

DAI-DEPUR: an integrated and distributed architecture for wastewater treatment plants supervision

Miquel Sànchez, Ulises Cortés

*Secció d'Intel·ligència Artificial, Departament de Llenguatges i Sistemes Informàtics,
Universitat Politècnica de Catalunya, C. Pau Gargallo 5, 08028 Barcelona, Spain*

Javier Lafuente

*Unitat d'Enginyeria Química, Departament de Química, Universitat Autònoma de Barcelona,
Edifici C. 08193, Bellaterra, Barcelona, Spain*

Ignasi R. Roda & Manel Poch

Laboratori d'Enginyeria Química i Ambiental, Universitat de Girona, Plaça Hospital 6, 17071 Girona, Spain

(Received 7 March 1995; revised version received 6 October 1995; accepted 18 January 1996)

The activated sludge process — the main biological technology usually applied to wastewater treatment plants (WWTP) — directly depends on live beings (microorganisms), and therefore on unforeseen changes produced by them. It could be possible to get a good plant operation if the supervisory control system is able to react to the changes and deviations in the system and can take the necessary actions to restore the system's performance. These decisions are often based both on physical, chemical, microbiological principles (suitable to be modelled by conventional control algorithms) and on some knowledge (suitable to be modelled by knowledge-based systems). But one of the key problems in knowledge-based control systems design is the development of an architecture able to manage efficiently the different elements of the process (*integrated architecture*), to learn from previous cases (*specific experimental knowledge*) and to acquire the domain knowledge (*general expert knowledge*). These problems increase when the process belongs to an *ill-structured domain* and is composed of several complex operational units. Therefore, an integrated and distributed AI architecture seems to be a good choice. This paper proposes an *integrated and distributed supervisory multi-level architecture* for the supervision of WWTP, that overcomes some of the main troubles of classical control techniques and those of knowledge-based systems applied to real world systems.

Key words: integrated AI systems, distributed AI systems, knowledge-based systems, real-time supervision and control, wastewater treatment, knowledge engineering, environmental engineering.

1 INTRODUCTION

The motivations of this work were originated by a double source of unsolved problems: the proved insufficiency of chemical engineering classical control methods applied to WWTP, and on the other hand, some limitations of knowledge-based systems (KBS), in artificial intelligence as we will explain in sections 1.2 and 1.3.

1.1 Wastewater treatment plants domain

The main goal of a wastewater treatment plant is to reduce the pollution level of the wastewater at the lowest cost, that is, to remove — within the possible measure — foreign compounds (pollutants) of the inflow water to the plant prior to discharge to the environment, so that the effluent water has the lower levels of pollutants

as possible (in any case, lower than the maximum allowed by law).

The plants taken as models, in this study, are based on the main biological technology usually applied: the *activated sludge* process. The target wastewater plant studied is located in Manresa, near Barcelona (Catalonia). This plant receives about 30 000 m³/day inflow from 75 000 inhabitants.

A wastewater treatment plant (WWTP) is usually composed of two stand-alone but interactive sub-systems:¹ *sludge line* and *water line*. The water line is formed by a sequence of unit process operations where the effluent of one process becomes the inflow to the next one. Usually, there are three major processes: primary treatment, secondary treatment and tertiary treatment. Each one reduces or removes the concentration of several specific pollutants. The sludge line is commonly composed of three sequential processes in order to get an innocuous and compacted sludge: thickening, anaerobic treatment and drying processes. The flowsheet of a plant is depicted in Fig. 1.

The activated sludge process directly depends on live beings (microorganisms), and therefore on unforeseen changes produced by them. It could be possible to get a good plant operation if the supervisory control system is able to react to the changes and deviations of the system,

and can take the necessary actions to restore the system's performance.

1.2 WWTP control and supervision problems

The *complexity of the process* — composed of several operational units — makes difficult the implementation of an automatic process control over the wastewater treatment plant system. There are many factors influencing the system (most of them cannot be controlled, as for example the water temperature, flow variations, peaks, toxic loading, etc.) and, furthermore, the domain is ill-structured; there is a lack of understanding of the true mechanisms of the biochemical processes involved in wastewater treatment plants, and the relationships among different phenomena, which characterize the system, are not well-enough known, although different mathematical models have been put forward to describe them.

Most information is neither numeric nor quantified; qualitative information cannot be used in the context of a conventional control model, as for example microbiological information, water smell and appearance or state of the flocculation during sedimentation. This kind of information is essential for the operator of the plant, but is not suitable to be included in the context of a classical numerical control model. Another added difficulty is the uncertainty or approximate knowledge; the variables which describe the process are global and most of them cannot be obtained on-line. Therefore, the expert also has to take into account subjective information, based on local experience, which enables him or her to identify certain states of the plant.

Finally, the system is dynamic; it is under continuous changes that can directly modify the performance of the process. The plant is never working in steady-state conditions. Also, there is no possibility to control the feed to the plant, which has a wide variability in flow and concentration, or to control the microorganisms' behaviour. The WWTP system must be able to keep the outflow under environmental law limits, and minimize environmental effects.

So, all these features reveal that supervision and control of activated sludge processes could only be treated in a multi-disciplinary way,² that includes: *monitoring* (sensor developing, continuous analysis equipment), *modelling* (equations that model the bio-reactors' behaviour), *control* (maintaining good effluent water quality and reducing operation costs), *qualitative information* (microbiological information, water colour and odour, water appearance, etc.), *expert knowledge* (supplied by the much experience from plant managers, biologists and operators) and *experimental knowledge* (specific knowledge supplied by the previously solved problems in the concrete plant). The last three features commonly provide the systems with incomplete, uncertain or approximate information.

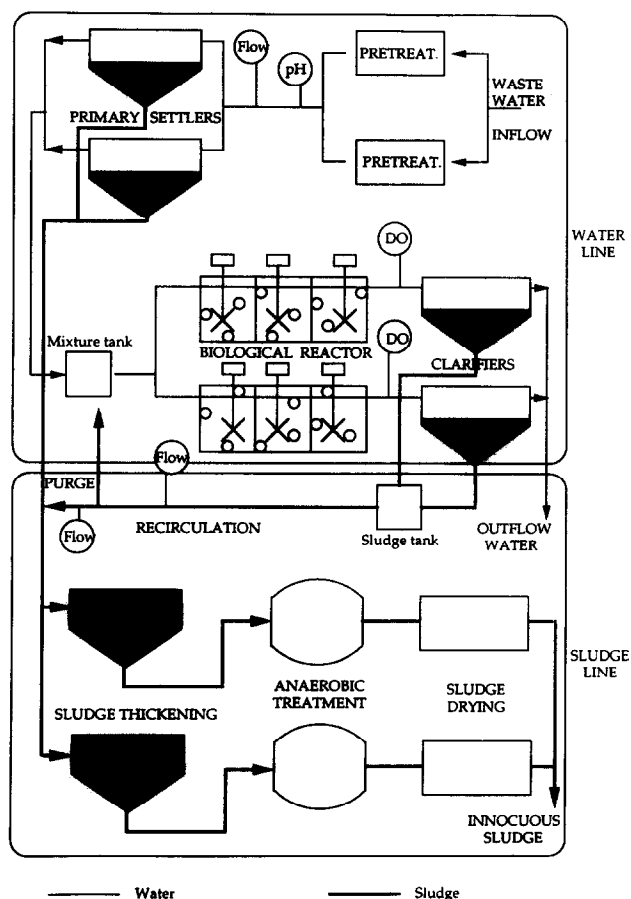


Fig. 1. Flowsheet of a wastewater treatment plant.

1.3 Knowledge-based control and supervision

The main characteristics of knowledge-based systems (KBS) point to the fact that they could be used for the supervision and control of wastewater treatment plants:³ usefulness in concrete domains, supporting numerical and/or symbolic information, specially useful in ill-structured domains, the treatment of uncertain or approximate reasoning.

Nevertheless, KBS do not incorporate some desired features from human intelligence and have some technical difficulties in their development:⁴ most KBS do not learn from previous cases, and the use of these previously solved cases is a valuable feature to be integrated in KBS; the *knowledge acquisition problem*, there are some difficulties in extracting the knowledge and experience from knowledge's sources; *brittleness*, their scope is limited to the forecasted situations in the domain, and they are not reliable when applied to unexpected situations; the increasing complexity of the systems, as the systems grow, it is more difficult to manage information and knowledge contained in them; lack of reusability, knowledge acquisition strongly depends on both the experts and the concrete domain, thus, it is very difficult for partial or global sharing and reuse of knowledge bases.

On the other hand, it is generally agreed that more powerful *knowledge acquisition* and *learning* tools and techniques are needed in order to increase both the quality and the quantity of KBS for real world supervisory applications. By this, we mean systems that exhibit a certain level of complexity, that sometimes have to cope with problems on the border (or slightly outside) of their special domain of competence (not brittle), and have to be properly self-updated and maintained (learning) in order not to degrade over time.

1.4 DAI-DEPUR's integrated and distributed architecture: an overview

One of the key problems in the design of knowledge-based control and supervision systems is the development of an architecture able to efficiently manage the different elements of the process⁵ (integrated architecture), to learn from previously solved cases (specific experimental knowledge) and to acquire the domain knowledge (general expert knowledge). These problems increase when the process is composed of several complex operational units. Therefore, an integrated and distributed problem solving architecture seems to be a good choice.

In Section 2 is presented some related work on chemical engineering classical control methods and, on the other hand, some knowledge-based approaches applied to WWTP control and supervision.

In this paper we will focus on the integrated design issues of our proposal (Section 3), which tries to overcome some of the troubles from knowledge-based

systems and those from classical control systems, in order to build-up a more efficient and robust system for the supervision and control of wastewater treatment plants. The supervisory integrated and distributed architecture proposes the integration of several interacting subsystems or agents, and the combination of problem solving capabilities, reasoning, as well as learning tasks in a single structure.

In Section 4, we describe an application of this architecture to a real plant. Section 5 discusses some of the results and provides some means of evaluation of the system. Section 6 gives the conclusions and suggests future research lines.

2 RELATED WORK

2.1 Chemical engineering classical control methods

The following methods are the different automatic process control configurations which have been applied to WWTP:

- *Feedback control*. This uses direct measurements of the controlled variables to adjust the values of the manipulated variables. The objective is to keep the controlled variables at desired levels (set points). It is commonly used to control the dissolved oxygen (DO) level in aeration tanks and it has also been studied to control substrate and biomass.⁶ A feed-back controller reacts only after it has detected a deviation in the value of the output from the desired set point.
- *Feedforward control*. Unlike the feedback systems, a feedforward control uses direct measurements of the disturbances to adjust the values of the manipulated variables. An example of feedforward control configuration was implemented in Luggage Point WWTP.⁷
- *Adaptive control*. There are two main reasons for using an adaptive controller in a WWTP. First, the process is non-linear (as the desired steady-state operation of the process changes, the best values of the controller's parameters change). Second, the process is nonstationary (its characteristics change with time). Different configurations of adaptive control have been proposed.⁸
- *Optimal control*⁹ and *predictive control*¹⁰ complete the list of automatic process control configurations applied to WWTP.

These conventional process control systems cannot work properly in a WWTP when there are either abnormal situations (storm, bulking, rising, etc.), or when there are unforeseen situations such as a mechanical fault (turbines, bridge of the clarifier, etc.) or when the available information is incomplete. Classical control processes have problems to be tuned and, sometimes, are not able to keep the system under control. Moreover,

most of them require numerical information; they need precise knowledge over the domain; they do not include the experts knowledge about the process, and their model of the system is rather static.

2.2 Knowledge-based systems approach

Some of the problems within the conventional process control systems described in section 2.1 have been the focus, during the last years, on much of the research efforts in artificial intelligence — specially in KBS — in many supervisory areas (monitoring of continuous processes, statistical process control, control of sun powered plants, control of petrochemical plants, etc.).

Specially related to the wastewater treatment area, KBS technologies have been developed as off-line consultations for: diagnosis,^{11–13} design,¹⁴ process optimization,¹⁵ etc. But all these approaches solve only certain aspects of the overall WWTP management process and cope with the KBS problems mentioned in section 1.3.

Thus it seems that the management of the whole WWTP process: system evaluation, diagnosis, supervision, actuation, etc., could be more easily achieved within an integrated and distributed AI architecture. *Distributed AI* (DAI) encompasses the research, analysis and development of 'intelligent communities' that integrate a coordinated set of knowledge-based processes, usually called *agents* (or actors or knowledge sources) that interact either by cooperation, by coexistence or by competition, in order to reach common objectives. The main reasons for distributing an AI system into a distributed problem solving architecture are:¹⁶ geographic distribution in the domain of application, functional decomposition, faster processing speed by means of parallel execution, modularity and extendibility, controlling the increasing complexity of AI systems and increasing the power of the resulting system. Also, there are some problems with a distributed architecture: how to maintain the global *coherence* of several agents involved in the architecture? How to *plan* a concrete solution for a given problem? How to *coordinate* agents' communication?

Although there are some areas where distributed AI systems have successfully been applied such as air traffic control, robotic systems, man-machine cooperation and office information systems, design, medical diagnosis, speech and natural language processing, etc., we have no knowledge of an integrated and distributed AI system applied to wastewater treatment plant control and supervision.

3 DAI-DEPUR ARCHITECTURE

3.1 Introduction

There are many kinds of distributed AI architectures which can be grouped into four main classes:¹⁷

blackboard systems (BBS), *supervisory systems* (SVS), *contract nets* (CN), *non-explicitly coordinated systems* (NECS). The main reason to choose a supervisory integrated and distributed AI system is because for WWTP there is a set of fixed abnormal situations such as storm, bulking, toxic loading, etc., that may be solved with a predetermined plan or strategy in a more efficient way than with other types of DAI architectures such as blackboard systems or contract nets. Furthermore, the integration of several techniques such as knowledge-based reasoning, case-based reasoning, learning, distributed problem solving and numerical control methods tries to cope with the troubles of single technologies that have been applied to real-world systems and to WWTP supervisory systems.

The integrated and distributed supervisory system (as shown in Fig. 2) is formed of several interacting subsystems (agents) that can be executed in parallel processing. Distribution criteria are based on spatial and semantic distances.¹⁸ In a WWTP there are some subsystems that are in spatially distributed locations: primary settlers, secondary-settlers, biological reactors, etc. (see Fig. 1). Thus, all of these agents are specialized, that focus on different aspects of the system. Each one has its own knowledge base.

The *water line subsystem* is a set of seven knowledge-based agents which diagnose and supervise the state of the water line subsystem. This task is done through their activation, which recalls information from the evolutionary real time data base and make their own

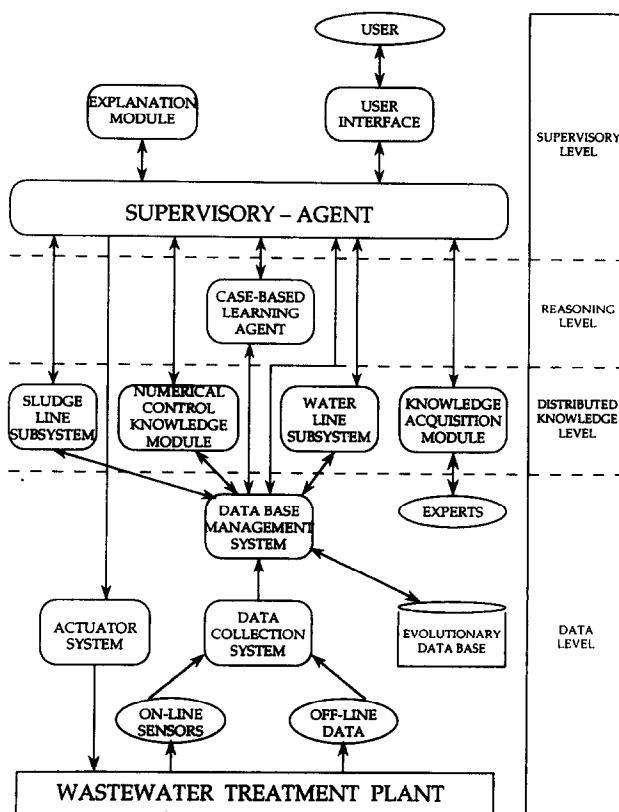


Fig. 2. DAI-DEPUR architecture.

inferences. This subsystem has already been developed, but is being re-structured.

The *sludge line subsystem* is a set of four knowledge-based agents which diagnose and supervise the state of the sludge line subsystem. This process will be done by means of their activation, which will recall information from the evolutionary real time data base and will come to their own conclusions. This subsystem is under development.

3.2 Levels of the architecture

The integrated^{19,20} multi-level architecture is composed of four levels as shown in Fig. 2: data level, distributed knowledge level, reasoning level and supervisory level. The multi-level feature provides the architecture with some independence among the levels.

3.2.1 Data level

The *data base management system* controls the access of the several KBS, CBRL agent, supervisory-KBS to the evolutionary (real-time) data base to guarantee the consistency of the system. The *data collection system* periodically receives data from the on-line sensors of the plant and sends them to the data base through the data base management system. The *actuator system* lets the system modify the on-line working parameters of the plant (turbines on/off, recirculating flow, etc.) under the supervision of the operator. This data level ensures a real-time environment for the WWTP system if there are real-time interfaces. If not, off-line supervision is required.

3.2.2 Distributed knowledge level

This general knowledge is obtained both from semi-automatic knowledge acquisition techniques and from expert knowledge. Agents forming the Water line subsystem are:

- Screen-KBS: agent supervising physical units that remove gross pollutants from the inflow of the WWTP.
- Grit removal-KBS: agent supervising physical units that remove grit to prevent abrasion and wear of mechanical equipment.
- Primary settler-KBS: agent supervising physical units with a long residence time that remove suspended particulates heavier than water.
- Biological reactors-KBS: agent supervising biological units that, using aerobic microorganisms (biomass or sludge), convert soluble BOD (biological oxygen demand) to new microorganisms.
- Secondary settler-KBS: agent supervising physical units with a long residence time that separate biomass from the liquid phase.
- Chlorination-KBS: agent supervising chemical units that allow, if necessary, to disinfect the outflow prior to discharge.

- Recirculation-KBS: agent supervising pumping systems to keep a certain level of biomass in the biological reactors.

Agents forming the sludge line subsystem are:

- Waste-KBS: agent supervising pumping systems that determine the sludge age (mean cell residence time, MCRT).
- Thickening-KBS: agent supervising physical units to increase the sludge concentration.
- Anaerobic treatment-KBS: agent supervising biological units that, using anaerobic microorganisms, convert biomass in methane (biogas), decreasing and percentage of BOD.
- Drying-KBS: agent supervising physical units that dry the sludge prior to discharge.

Each of these agents includes concrete knowledge about its local system, and all necessary elements for processing, validating and monitoring the information from this area of the process. This knowledge has been obtained by means of the knowledge acquisition module as will be described in Section 3.2.2.2.

From these measurements and observations, agents describe the behaviour of their local units and communicate their inferences to the supervisory agent. It uses all this information to infer the whole plant state. Once this whole plant state is diagnosed, the supervisory agent starts a procedure to execute an adequate strategy.

Numerical control knowledge. The numerical control knowledge module allows the system to simulate the actual plant operation, obtain simulated values for some required variables and implement a dissolved oxygen (DO) control scheme²¹ based on four main blocks.

- A mathematical model of the process.
- A software sensor to estimate the oxygen uptake rate (OUR).
- A continuous-range optimization procedure.
- An algorithm that, using the continuous-range optimal control value computed by the previous block, generates a discrete-range suboptimal control value, suitable to be applied by the aeration motors.

Knowledge acquisition module. The knowledge acquisition module is based in recent developments in knowledge acquisition. This module uses the software μ , which is the merging of LINNEO⁺²² and GAR²³ for automatic generation of inference rules as the result of a previous classification process of attributes and observations, defined by experts.²⁴

LINNEO⁺ is a knowledge acquisition tool that works incrementally with an unsupervised learning strategy which accepts a stream of observations and discovers a classification scheme on the data set. As a control strategy, it retains only the best hypotheses which are consistent with the observation given a similarity criterion. Part of the LINNEO⁺ methodology could

be considered as a *conceptual clustering* method with two critically important tasks:

- *Clustering*, which determines useful subsets of data using a fuzzy set approach, and *characterization*, which determines a concept for each extensionally defined set discovered by clustering.
- *Validation* from an expert to accept or reject the resulting clusters. Other modules try to exploit observational knowledge from the data set, or take advantage of the expert's knowledge if available. This knowledge is called domain theory (DT) and it is used to bias the process.

The main objective of LINNEO⁺ is to build classifications for ill-structured domains; where much imprecise information exists. It is assumed that observations vary in their degree of membership with regard to each class. Bearing all this in mind, the use of the conventional concept of distance as a fuzzy similarity value is used.

GAR (automatic rule generator) is used to generate a set of classification rules from LINNEO⁺'s output (a representation of the concept structure of the domain in terms of classes). GAR can generate both conjunctive and disjunctive rules, but after having analysed and compared several kinds of classification rules, one arrives at the following conclusions:

- The effectiveness of rule generation (defined as the specificity of a rule normalized in time) for conjunctive rules is the highest one.
- When delivered to experts, conjunctive rules are qualified as more understandable than other sorts of rules.
- When applying conjunctive rules, the reasoning process is faster.
- Conjunctive rules structure knowledge in a more modular way.

Therefore, these facts drive the system to output conjunctive rules. The algorithm for conjunctive rule generation could be summarized in this way:

- (1) select the best term;
- (2) add such term to the 'up to now' conjunctive premise;
- (3) reduce the set of possible terms;
- (4) repeat steps 1–3 while the rule is not completed.

For instance, a rule generated by GAR is:

```
(IF(>323.0 AFTER-PRIMARY-SETTLER-
    CHEMICAL OXYGEN DEMAND(COD))
(<7.7 INFLOW-pH)
(<93.0 GENERAL-CLEANSING-
    PERCENTAGE-BOD)
(>300 INFLOW-BOD))
→
Class-12)
```

which describes a high in-plant-overloading situation plus a poor sedimentation process in the primary settler.

3.2.3 Reasoning level

The case-based learning and reasoning agent manages a Case Library. This Case Library contains information about previously detected situations and solutions given to them as well as their efficiency (experimental or specific knowledge). A case-based reasoning²⁵ is performed in order to get benefit from these past experiences and cases. The Case Library is modified accordingly with the new information. The cases are previously experienced situations, which have been captured and learned, in such a way, that they can be reused in the solution of future situations.

The reasoning process in the case-based reasoning and learning agent²⁶ is performed by the following steps:

- *Retrieving* the most similar case(s) (previous working situations) by means of some heuristic functions or distances, possibly domain dependent. The normalized weighted distance — after a wide performance study — used to rank the best cases is:

$$d(C_i, C_j) = \frac{\sum_{k=1}^u e^{W_k} d(A_{ki}, A_{kj})}{\sum_{k=1}^u e^{W_k}}$$

where

$$d(A_{ki}, A_{kj}) = \frac{|\text{quantval}(A_{ki}) - \text{quantval}(A_{kj})|}{\div (\text{upperval}(A_k) - \text{lowerval}(A_k))}$$

if A_k is a lineal (ordered) attribute and $W_k \leq \alpha$

$$d(A_{ki}, A_{kj}) = \frac{|\text{qualval}(A_{ki}) - \text{qualval}(A_{kj})|}{\div (\# \text{mod}(A_k) - 1)}$$

if A_k is a lineal (ordered) attribute and $W_k > \alpha$

$$d(A_{ki}, A_{kj}) = 1 - \delta_{\text{qualval}(A_{ki}), \text{qualval}(A_{kj})}$$

if A_k is a categorical (not ordered) attribute

and,

C_i is the case i

C_j is the case j

W_k is the weight of attribute k

A_{ki} is the attribute k in the case i

A_{kj} is the attribute k in the case j

$\text{quantval}(A_{ki})$ is the quantitative value of A_{ki}

$\text{quantval}(A_{kj})$ is the quantitative value of A_{kj}

A_k is the attribute k

$\text{upperval}(A_k)$ is the upper quantitative value of A_k

$\text{lowerval}(A_k)$ is the lower quantitative value of A_k

α is a cut point on the weight of the attributes

$\text{qualval}(A_{ki})$ is the qualitative value of A_{ki}

$\text{qualval}(A_{kj})$ is the qualitative value of A_{kj}

$\# \text{mod}(A_k)$ is the number of modalities (categories) of A_k

$\delta_{\text{qualval}(A_{ki}), \text{qualval}(A_{kj})}$ is the δ of Kronecker

- *Adapting* or reusing the information and knowledge in that case to solve the new problem (the current working situation of the plant).
- *Evaluation* of the proposed solution. Usually, it is performed by simulation or by questioning a human oracle (in this case the human oracle is the plant manager). In the future, the evaluation will be done through automatic checking of the effectiveness of past solved cases.
- *Learning* the parts of this experience likely to be useful for future problem solving. The agent can learn both from successful solutions and from failed ones. This record is made by updating the Case Library accordingly.

The Case Library is implemented as a prioritized discrimination tree, where the priority of node-attributes is obtained from experts' judgement and from an inductive learning method (such as ID3). The Case Library is initialized with some prototypical situations obtained with LINNEO⁺ classification, from real data (see section 5):

- normal
- toxic-substances-loading
- storm
- plant-problems
- solids-shock
- primary-treatment-problems
- secondary-treatment-problems

It evolves from initial contents and captures the experimental knowledge of the concrete plant under control. Evaluation and adaptation steps are under development for DAI-DEPUR.

3.2.4 Supervisory level

The *supervisory-KBS* agent is the manager of the distributed system and acts as a master. It receives diagnosis information from the water line and from the sludge line subsystems and the most similar case retrieved from the case-based learning and reasoning agent. If the diagnosed working situation of the plant is *normal*, then automatic numerical control is activated or maintained. Otherwise, the supervisory agent notifies the operator of the plant of the current situation, suggesting the possible solution to take into account or directly acting over the system. About 20 different working situations of the plant have been defined.²⁴ Each one could be defined in terms of raw descriptions and relationships that effectively match the results obtained using LINNEO⁺. For example, the *bulking-non-filamentous* situation was defined as follows:

Outflow-chemical oxygen demand (COD)	→ high
Sludge age	→ old
Filamentous-presence	→ normal
Sludge volumetric index (SVI)	→ high
Recirculation-volatile suspended solids (RVSS)	→ low
All-other-attributes	→ nought-value (don't care)

The *user interface module* provides interaction between the operator and the system through visualization of a chart of the plant, asking and answering the operator and/or system inquiries, evolution of the system, diagnosed working situation, integration of process data obtained off-line in the laboratory, etc. The *explanation module* gives some explanations about conclusions reached of the different KBS agents of the system (supervisory agent, screen agent, thickening agent, etc.) and could give some required reports about deductive processes.

3.3 The supervisory cycle

The interaction of the several processes involved is done through communication channels and centralized by the supervisory agent process. Other processes interact with it asynchronously at different steps of the supervisory time cycle. These communications are shown in Fig. 3. The system activates a new supervisory cycle at fixed intervals of time. Each cycle is formed of four steps: diagnosis or evaluation phase, learning and reasoning phase, supervisory and communication phase, and actuation phase.

3.3.1 Diagnosis or evaluation phase

In a new cycle, the supervisory agent activates the water line agents to diagnose the state of the different subsystems of the plant (biological reactor, primary settler, recirculation, etc.). This means concurrent

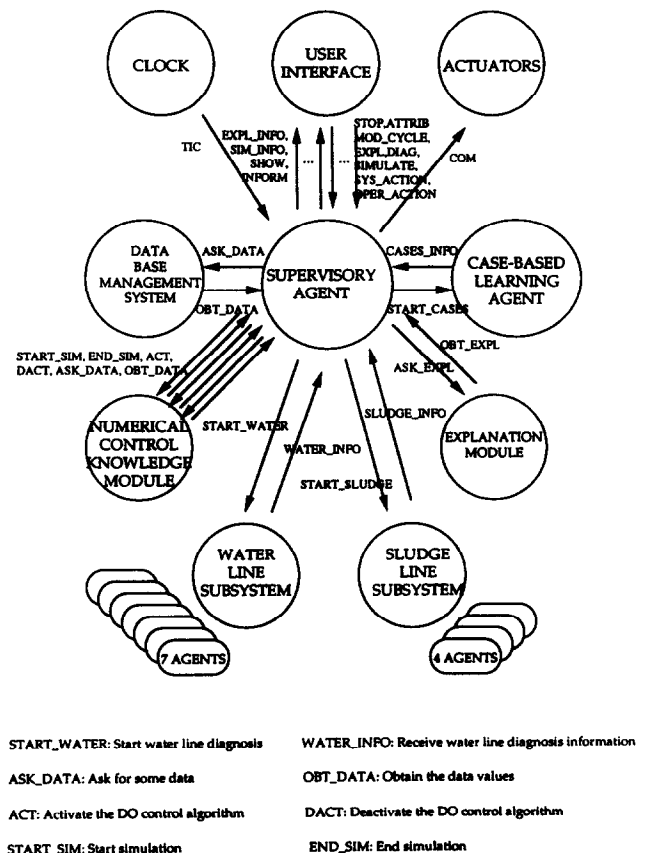


Fig. 3. Concurrent processes interaction.

processing of all agents involved. For this purpose it is necessary to know some values for certain variables of the process. All this data can be extracted from the evolutionary data base, fitted either with the on-line sensor values coming from the data collecting systems or with some other features provided by the operator (such as a microbiological laboratory analysis, qualitative observation, etc.).

3.3.2 Learning and reasoning phase

At the same time as the diagnose phase, in concurrent execution, the case-based reasoning and learning agent (CBRL) is activated to retrieve similar cases recorded in the Case Library. Next, the most similar one is updated in order to adapt it to the current situation of the plant. During this task access to the data base is required. The results are communicated to the supervisory agent. Afterwards, the agent can learn from its successes or failures.

3.3.3 Supervisory and communication phase

The supervisory agent combines all information coming from the several KBS agents (*general knowledge*) and from the CBRL agent (*specific knowledge*) to infer the current global situation of the plant and the suggested actions to be taken. It sends this information to the operator through the user interface module. The system can be used for explanations, retrieval of certain values, etc.

3.3.4 Actuation phase

If the *normal* situation has been detected the automatic control is maintained or activated. In *abnormal* situations, the supervisory agent waits for the operator's validation of the actions in order to update the current working state of the plant. If there are on-line actuators, the plant can be automatically updated through the actuator system. If not, manual operation is required.

4 APPLICATION

As an example in the present paper, a case describing the integrated reasoning system of DAI-DEPUR is presented. Focusing on the secondary settler agent, the information considered is (see Fig. 4): inflow, temperature, inflow COD (chemical oxygen demand), inflow biomass, outflow COD, outflow suspended solids, recirculation flow, recirculation biomass, sludge level and qualitative observations. With all this information, the secondary settler agent may detect different local situations:

- increase or decrease in the sludge level (slowly, quickly, etc.)
- poor sedimentation
- presence of bubbles
- change in the efficiency of the solids removal

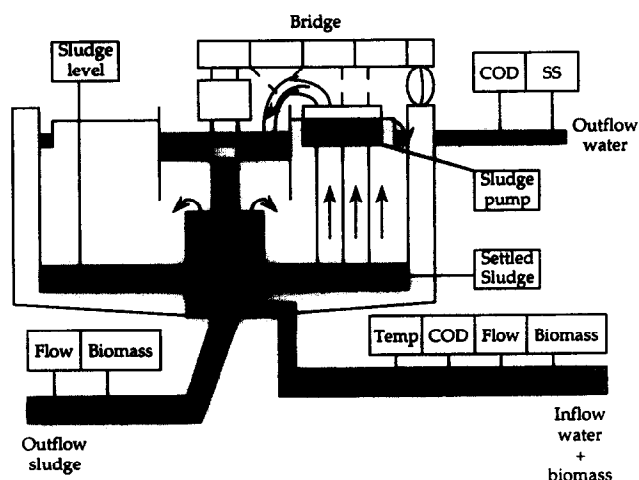


Fig. 4. Secondary settler or clarifier.

- some mechanical problem (pumps, bridge, etc.)
- flow bursts

Following the example, if the secondary settler agent detects a slight increase in the sludge level combined with a decrease in the biomass concentration in the recirculation flow and a clear sobrenatant, he concludes a *bulking* suspect (poor sedimentation of the biomass). This alarm arrives at DAI-DEPUR supervisory agent, which requests — from the biological reactors agent — the behaviour of the sludge volumetric index (SVI) value and microscopical observations to conclude the *bulking situation*. Then it starts a planning strategy to determine the causes of this situation using information coming from all agents. Also, the CBRL agent is simultaneously activated to retrieve any similar previous cases in the past.

In particular for the present system, the bulking causes considered in the supervisory agent are:

- low DO level
- nutrient deficiency
- low F/M (food/biomass) ratio
- sulphur presence
- high pH variability

Immediately after problem detection, the supervisory agent executes the bulking strategy to avoid biomass decrease in the system. These strategies are different for each kind of bulking. Each of them activates a supervisory agent's specific area which contains some knowledge to control the bulking problem.

For example, if DAI-DEPUR has detected that a possible cause of the bulking is low-DO-level (due to the presence of filamentous microorganisms *S. Natans*, type 1701 or *H. Hydrossis*) then, it requires, in Manresa's plant, the operator to check DO-sensors or the DO-set-point executing the next rules:

```

RBU04  IF      (bulking is true and the cause is low-
                DO-level)
        THEN    (check the status of the DO-sensor)
RBU05  IF      (the status of DO-sensor is OK)

```



```

      THEN (increase the set-point of DO-
RBU06 IF (the status of DO-sensor is wrong)
      THEN (calibrate the DO-sensor keep the
            DO-control)
RBU07 IF (the status of DO-sensor is not-
            possible-to-check)
      THEN (increase the set-point of DO-control
            keep the alarm-situation
            send message-34 to operators)

```

The main difference between this system and a classical control one is that the second just could act on the set-point of the controller, but DAI-DEPUR may activate or modify the classical control strategy, or if it is necessary look for operational problems and errors in sensors. In addition, it can retrieve some similar previously detected situations in the past operation of the plant, and update them to solve this new situation. Finally, it can learn from this new experimented situation (case).

5 EXPERIMENTAL RESULTS AND EVALUATION

The experimental evaluation and validation of the system is incrementally being done at several points during its development:

- Validation of each KBS agent and modules: knowledge bases obtained from the knowledge acquisition module, CBRL agent, the numerical control knowledge module, etc., are tested with the experts' opinion and with some real data from the plant taken as a model.²⁴
- Validation of the whole system. The whole system is being validated at three levels: (a) simulation of the plant in real time, (b) building-up and testing on a pilot scale plant and (c) validation on a real plant.

The real data used to build-up the LINNEO⁺ classifications prior to initializing the Case Library are available via anonymous ftp from the UCI Machine Learning Repository of data bases (ftp.ics.uci.edu). They are 527 data (days) corresponding to the period 1990–1991. Each piece data is described by means of the daily mean of 38 variables. Of these 38 variables, 29 correspond to measurements taken at different points in the plant, while the remaining 9 variables correspond to the calculated performance of the primary and secondary treatments for the whole plant. Now, a new real data set is being studied from another WWTP which uses a slightly different technology. That new plant is located in Cassà de la Selva-Llagostera, near Girona (Catalonia), and receives about 2500 m³/day inflow from 10 000 inhabitants. WWTP operation is being simulated for validation purposes with a GPS-X simulation package²⁷.

A wider evaluation of DAI-DEPUR will be possible

due to the construction of a pilot scale WWTP that will be connected in parallel to the Manresa plant. So it will be possible to create non-standard and/or dangerous situations — those that cannot be tested at the plant lest the environment may be damaged — and to measure our system's performance with real data properly scaled.

6 CONCLUSIONS AND FUTURE LINES

An integrated and distributed supervisory architecture (DAI-DEPUR) for wastewater treatment plant supervision has been designed and is currently being developed. It is composed of several knowledge-based systems (agents) and other processes (CBRL, DBMS, etc.) that can be executed concurrently. This integrated architecture approach has several advantages that make it more powerful than other single technologies applied to wastewater treatment plants as knowledge-based approaches (see section 2.2), the same as to other complex ill-structured domains:

- It makes it possible to reason in an ill-structured domain, where other kinds of reasoning like model-based reasoning or algorithmic reasoning would not be possible or easily formulated.
- It allows the system to learn from previously solved problems and to adapt the available experimental knowledge over the domain (dynamic learning environment).
- It overcomes the brittleness of KBS in coping with unforeseen situations (not previously considered by the general expert knowledge), trying to solve them by means of the most similar situation in the Case Library.
- It captures the knowledge provided by the experts (knowledge acquisition) which is very important, although subjective, to get a central corpus of knowledge about the domain.
- The integration of problem solving capabilities, reasoning and learning tasks in a single system.
- The cooperation of knowledge-based reasoning (general expert knowledge) and case-based reasoning (specific experimental knowledge) to deal with either prototypical or idiosyncratic situations.
- Due to the dynamic learning environment, the system is able to self-adapt to different wastewater treatment plants, making the system reusable in any plant with some minor changes. It is only necessary to fill the Case Library with an initial set of specific cases (operating situations of the concrete WWTP), which can be obtained semi-automatically from real operational data.

On the other hand, the design and implementation of the integrated and distributed problem solving architecture is more difficult and complex than other single-technology approaches applied to WWTP supervision. With this integrated approach, the plant can be

controlled in normal situations (mathematical control), usual abnormal situations (expert control) and unusual abnormal situations (experimental control).

There are some features to be considered for the future:

- New research direction points to consider that knowledge bases would not be static but dynamic ones. Inference rules (control or expert ones) could be adapted in the same way as a CBRL agent updates the Case Library according to new experience.
- Perhaps, the study of feasibility of a model-based reasoning to increase the power of the system, complementing the knowledge-based reasoning and the case-based reasoning, although it would have more potential if the WWTP domain were a well-structured domain.

Also, there are some related questions that research in WWTP field has to bear in mind, such as: the development of an automatic pattern recognition of microbiological images that can help to capture this useful information into the system²⁸ and to apply the data set and knowledge bases as retrofitting information to optimization in the design of wastewater treatment plants.²⁹

ACKNOWLEDGEMENTS

Currently, the project is partially supported by the 'Junta de Sanejament de la Generalitat de Catalunya' and the Spanish CICYT project ROB94-679 and EEC project VIM ERBCHRXCT 930401. The authors wish to acknowledge the gentle cooperation of Ricard Tomàs, manager of the Manresa wastewater treatment plant. Also, they wish to acknowledge the efforts and interesting comments and suggestions of the anonymous referees on our previous version of the paper.

REFERENCES

1. Metcalf & Eddy Inc., *Wastewater Engineering: Treatment/Disposal/Reuse*, 3rd edn. McGraw-Hill, New York, 1991.
2. Venkatasubramanian, V., Towards integrated process supervision: current status and future directions. In *Proc. of 2nd IFAC Workshop on Computer Software Structures Integrating AI/KBS Systems in Process Control*, Lund, Sweden, 1994, pp. 9–21.
3. Stephanopoulos, G. & Stephanopoulos, G., Artificial intelligence in the development and design of biochemical processes. *Trends Biotechnol.*, 1986, 4(9), 241–9.
4. Steels, L., Components of expertise. *AI Magazine*, 1990, 11(2), 28–49.
5. Plaza, E., Aamodt, A., Ram, A., Van de Velde, W. & Van Someren, M., Integrated learning architectures. *Proc. European Conf. on Machine Learning (ECML-93)*. Lecture Notes in Artificial Intelligence 667. Springer-Verlag, Berlin, 1993, pp. 429–41.
6. Marsili-Libelli, S., Optimal control strategies for biological wastewater treatment. In *Environmental Systems Analysis and Management*, ed. S. Rinaldi, North-Holland, Amsterdam, 1982, pp. 279–87.
7. Corder, G. D. & Lee, P. L., Feedforward control of a wastewater plant. *Water Res.*, 1986, 20, 301–9.
8. Dochain, D., Design of adaptive controllers for non-linear stirred tank bioreactors: extension to the MIMO situation. *J. Process Control*, 1991, 1, 41–8.
9. Beck, M. B., Identification, estimation and control of biological waste-water treatment processes. *IEEE Proc.*, 1986, 133, 254–64.
10. Moreno, R., de Prada, C., Lafuente, J., Poch, M. & Montague, G., Non-linear predictive control of dissolved oxygen in the activated sludge process. *ICCAFT 5/IFAC-BIO 2 Conference*, Keystone (CO), USA, 1992.
11. Gall, R. & Patry, G., Knowledge-based system for the diagnosis of an activated sludge plant. In *Dynamic Modelling and Expert Systems in Wastewater Engineering*, ed. G. Patry & D. Chapman. Lewis Publishers, Chelsea, MI, 1989.
12. Krichen, D. J., Wilson, K. D. & Tracy, K. D., Expert systems guide biological phosphorus removal. *Water Environm. Technol.*, 1991, 3(10), 60–4.
13. Serra, P., Sánchez, M., Lafuente, J., Cortés, U. & Poch, M., DEPUR: a knowledge based tool for wastewater treatment plants. *Engng Applic. Artificial Intelligence*, 1994, 7(1), 23–30.
14. Coleman, J. J., Krovvidy, S., Scott Summers, R. & Wee, W. G., An AI approach for wastewater treatment systems. *Appl. Intelligence*, 1991, 1(3), 247–61.
15. Huang, Y. L., Sundar, G. & Fan, L. T., Min-cyanide: an expert system for cyanide waste minimization in electroplating plants. *Environm. Prog.*, 1991, 10(2), 89–95.
16. Chaib-draa, B., Mandiau, R. & Millot, P., Distributed artificial intelligence: an annotated bibliography. *ACM SIGART Bull.*, 1992, 3(3), 20–37.
17. Kirn, S. & Schneider, J., STRICT: Selecting the "Right" architecture. *Proc. 5th Int. Conf. on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE-92)*. Springer-Verlag, Berlin, 1992, pp. 391–400.
18. Bond, A. H. & Gasser, L. (eds), *Readings in Distributed Artificial Intelligence*. Morgan Kaufmann Publishers, San Mateo, CA, 1988.
19. Sánchez, M., R.-Roda, I., Lafuente, J., Cortés, U. & Poch, M., DAI-DEPUR architecture: distributed agents for real-time WWTP supervision and control. *2nd IFAC/IFIP/IMACS Int. Symp. on Artificial Intelligence in Real Time Control (AIRTIC '94)*. València, Spain, 1994, pp. 179–84.
20. Sánchez, M., Cortés, U., R.-Roda, I., Lafuente, J. & Poch, M., Integrating general expert knowledge and specific experimental knowledge in WWTP. *IJCAI '95 Workshop on Artificial Intelligence and the Environment*, Montréal, Canada, 1995, pp. 75–81.
21. Serra, P., Lafuente, J., Moreno, R., de Prada, C. & Poch, M., Development of a real-time expert system for wastewater treatment plants control. *Control Engng Pract.*, 1993, 1(2), 329–35.
22. Béjar, J., Cortés, U. & Domingo, M., Using domain theory to bias classification processes in ill-domains. *Proc. IV Congreso Iberoamericano de Inteligencia Artificial (IBERAMIA-94)*, 1994, pp. 187–99.
23. Riaño, D., Automatic knowledge generation from data in classification domains. Master's thesis, LSI-94 research report, Dept. de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, 1994, (in English).
24. Sánchez, M., Cortés, U., Béjar, J., de Gràcia, J., Lafuente, J. & Poch, M., Concept formation in WWTP by means of

- classification techniques: a compared study. Accepted to appear in *Appl. Intelligence*, 1988, 7(2).
25. Kolodner, J., *Case-Based Reasoning*. Morgan Kaufmann, San Mateo, CA, 1993.
26. Sánchez, M., Cortés, U., R.-Roda, I., Poch, M. & Lafuente, J., Learning and adaptation in WWTP through case-based reasoning. Submitted to a special issue on machine learning of *Microcomp. Civ. Engng*, 1995.
27. Hydromantis, Inc., *GPS-X User's Guide*. Hydromantis Inc., Hamilton, Ontario, Canada, 1995.
28. Dellepiane, S. G., Venturi, G. & Vernazza, G. L., Model generation and model matching of medical images by a fuzzy approach. *Patt. Recogn.*, 1992, 25(2), 115–37.
29. Bañares-Alcántara, R. & Ponton, J. W., Artificial intelligence techniques in chemical engineering process design. In *Applications in Artificial Intelligence in Engineering, VII*, Computational Mechanics Publications, Bath, 1992, pp. 581–607.