out. He/she (acting as an expert) can maintain, modify or deactivate the automatic control over the plant (i.e. a closed loop for controlling the dissolved oxygen level in the aeration tank), supported by the actuation strategy that the EDSS suggested (expert, based on the reasoning procedure that the plant manager would do, or experiential, based on historical cases (real experiences) occurred in the plant). In fact, levels fourth (see §2.4 and fifth §2.5 are designed to either to support or to replace the plant manager in the decision making process.

This module also raises the interaction of the users with the computer system throughout an interactive and graphical user-machine interface. Moreover, at any moment and until a new supervisory cycle starts, the user can consult the conclusions of the EDSS as well as any quantitative or qualitative variable to know the state and trend of the plant.

### 2.4 Fourth Level

In this level plans are built and showed to the manager to guide the solution of problems. The action plans are structured as specific and/or non-specific depending whether the cause of the problem is well determined or not. The plans proposed by the ES are a list of general actions suggested in literature to solve an specific problem, which often do not fit well or cannot be directly applied to the actual facility. In that case, the particular strategy proven and validated by the manager of the studied plant must be collected.

As an example of a generic plan we have that for solving a Filamentous Bulking problem caused by the filament type `021N` due to a low `F:M ratio`, the utilisation of an anoxic zone to induce the selector effect could be recommended as a general action. But this plan has been shown not to be really effective for type `021N` control in the Granollers WWTP. Instead or at least, as complement for it, a gradual increase of wasting flow rate is needed and this action has been included by the manager in the learning phase of the CBR.

The set of available actions is described bellow. Those actions are contained in the Task Ontology and therefore they can be adapted to the actual situation, if needed. Also, this Ontology can be used to explain each action and its effect on the process. We have an experimental approach using a domain specific ontology, called WaWO [3], to help in the resolution of diagnosis impasses and to represent the transitions between two states of the process.

### 2.5 Fifth Level

The set of actions to be performed to solve problems in this domain have been defined for a generic facility and they could be easily adapted to others of the same architecture and technology. The set of these actions could be classified in three categories: a) Modify the state, b) Regulate a variable and, c) Add materials.

In the first category, for example, we have: Modify set control points and act upon the mechanical elements of the facility; in the second we have: Regulate aeration flow, waste flow, internal/ external sludge recirculation flow; and, in the last we have: Add chemical reagents, nutrients and alcalinity.

Of course the system *recommends* not only the action, or the sequence of actions (a plan), but a value that has to be accepted by the facility manager. This is the last level in the architecture that closes the loop. Its depicted in figure 1 as the box of `actions`. In §3.1 some data about the acceptance of these recommendations are given.

## 3 Results

The EDSS is performing real-time support to process operation in the Granollers WWTP since September 1999. During the first four months the system was validated. The main objectives of the validation process were to guarantee the right performance of the EDSS prototype, while checking for compliance with user requirement specifications. The methodology to validate the system was carried out through a two-stage validation procedure: Laboratory testing, which involves the execution of series of experiments to validate the correctness, consistency, and usability of each module. Face validation and historical cases techniques were combined to discover inaccuracy and inconsistency of the reasoning modules. Field testing of the overall EDSS, which was faced up to real cases to detect module integration errors, and to assure the system could deal with real qualitative variables and missing information. The flow of information throughout the system was strictly followed to detect weak reasoning. When necessary, the Knowledge Base and the Case library were refined, adjusted, corrected and/or extended. The ES module was tested in the laboratory with historical real cases.

The standard reference was the expert criteria (i.e. the answers of the ES were compared to the solutions given by different WWTP experts). The reasoning paths were followed, compared, and discussed with the assistance provided by the expert of the plant. This comparison enabled us to discover which of the rules and procedures had been correctly fired, checking for unnecessary IF conditions, and redundant, conflicting, subsumed, circular, dead-end, unreachable and missing rules. As a result, during this period, new rules, procedures or facts were added, whilst others were modified or deleted (rule refinement, reformulation and revision). On the other hand, the CBR module was validated according accuracy (set of acceptable responses) and adequacy to the domain knowledge covered. Expert experiences and historical cases with known solutions were used as standard reference. Several experiments were carried out to validate the hierarchical structure of the case library, the similarity measurement used, and the accuracy of using meta-libraries in the retrieval phase. If the CBR response differed from the expected output (i.e. the retrieved case suggested a diagnosis that was not similar enough to the standard reference diagnosis), details of the CBR module were revised (e.g. the relevant variables, their weight, the similarity criteria, the structure of the case library, etc.).

The second phase of the validation phase was carried out by means of field tests in the real facility. The main objective was to test the system within its real environment and to identify the need for further modifications. It was necessary to refine some modules of the EDSS, still uncovering unexpected errors or dealing with new cases not provided by the system. During this new 4-month period of validation, the EDSS was able to identify 123 different problem situations. From those, 79.7% were successfully identified (about one third in advance and the rest the same day), and 8.1% were wrong identifications (10 situations) while 12.2% were not identified (15 situations).

After these evaluation phases the system has been autonomously working in the facility.

### 3.1 Results of year 2001

After 16 months of autonomous functioning in the Granollers facility we collected a new set of resulting data for year 2001. In this period the ES performance was 79% and the CBR 82%. In this case we observed that the CBR has been improving its performance due to its ability to learn and adapt itself to new *situations* ocurred in the WWTP.

The list of correct situations detected during 2001 includes: Hydraulic Shock 21.3%, Non-biological secondary settler problems 15.9%, Primary settler problems 9.7%, Underloading 15.5%, Filamentous Bulking 12.6%, Rising Sludge 6.3%, Bulking tendency 4.3%, Transition state 3.9%, Overloading 2.9%, and other as Ammonia schock, organic matter schock, overloading tendency, etc.

In a 75% of the cases the actions proposed by Dai-Depur were accepted by the facility operator.

## 4 Conclusions

A hybrid knowledge-based supervisory system was developed and implemented to support the operation of a real WWTP. The system, which links classical control and numerical modelling to intelligent techniques, was structured into an agent-based architecture with three different levels: i) data gathering; ii) reasoning and diagnosis; iii) integration and decision support. In spite of certain reservations of the scientific community about the use of these techniques, the system has been successfully supervising the operation of the Granollers WWTP since September 1999. An 80% of correct identification of the process situation (one third of them in advance) during the first four-month validation process corroborates this conclusion. Validation of the proposed actuation strategies could not be evaluated because the EDSS still does not act directly over the process. A deeper revision of the whole system has been carried out during the last year (2001), in order to optimize the reasoning mechanisms of the tool, which is more robust enabling DAI-Depur to act automatically over the process. But this phase still under consideration.

Concerning the transferability of the EDSS to another facility, both technical and human bottlenecks must be seriously considered, in order to minimize time and cost effort. Among the long list of tasks to be scheduled, the key steps are: acquisition of specific knowledge of the facility, definition of the case and CBR calibration, adaptation of the ES Knowledge Base, communication of the EDSS to SCADA system and peripherals, validation of the performance, and delivery to the owners of the facility. Main human bottlenecks include skill and motivation of the operators, trust of the head of the plant in the EDSS, and expert paradigm.

## ACKNOWLEDGEMENTS

We would like to thank Consorci per a la Defensa de la Conca del riu Besòs, owner of the Granollers WWTP. U. Cortés, and M. Sànchez-Marrè like to acknowledge the IST-2000-28385 Agentcities.NET and the IST-1999-10176 A-TEAM projects. The views in this paper are not necessarily those of Agentcities.NET and A-TEAM consortia. We also had the partial support of CICyT TIC2000-1011 and CICyT DPI2000-0665-C02-01 grants.

## REFERENCES

[1] J. Baeza, *Desarrollo e implementación de un sistema supervisor para la gestión y control de una EDAR*, Ph.D. dissertation, Universitat Autònoma de Barcelona, 1999.

[2] Ll. Belanche, J.J. Valdés, J. Comas, I. R.-Roda., and M. Poch, ‘A soft computing techniques study in wastewater treatment plants’, *Artificial Intelligence in Engineering*, **14**(4), 307–317, (2000).

[3] L. Ceccaroni, *ONTOWEDSS. An ontology-base environmental decision-support system for the management of wastewater treatment plants*, Ph.D. dissertation, Universitat Politècnica de Catalunya, 2001.

[4] U. Cortés, M. Sànchez-Marrè, L. Cecaronni, I. R.-Roda., and M. Poch, ‘Artificial intelligence and environmental decision support systems.’, *Applied Intelligence*.

[5] Gensym, ‘G2 classic 5.1 rev. 7’, Technical report, Gensym Corporation, (2000).

[6] G. Guariso and H. Werthner, *Environmental Decision Support Systems*, Ellis Horwood-Wiley, 1989.

[7] M. Henze, C.P.L. Grady Jr.and W. Gujer, G.v.R. Marais G.v.R., and T. Matsuo. Activated sludge model no. 1. IAWPRC Scientific and Technical Reports No. 1, December 1987. IAWPRC.

[8] Hydromantis, ‘Gps-x version 3.0’, Technical report, Hydromantis Inc., (1999).

[9] G. Olsson, H. Aspegren, and M. K. Nielsen, ‘Operation and control of wastewater treatment - a scandinavian perspective over 20 years.’, *Water Science and Technology*, **37**(12), (1998).

[10] I. R-Roda, J. Comas, J. Colprim, M. Poch, Sànchez Marrè, U. Cortés, J. Baeza, and J. Lafuente, ‘A hybrid supervisory system to support wastewater treatment plant operation: implementation and validation’, *Water Science and Technology*. In press.

[11] I. R-Roda, Sànchez Marrè, U. Cortés, J. Lafuente, and M. Poch, ‘Consider a case-based system for control of complex processes’, *Chemical Engineering Progress*, **95**(6), 39–45, (1999).

[12] I. R Roda J. Comas, M. Poch, M. Sànchez-Marrè, and U. Cortés, ‘Automatic knowledge acquisition from complex processes for the development of knowledge-based systems.’, *Industrial and Engineering Chemistry Research*, **40**, 3353–3360, (2001).

[13] C. Rosen and G. Olsson, ‘Disturbance detection in wastewater treatment plants.’, *Water Science and Technology*, **37**(12), 197–205, (1998).

[14] M. Sànchez Marrè, U. Cortés, I. R-Roda, and M. Poch, ‘Learning and adaptation in wastewater treatmentplants through case-based reasoning’, *Microcomputers in CivilEngineering. Journal of Computer-Aided Civil and InfrastructureEngineering*, **12**(12), 251–266, (1997).

[15] G. Stephanopoulos and C. Han, ‘Intelligent systems in process engineering: a review.’, *Computers Chem. Engng.*, **20**(6-7), 743–791, (1996).

[16] I. Takács, G.G. Patry, and D. Nolasco, ‘A dynamic model of theclarification-thickening process’, *Water Research*, **25**(10), 1263–1271, (1991).