A Multi-paradigm Decision Support System to improve Wastewater Treatment Plant Operation

Ignasi R.-Roda, Joaquim Comas and Jesús Colprim

Laboratori d'Enginyeria Química i Ambiental. Universitat de Girona.
Campus Montilivi, E-17071 Girona.
Catalonia, Spain.
{ignasi, jcomas, jesus}@lequia1.udg.es

Juan Baeza

Dept. d'Enginyeria Química
Universitat Autònoma de Barcelona
Campus Bellaterra, E-08193 Cerdanyola.
Catalonia, Spain
baeza@uab-eq.uab.es

Miquel Sànchez-Marrè and Ulises Cortés

Dept. de Llenguatges i Sistemes Informàtics.
Universitat Politècnica de Catalunya
Edifici C5-Campus Nord. Jordi Salgado 1-3,
E-08034 Barcelona. Catalonia, Spain.
{miquel, ulises}@lsi.upc.es

Abstract

The paper suggests the integration of different fields in a multi-paradigm approach as the most optimal solution to guaranty the successful control of a complex environmental process such as Wastewater Treatment Plants (WWTP). This integration is able to supervise the different tasks of the system, to deal with any kind of data gathered from the process (quantitative and qualitative), and to include different kinds of knowledge: numerical, expert, experiential, and predictive. It is also detailed the behavior of the supervisory cycle of the DSS bringing face to a real case study, and how every kind of module including a mechanistic model of the process, an expert system, a case-based reasoning system, and a transition network works to reach an optimal solution.

Introduction

The increasing degradation of the environment has forced society to consider whether human beings are changing the conditions essential to life on earth. This sensation has encouraged research, and an enormous amount of effort has gone trying to understand and protect environmental degradation.

Environmental systems have several characteristics which difficult their understanding and control. Guariso & Werthner listed in (Guariso and Werthner, 1989) some of these distinctive features: evolve over time, involve processes which take place in a 3-dimensional space, are complex, involve interactions between physico-chemical and biological processes, are stochastic, and many times are periodic in time.

Water is the very essence of life in the Earth, but water is not only becoming scarcer, its quality is also being degraded. Everywhere, it seems, surface waters are being polluted with a frightening assortment of municipal, industrial and agricultural wastes. These wastewater contain a wide variety of contaminants, mainly suspended solids, organic matter, pathogens, and nutrients. Safe treatment of this wastewater is necessary to protect the health of the people, and also to prevent the occurrence of certain nuisances.

The main objective of wastewater treatment plants (WWTP) is to provide a regulated outflow of water with limited quantity of contaminants in order to maintain natural water systems at as high a quality level as possible, thus ensuring a good quality of life. Activated sludge is the most popular technology applied to treat biologically these wastewater. The basis of this technology lies in oxidizing biodegradable organics from wastewater into stabilized, low-energy compounds, maintaining a mixture of microorganisms (also called biomass) and supplying oxygen by aerators. Understanding and automatic control of WWTP is not well established. Apart from distinctive features of environmental systems listed above, there are some added characteristics that even difficult more their right operation: the lack of reliable on-line sensors, most of the data related to the process is subjective and cannot be numerically quantified, and the delay of some analytical determinations in laboratory. Today, effective control of WWTP is of critical importance not only for economic reasons but also to satisfy stringent environmental constraints.

This complexity of environmental systems, and the multi-faceted nature of many environmental problems, suggests that their suitable management can not be based on a single technique. To improve the control of such a complex processes, Stephanopoulos & Han in (Stephanopoulos and Han, 1996) remarked the necessity to develop a framework to integrate the processing of diverse forms of knowledge. This framework should include an array of specific supervisory intelligent systems (for the logical analysis and reasoning) and numerical computations for detailed engineering.

Following this idea, our proposal to improve WWTP control reflects the integration of classical methods with artificial intelligence techniques, including different kinds of knowledge: numerical, expert, experiential, and predictive. This integrated control system is structured into separate levels, providing an agent based architecture with additional modularity and independence, making the system easier to understand, to design, and to analyze (Sànchez et al., 1996). An schematic structure of our hybrid approach, clearly a Decision Support System (DSS), is shown in figure 1.

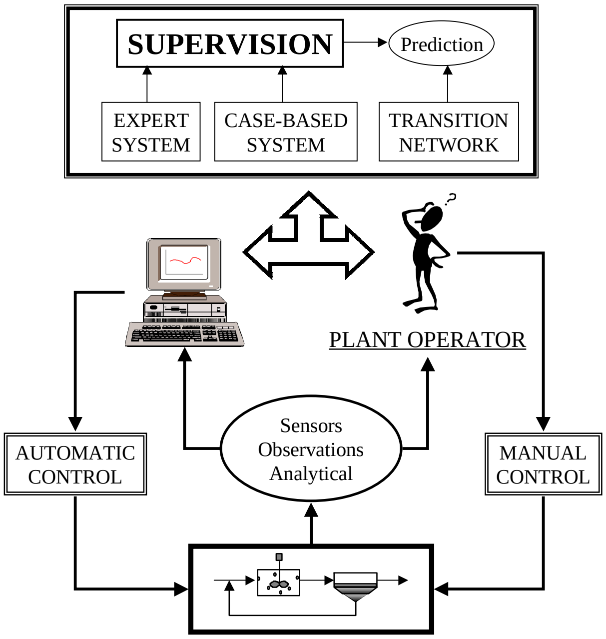

Figure1: Structure of the hybrid approach

The bottom of the figure shows the classical automatic control, which can be a simple control loop or an algorithm based on a mechanistic model of the process. This automatic control device, supervised by the manager or the plant operators, receives some of the information generated in the plant (provided from on-line sensors, from analytical determinations, and from different process observations). In the top of the figure appears the flowsheet of our hybrid approach, a DSS that collects the available information to infer the possible operating state of the plant. The Expert System (using heuristic knowledge), and the Case-Based Reasoning System (retrieving the most similar historical case stored in its library), running in parallel, send their conclusions to the supervision module, where the possible consequences of these actuation strategies can be evaluated by means of a predictive transition network. The final conclusion(s) of the DSS are sent to the operator through the computer interface or, when possible, the strategy can be applied automatically on the plant, modifying the set points of the automatic control device. This hybrid approach, which acts in supervisory cycles, requires an interface contact between the system and the manager or plant operators, to complement the data gathering with additional values and observations, or simply to check the inference way.

Some guidance to ease the development of this DSS and its application to any WWTP was presented in a protocol (R.-Roda, 1998). The protocol details a list of main tasks together with a prediction of time and requirements. Among these tasks: previous study of the plant, deep literature search, interviews with the operators, study of the historical plant database, Mechanical or black-box modeling, Expert System (ES) development, Case-Based Reasoning System (CBRS) development, Transition Network (TN) definition, integration, validation, application and final evaluation of the prototype.

The development of the DSS to supervise WWTP control is based on the knowledge acquisition from plant data base (specific knowledge that is used to adapt the knowledge base of the ES, to calibrate the mechanistic model of the plant, to define the initial seed of cases for the CBRS, and to model by means of a network the real transitions that took place among the different states of the plant), and from the literature (general knowledge that is mainly used as starting point when building the knowledge base of the ES).

The paper shows the current development of our DSS that is being carried out to improve the automatic control of a WWTP located in Catalonia. The behavior of its supervisory cycle bringing face to a real case study, and how every kind of knowledge included in the system help to infer the solution are also presented.

The WWTP and the Data Acquisition

The treatment plant selected to develop and apply our DSS proposal is situated in the basin of the River Besòs in Catalonia (Spain). This plant provides pre-, primary and secondary treatment using the activated sludge process to remove organic load and suspension solids contained in the raw water of about 130,000 inhabitants-equivalents.

This raw influent comes from a sewer collecting together urban and industrial wastewater. An exhaustive characterization of the wastewater quality is carried out in the plant, including both quantitative and qualitative data. Quantitative data is provided by on-line sensors (mainly flows and physical parameters like pH, conductivity or dissolved oxygen) and by the analytical determinations from the samples collected daily in the plant (measuring organic matter, nutrients, oils and biomass concentration). Some global parameters like SRT (Sludge Retention Time), SVI (Sludge Volume Index) or F/M (Food to Microorganism ratio) are inferred when requested. Qualitative data -measured in a lower frequency- includes microscopic determinations of the biomass (floc characterization, microfauna identification and counting, and filamentous bacteria identification and counting) and process observations of the operators (like presence and color of foam and sludge, or appearance of the settler supernatant and effluent).

A data acquisition system for the DSS has been successfully developed. The information generated by the sensors and the state of all the equipment distributed in the WWTP is monitored by a PLC (Programmable Logic Computer) network. These PLCs are connected to a Master-PLC, linked via RS-232 to the PC where the DSS resides. In the PC, a Data-Server program is in charge of communicate with the Master-PLC via RS-232 using a set of predefined messages. The Data-Server is built with G2-Gateway, a tool to develop extern data acquisition systems for G2 (Gensym, 1997). It is able to receive a data petition from G2 at any moment, and answering immediately. This communication is established via TCP/IP. This architecture permits to have different data servers in any computer connected via TCP/IP with the computer running G2 although, in our case, the Data-Server and G2 are running in the same PC. Analytical determinations and operator observations must be introduced by means of the PC interface.

The Multi-Paradigm Supervisory Decision Support System

Following the guides of the mentioned protocol, the DSS is being developed. Four tasks are focusing our main efforts: mechanistic process modeling, ES, CBS and TN. While the data acquisition system and the three first tasks have been successfully developed, the Transition Network is still in a preliminary stage and will not be presented in the paper.

Model

A mechanistic model of the treatment process of the plant was developed, using a commercial software -GPS-X (Hydromantis, 1995)- to run the simulations. Biological reactor was assumed to be as four Continuous Stirred Tank Reactors, while primary and secondary settlers was assumed to be one dimension tank with 10 layer solids flux model with biological reaction. It was conducted a preliminary simulation using default parameters of Mantis, a modified ASM1 model (Henze et al., 1987), process operational conditions and influent profiles. Then, an iterative calibration of the parameters was done to adjust the results of the simulation with real values determined in the laboratory. Biological reactor was assumed to be as four Continuous Stirred Tank Reactors, while primary and secondary settlers was assumed to be one dimension tank with 10 layer solids flux model with biological reaction.

Expert System

General and specific knowledge of WWTP has been extracted from an exhaustive literature review, interviews with the manager and plant operators, and an automatic classification of the historic database. This information has been gathered, structured and synthesized in a group of diagnosis and actuation logic decision trees.

Once validated by the experts of the process (in our case the manager of the plant), these logic trees were codified by means of heuristic rules conforming the Knowledge Base (KB) of a real time expert system. A commercial shell with a user-friend interface -G2- was selected as the suitable inference engine where the ES could be built.

The KB covers troubleshooting for primary sedimentation problems and secondary treatment problems, including non-biological problems (rainy days, storms, mechanical and electrical problems, overloading, underloading, aeration problems, and imbalanced flow rates) and problems with biological origin (among them, bulking sludge, foaming, rising, pinpoint floc, slime viscous bulking, and dispersed growth).

Case Based Reasoning System

A case is a codified description of a specific state or experience within the process in a storable, easily retrievable form. In our approach, the case was defined to correspond to a 24-hour period of the WWTP operation. According to the manager of the plant, the case was codified through the 16-most relevant variables measured routinely in the plant. These variables were prioritized and discretized into several modalities, such as low (L), normal (N), and high (H), and finally, as a result of an iterative process, a weight and a discriminant order were assigned showing their relevance in the characterization of the situation. The goal of this process is the organization of the case library into a hierarchical tree, which would optimize the retrieval phase. For more details of the case library see (Sànchez-Marrè, 1997).

Our CBRS proposal, which code has been developed in Common-Lisp, is based on a working cycle that consists of the following steps: gathering and processing data from the process to define the current case, searching the case library and retrieving the case that best fits the current one, adapting the solution if the retrieved case does not perfectly match the current case, applying the adapted solution to the process, evaluating its consequences, and learning details about the new experience.

Among the whole historical days stored in the data base of the plant, twenty-five days clearly representing real WWTP typical situations were selected as initial cases to seed the case library (e.g. storm, filamentous bulking, rising, foaming, overloading, and normal situation). Initializing the case library with a set of common situations, the CBRS is ready right from the start to propose solutions to problems similar to those in this initial "seed". The 16-most relevant variables selected in our application, as well as its discretization, discriminant order, and weight assignment, are showed in table 1.

<table>
<tr><th rowspan="2">Variable</th><th rowspan="2">Interpretation (units)</th><th rowspan="2">Discriminant order</th><th rowspan="2">Weight</th><th colspan="5">Modalities</th></tr>
<tr><th>Low (L)</th><th colspan="2">Normal (N)</th><th colspan="2">High (H)</th></tr>
<tr><td>Q-I</td><td>Inflow wastewater (m^3/d)</td><td>8</td><td>6</td><td>(< 6500)</td><td colspan="2">(6500 - 25000)</td><td colspan="2">(> 25000)</td></tr>
<tr><td>COD-I</td><td>Chemical oxidizable organic matter at the input (mg O_2/l)</td><td>4</td><td>5</td><td>(< 400)</td><td colspan="2">(400 - 950)</td><td colspan="2">(> 950)</td></tr>
<tr><td>NH3-I</td><td>Ammonia at the input (mg N/l)</td><td>13</td><td>5</td><td>(< 20)</td><td colspan="2">(20 - 100)</td><td colspan="2">(> 100)</td></tr>
<tr><td>COND-I</td><td>Waterflow at the output of primary settler</td><td>7</td><td>5</td><td>(< 1000)</td><td colspan="2">(1000 - 5000)</td><td colspan="2">(> 5000)</td></tr>
<tr><td>Q-P</td><td>(m^3/d)</td><td>10</td><td>2</td><td>(< 6500)</td><td colspan="2">(6500 – 25000)</td><td colspan="2">(> 25000)</td></tr>
<tr><td>COD-P</td><td>Chemical oxidizable organic matter at the output of primary settler (mg O_2/l)</td><td>9</td><td>5</td><td>(< 250)</td><td colspan="2">(250 - 550)</td><td colspan="2">(> 550)</td></tr>
<tr><td>COD-E</td><td>Chemical oxidizable organic matter at the effluent (mg O_2/l)</td><td>1</td><td>8</td><td>(< 40)</td><td colspan="2">(40 – 150)</td><td colspan="2">(> 150)</td></tr>
<tr><td>NO3-E</td><td>Nitrate at the effluent (mg N/l)</td><td>11</td><td>7</td><td>(< 1)</td><td colspan="2">(1 - 6)</td><td colspan="2">(> 6)</td></tr>
<tr><td>NH4-E</td><td>Ammonia at the effluent (mg N/l)</td><td>12</td><td>7</td><td>(< 25)</td><td colspan="2">(20 - 70)</td><td colspan="2">(> 70)</td></tr>
<tr><td>SVI</td><td>Measure of the sedimentability of the activated sludge (ml/g)</td><td>2</td><td>10</td><td>(< 50)</td><td colspan="2">(50 – 150)</td><td colspan="2">(> 150)</td></tr>
<tr><td>F/M</td><td>Food to microorganism ratio (kg/kg· d)</td><td>16</td><td>8</td><td>(< 0.25)</td><td colspan="2">(0.25 – 0.5)</td><td colspan="2">(> 0.5)</td></tr>
<tr><td>FOAM-AS</td><td>Presence of foam at the aeration tank</td><td>3</td><td>9</td><td>-</td><td colspan="2">(≤common)</td><td colspan="2">(> common)</td></tr>
<tr><td>BIODIV</td><td>Biodiversity of microorganims</td><td>6</td><td>10</td><td>(0 – 2)</td><td colspan="2">(2 – 6.5)</td><td colspan="2">(> 6.5)</td></tr>
<tr><td>BAC-PRES</td><td>Filamentous bacteria presence</td><td>15</td><td>8</td><td>-</td><td colspan="2">(≤common)</td><td colspan="2">(> common)</td></tr>
<tr><td>PRED-MIC</td><td>Predominant specie of microfauna</td><td>14</td><td>8</td><td>Flag<20</td><td>Attatched ciliates</td><td>Crawling ciliates</td><td>None</td><td>Rotifer</td></tr>
<tr><td>PRED-BAC</td><td>Predominant filamentous bacteria</td><td>5</td><td>10</td><td>S. natans</td><td>M.Parvicella</td><td>021/thiotrix</td><td>Nocardia</td><td>None</td></tr>
</table>

Table 1: Variable discretization and prioritization.

A Case Study

As an example in the present paper, a case study describing the supervisory cycle of our DSS proposal is presented. The case study is based on real data and corresponds to a problematic period of the WWTP performance.

As any other day, the supervisory system gathers all kind of information collected in the plant: on-line and analytical determinations from different sampling points (influent, primary treatment effluent, activated sludge, and effluent) and qualitative data corresponding to biomass microscopic examination and operator observations. Table 2 contains a selected piece of this information, showing only the information that can be useful to follow and understand the case study.

The supervisory cycle is routinely executed once a day, though it could be started manually whenever the operator requires it, or when there is a sudden alarm in the plant. Once the information has been gathered and processed, the two modules developed that are based on knowledge (ES and CBRS) are executed concurrently, using both the information, but without any kind of interaction between them. Next subsections show the different behavior of each module processing these data in order to conclude the diagnosis and to propose the solution.

On-line	Inflow: 18532 m^3/d Conductivity$_{Inflow}$: 3150 µS/cm Waste_flow_rate: 926 m^3/d
Analytical	COD$_{Inflow}$: 864 mg/l Ammonia$_{Inflow}$: 84 mg/l COD$_{Primary}$: 543 mg/l COD$_{Outflow}$: 128 mg/l Ammonia$_{Outflow}$: 58 mg/l Nitrate$_{Outflow}$: 1 mg/l Biomass concentration: 1423 mg/l
Global (inferred)	SRT: 9.6 days F/M: 0.22 Kg/Kg·d SVI: 117 ml/g
Microscopic	Floc_appearance: slightly-disperse Biodiversity: 6 Predominant_specie: attached_ciliates Filamentous: abundant Predominant_filamentous: *Nocardia* *Nocardia*: abundant
Observations	Foam_bioreactor: abundant Supernatant_clouds_presence: none

Table 2: Relevant data for the case study

Expert System

The ES can infer the situation of the WWTP following different paths of the logic decision trees that constitute its knowledge base.

In our case study, the diagnosis follows the paths based on the values of some numerical variables related to the global state of the process (low F:M ratio and high SRT) and on some qualitative information (referring to the abundant presence of foam on the bioreactor surface, the proliferation of filamentous bacteria, and an slightly disperse floc). As figure 2 shows, the ES activates an intermediate alarm, pointing the risk of filamentous organism proliferation.

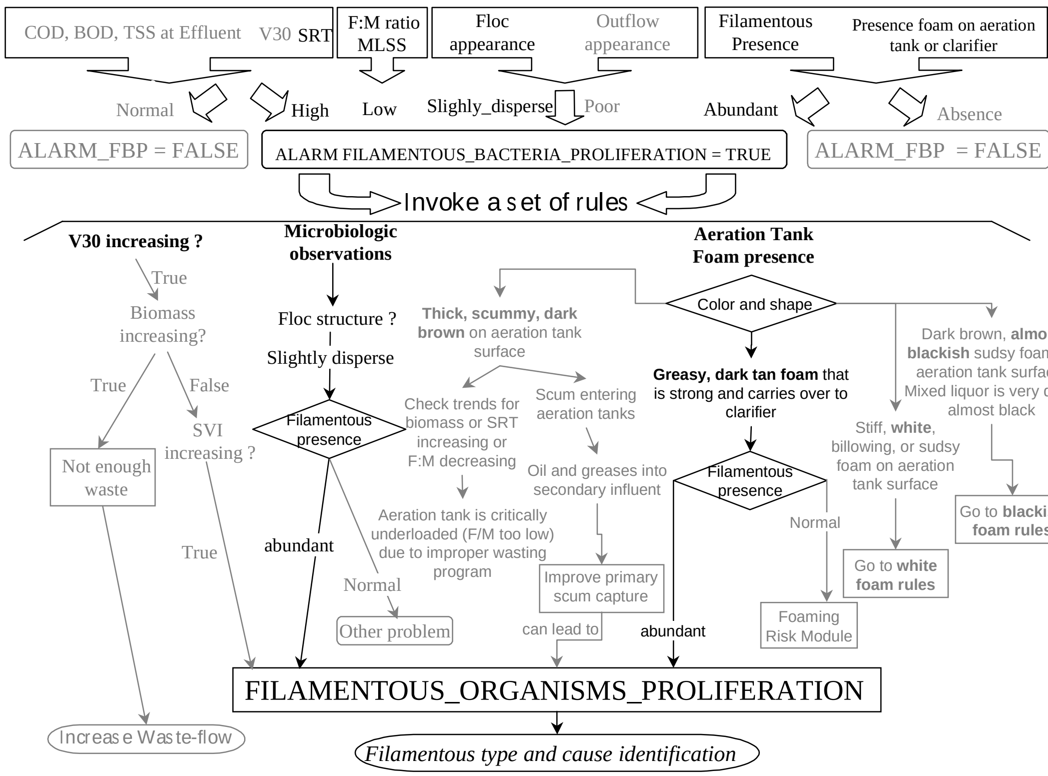

Figure 2: Decision tree for filamentous bacteria proliferation

In contrast, other possible conditions requiring quantitative or qualitative variables related to influent and effluent water quality (*e.g.* normal values of other analytical parameters like Total Suspended Solids (TSS), organic matter measured as chemical -COD- and biochemical oxygen demand -BOD- at the effluent, 30-minutes sludge settling volume (V30), and outflow_appearance) do not infer any diagnosis path.

When this intermediate alarm is activated, the inference engine of the ES invokes different diagnosis routes through the metarules to infer the situation (bold branches in figure 2), while other rules are not invoked (gray branches in figure 2). When the situation is identified as proliferation of filamentous organisms, then the ES activates a similar set of rules, corresponding to filamentous identification.

With the information detailed in table 2, the ES detects an abundant presence of *Nocardia* and common of *Microthrix Parvicella*, inferring the *Foaming* situation caused by these filamentous bacteria. Then, the inference engine of the ES calls the rules that must determine the causes of *Nocardia* and *M. Parvicella* proliferation. It is important to remark that the ES is still using the same information collected routinely in the plant, that is with minimum additional information required to the user through the computer interface.

If the right cause is determined, an specific actuation is recommended. In case the cause is not successfully determined, a non-specific actuation must be proposed.
Focusing again in the case study, the ES is able to conclude the causes of foaming (old sludge or high value of SRT and a sudden decrease in F/M ratio) and then, it could activate the corresponding specific actuation rules.

Case-Based Reasoning System

Case Based Reasoning System do not use all the available information collected daily, but only the 16-most relevant variables considered in the case library development (see table 1).

The first step of the CBRS cycle consists of searching and retrieving the most similar case recorded in the case library. This comparison is based on a similarity criteria. Our approach proposes a normalized exponential weight-sensitive distance function based on Manhattan criteria (for more details see Sànchez et al., 1996). In the case study, the best historic case included in the initial set correspond to Case #20 with a distance of 0.0587, a severe episode of *Foaming* caused by *Nocardia*. The case library does not only contain the case description, but also the expert diagnosis, a proposal of recommendations on how to keep the process under control, and the effect of applying the proposed solution, that is, whether the result was a success or a failure.

Once retrieved, Case #20 identifies foaming cause as "Oils in the influent", proposing "Physical removal of scums from bioreactor and settler, slight increasing of the waste flow rate, and checking for pretreatment operation and trends of oils in influent during the following 15 days. Check also for *Nocardia* trends during a week". The evaluation of the solution was "Successful: In five days *Nocardia* population starts to decrease".

Supervision and Prediction Process

The integration of the different information and solutions provided by the ES and by the CBRS is accomplished by the supervisor module, the top component of our DSS (see figure 1). This module is made up of a set of rules that must process the information received from the ES and the CBRS, evaluating any possible conflict. Moreover, this supervisor module suggests an actuation plan resulted of integrating the expert recommendations sent by the ES and the experiential actuation retrieved by the CBS.

In this case study, the situation inferred is the following:

"**FOAMING** caused by *Nocardia* : The proliferation of free and inside of floc filamentous bacteria with hydrophobs components, as *Nocardia* and *Microthrix Parvicella,* induces the apparition of foam and scums of biological origin at the biological reactor surface which could extend throughout the secondary settler and sometimes could provoke unintentional loss of biomass, reducing its concentration and causing bioreactor overloading".

While the solution proposes the following actuation strategy:

"According to the cause detected :

- Increase waste flow rate about 10%/day, until process approaches normal control parameters.
- Remove foam and scum physically from aeration tank and settler.
- Check for trends on *Nocardia* population every day.
- If foaming persists, consider the possibility of using the selector effect and the addition of chlorine into the recycled activated sludge stream. Send a message to operators recommending how to remove physically scums and how to add chlorine (doses, location, and test)"

Sometimes the integration of both proposals (from the ES and from the CBRS) is not so clear. In these cases, the effects of the solutions can be evaluated by means of the transition network. This prediction can help the operators to take the final decision, which can be based upon the expert (ES), experiential (CBRS) or numerical control (the automatic control loop based on the model). The latter option is preferred when the system infers a normal situation meaning a correct WWTP operation.

Finally, finishing the day or later, the manager of the plant makes an evaluation of the result of the actuation plan carried out, which allow to close the CBRS cycle.

Conclusions

A multi-paradigm approach for decision support systems has been described in this paper. This integration is able to supervise the different tasks of the system, to deal with any kind of data gathered from the process (quantitative and qualitative), and to include different kinds of knowledge: numerical, expert, experiential, and predictive. Also, a case study of a real application has beeen explained to show the potential of the approach. The designed architecture is currently being applied to the WWTP operation domain, with good and promising results.

Acknowledgements

The authors wish to acknowledge the cooperation of Consorci per a la Defensa de Conca del Besòs for their kindly contribution. This work has been supported by the Spanish CICyT projects BIO96-1229 and AMB97-889, and by the CIRIT SGR 177 (1997) of the Generalitat de Catalunya as a "Grup de Recerca Consolidat".

References

Gensym 1997. G2 Reference Manual 5.0. Gensym Corporation, Cambridge.

Guariso, G. and Werthner, H. 1989. *Environmental Decision Support Systems*. Ellis Horwood-Willey.

Henze, M.; Grady, C. P. L. Jr.; Gujer W.; Marais, V and Matsuo T. 1987. Activated Sludge Model No. 1, IAWPRC Scientific and Technical Reports, 1, IAWPRC, London.

Hydromantis. 1995. GPS-X Technical Reference. Hydromantis, Inc., Ontario.

R.-Roda, I. 1998. Development of a protocol for the application of knowledge based systems to wastewater treatment plants management. Ph.D. Thesis. Universitat de Girona (In Catalan).

Sànchez M.; Cortés, U.; Lafuente, J.; R.-Roda, I. and Poch, M. 1996. DAI-DEPUR: a Distributed Architecture for Wastewater Treatment Plants Supervision. *Artificial Intelligence in Engineering*, **10**(3):275-285.

Sànchez-Marrè M.; Cortés, U.; R.-Roda, I.; Poch, M. and Lafuente, J. 1997. Learning and adaptation in WWTP through case-based reasoning. *Microcomputers in Civil Engineering*, (Special issue: Machine Learning), 12(4):251-266.

Stephanopoulos, G. and Han, C. 1996. Intelligent Systems in process Engineering: a review. *Computers chem. Engng*, 20(6/7):743-791.