DATA MINING AND KNOWLEDGE DISCOVERY FOR MONITORING AND INTELLIGENT CONTROL OF A WASTEWATER TREATMENT PLANT

S. Manesis, V. Deligiannis and M. Koutri
University of Patras,Electrical & Computer Engineering Department, Patras 26500, Greece
{stam.manesis, bdeligiannis, mkoutri}@ece.upatras.gr

Keywords: Intelligent control, Knowledge-based systems, Data mining, Wastewater treatment, Expert-Fuzzy Logic.

Abstract: Intelligent control of medium-scale industrial processes has been applied with success but, as a method of advanced control, can be further improved. Since intelligent control makes use of knowledge-based techniques (such as expert systems, fuzzy logic, neural networks, etc.), a data mining and knowledge discovery subsystem embedded in a control system can support an intelligent controller to achieve a more reliable and robust operation of the controlled process. This paper proposes a combined intelligent control and data mining scheme for monitoring and mainly for controlling a wastewater treatment plant. The intelligent control system is implemented in a programmable logic controller, while the data mining and knowledge discovery system in a personal computer. The entire control system is basically a knowledge-based system which improves drastically the behavior of the wastewater treatment plant.

1 INTRODUCTION

Data mining is a fast growing research field aiming at the extraction of valuable knowledge from massive databases. Due to the increasing use of computing in the context of several applications, data mining can actually be applied to various problems related to the operation of man-made systems and their interactions with other natural ones. These interactions are becoming significantly important as populations are growing and world’s sensitivity for the environment is increasing. An area of particular success has been in data mining for wastewater treatment systems and surface water systems (streams, lakes and rivers), where complex problems can be solved that are unsolvable by any other means (Condras et al. 2002; Condras and Roehl 1999). Data mining tools have been adapted for unsteady continuous systems, as wastewater treatment plants where the hydro-dynamical, biological and physical phenomena are highly coupled, in order to monitor the wastewater quality and detect dangerous faults of the process (Victor Ramos et al. 2004).

On the other hand, *intelligent control* (DeSilva 1995; Harris et al. 1993), the discipline that performs human-like tasks in environments of uncertainty and vagueness with minimal interaction with human operators, has had a significant impact in the process industry. The cement industry was, in fact, the first process industry to apply intelligent control techniques in the late 1970s in the form of fuzzy control, and today hundreds of industrial plants worldwide are controlled by such controllers (Boverie et al. 1991; Jamshidi et al. 1993; King 1992).

A fundamental attribute of intelligent control is its ability to work with symbolic, inexact and vague data which human operators comprehend best. Indeed, its ability to deal with incomplete and ill-defined information, an inherent characteristic of wastewater treatment plants, permits implementation of human-like control strategies which have hitherto defied solution by any of the conventional hard control techniques. Fuzzy logic and artificial neural networks (Harris et al. 1993) are two examples of soft computing which have migrated into the realm of industrial control over the last two decades. Chronologically, fuzzy control was the first and its application in the process industry has led to significant improvements in product quality, productivity and energy consumption. Fuzzy control is now firmly established as one of the leading advanced control techniques in use in industry. Over the last two decades or so wastewater engineering has undergone significant advances in both theory and practice. Experience gained from the operation of numerous wastewater plants, coupled with the results of recent research in the field, has led to improved plant design and wastewater management. Today, effective control of wastewater treatment plants (Rodriguez-Roda et al. 2002; Manesis et al. 1998; Katebi et al. 2000) is of critical importance not only for economic reasons but also to satisfy stringent environmental constraints.

Expert fuzzy control systems have been developed based on human operators' experience, the knowledge of which is acquired by way of extensive interviews. However, the heuristic knowledge of an operator, although it can contain some important consequences about the operational behavior of a plant, can not be based on a large number of measurements and trend diagrams. A data mining and knowledge discovery system can be used to improve or optimize, as well as to evaluate the behavior of the controller. Such a research and development project, called TELEMAC (Lambert 2004; Dixon et al. 2007), within the European IST program is now working on data mining which opens up the prospect of learning from data in order to manage wastewater treatment plants better. Classification techniques for concept acquisition have been also applied in order to build knowledge bases that can help human experts to manage wastewater treatment plants (Serra et al. 1994). Usually, a Data Mining and Knowledge Discovery (DMKD) system (Huang and Wang 1999; Sanguesa et al. 1997; Dixon et al. 2004; Dixon et al. 2007) for process monitoring and control is based on simple measurements of the controlled variables from which association rules may be found. As a valuable addition, a DMKD system based on both measurements and actions of a fuzzy controller is the main idea presented in this work. Particularly, the paper describes a new hybrid scheme in which the induction rules are a priori given but modifiable while the DMKD system searches on both inputs and outputs of the fuzzy controller, records the activity of rules and hence constitutes a part of the overall controller. Experimental results from a four months operation period of the treatment process are presented in last section.

2 WASTEWATER TREATMENT

Wastewater treatment plants typically have two principal stages as shown in Fig.1, the *primary* stage which includes the bar racks, grit chamber and primary settling tank whose objective is the removal of the organic load and solids in the wastewater to a degree of 30-50% and the *secondary* stage whose objective is the biological treatment of the organic load.

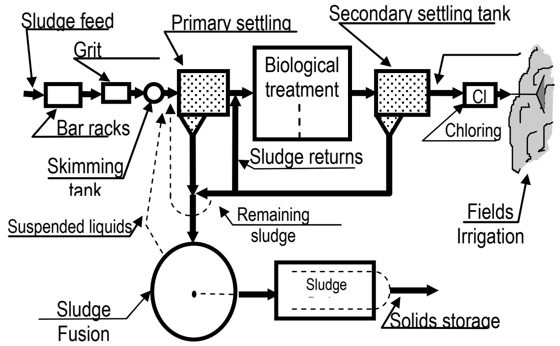

Figure 1: Schematic of a typical wastewater treatment plant.

The removal of organic load (biochemical oxygen demand or BOD, mixed liquid suspended solids or MLSS) in conjunction with secondary treatment performed in the final stage, leads to an overall treatment level of the order of 80-90%. In all wastewater treatment plants it is necessary the oxygen content in the aerated zone to be also subjected to close control. This is achieved by a suitable control strategy involving the following three *manipulated* variables:

(1) the oxygen supply to the aerated zone (O_2Feed)
(2) the mixed liquid returns rate from aerated zone to the anoxic one (R_ml)
(3) the sludge returns rate from settling tank to the biological reactor (R_Sludge).

The quantities which are appropriately measured by suitable instrumentation and constitute the controlled variables of the plant are:

(1) the ammonia concentration in the reactor ($N\text{-}NH_3$)
(2) the nitrate concentration in the reactor ($N\text{-}NO_3$)
(3) the dissolved oxygen in the reactor (DO)
(4) the temperature in the reactor ($TEMP$)
(5) the mixed liquid suspended solids concentration in the reactor ($MLSS$)
(6) the difference in biochemical oxygen demand between the entrance and exit of the secondary settling tank ($D(BOD)$)

These six variables constitute therefore the inputs to the intelligent controller. By the nature of the process and the interaction of the controlled variables it is obvious that effective control can only be achieved by means of a multivariable controller behind of which complex knowledge must exist.

3 INTELLIGENT CONTROL OF A WASTEWATER TREATMENT PLANT

Having established the principal controlled and manipulated variables, the next task in developing an intelligent controller for a wastewater treatment plant, using linguistic techniques, is to establish a set of linguistic descriptors for each manipulated variable. These are expressions of the type *Very high, High, Low, OK* etc. which are commonly used by plant operators. The *integrity* of an intelligent controller is directly related to the number of such descriptors, bur practical limitations place a limit on this number. The *granularity* of the controller is inversely proportional to the number of linguistic descriptors. Three descriptors are generally sufficient to describe the controller input variables, the *HI* (High), *OK* and *LO* (Low) descriptors all with trapezoidal membership functions. The locations of the centroids of the membership functions can be considered as the modal points of the fuzzy resolution while the number of such modal points corresponds to the number of fuzzy states of the variable. Also, the intermodal spacing of the membership functions is a measure of the resolution of the variable. It is obvious that the overall accuracy of the intelligent control system is directly related to this resolution. In a similar manner, the manipulated variables or controller outputs are allocated by five descriptors *VH* (VeryHigh), *HI* (High), *OK*, *LO* (Low) and *VL* (VeryLow) which provide sufficient fineness of control. For computational simplicity, singletons provide a convenient way to describe the membership functions of the controller outputs where high accuracy is not of paramount importance and also lead to a particularly simple arithmetic procedure for defuzzification.

The inference engine of the intelligent controller manipulates linguistic control rules of the form:

R: if ((*D(BOD)* is Y_1) and (*MLSS* is Y_2) and (*TEMP* is Y_3 and (*DO* is Y_4) and (*N-NH$_3$* is Y_5) and (*N-NO$_3$* is Y_6)) then ((*O$_2$Feed* is U_1) and (*R_Sludge* is U_2) and (*R_ml* is U_3))

where Y_m and U_n are the linguistic descriptors of the *m* controller inputs and *n* outputs respectively, where m ∈ {1,2,3,4,5,6} and n ∈ {1, 2, 3}. For the *k*th linguistic rule, the values of the membership functions corresponding to the process outputs (i.e. the controller inputs) are computed to form the array

$$\mu_1^k(D(BOD)),\ \mu_2^k(MLSS)\ldots\mu_6^k(N-NO_3)$$

the minimum element of which is the *degree of fulfillment* of that rule and is a measure of the contribution of that rule to the final control action, i.e.

$$\sigma_k = \min\left\{\mu_1^k(D(BOD)), \mu_2^k(MLSS)\ldots\mu_6^k(N-NO_3)\right\} \in [0,1]$$

The union of the weighted products of the corresponding membership functions of the controller output fuzzy sets $\{v(\cdot)\}$ is subsequently computed to form the resultant output membership functions. The membership function of the jth controller output is thus the result of the *max* operator:

$$\max\left\{\sigma_1 v_1^j(O_2Feed), \sigma_2 v_2^j(R\_Sludge), \sigma_3 v_3^j(R\_ml)\right\}$$

The engineering values of the controller outputs, necessary to drive the actuators of the plant under control are obtained following the defuzzification controller outputs are described by *p* singletons the centre of gravity *(COG)* of the jth controller output is simply the inner product procedure. When the membership functions of the:

$$y_j = \left\langle \varphi_j, z_j \right\rangle$$

where the coefficients

$$\varphi_j = \frac{\sigma_j}{\sum_{l=1}^{p} \sigma_l} \in [0,1]$$

are the *fractional degrees of fulfillment* while the array $\{z_j\}$ contains the locations of the singletons of the membership functions of the outputs.

A fuzzy system shell (FuzzyControl++ S7 ® by Siemens) was used to develop the intelligent controller of the wastewater treatment plant for a city of 120.000 PE in Greece. The shell uses trapezoidal membership functions, Mamdani *max-min* inference and *COG* defuzzification (Mamdani 1974; Patyra and Mylnak 1996). A Simatic S7-300 programmable logic controller equipped with digital and analogue input-output cards, to which the plant sensors and actuators are directly connected, has been selected. This actual controller is linked to a host computer necessary not only for the fuzzy system shell implementation but also for the data mining and knowledge discovery procedure which is described in next section.

4 DATA MINING AND KNOWLEDGE DISCOVERY FOR INTELLIGENT CONTROL

With intelligent controller being operated as described above, the controller accepts a stream of data and decides about the actions on the manipulated variables. The generated continuously over time actions, together with the operational conditions constitute a kind of knowledge, remarkably accepted or not, which nobody exploits. Hence, a DMKD system should be introduced in order to acquire and evaluate this knowledge. It is known that the operational data of any industrial process are used by plant operators and supervisors to develop an understanding of a plant operation through interpretation and analysis. In a first level, there are methodologies and tools that automate data interpretation and analysis derived by a large number of measurements. In industrial processes which already operate with an expert-fuzzy rule-based controller the requirements of this first level have been realized. Human operators are more concerned with the current status of the process and possible future behavior rather than the current values of individual variables. Apart from simple measurements, we need in a second level a furthermore analysis of the dataset consisting of the combined output actions and input measurements, which only the expert fuzzy controller can give us during its operation. Therefore, we propose the analysis of the set of the fuzzy control states with the use of DMKD techniques. *Data mining* refers to the extraction of interesting patterns from large amounts of data, while involves the use of techniques from multiple disciplines such as database technology, statistics, machine learning, neural networks, information retrieval, etc. *Knowledge Discovery in Databases* (KDD) concerns a systematic process consisting of a set of well-defined steps. Data mining constitutes a step in the whole knowledge discovery process (Comas et al. 2001; Gibert et al. 2005). The processes controlled by expert-fuzzy controllers are usually slow procedures and hence the DMKD process can operate off-line allowing the human confirmation. The data mining algorithmic procedure will operate as a computational component at the control center and will run periodically to discover knowledge and update the knowledge base.

4.1 Problem Ddefinition and Rrepresentation

Given a wastewater treatment plant along with the corresponding intelligent controller, we define the fuzzy control state as the vector

$$((m_1, m_2, m_3, m_4, m_5, m_6), (u_1, u_2, u_3), t_k)$$

where m_i, $i=1,\ldots,6$ are the measurements of the inputs for which the fuzzy controller decides the actions u_j, $j=1,\ldots,3$ for the corresponding outputs at the time stamp t_k, $k= 1, 2\ \ldots, \infty$. According to this description, a set of fuzzy control states could be stored in a relational database on which data mining could be performed. A relational database is a collection of tables, each of which is properly named. In our case there is one table representing the set of fuzzy states (see Table 1). As an example, the fuzzy state s_1 is defined as

$$s_{1:}\ ((HI, OK, OK, LO, OK, OK), (VH, OK, LO), 2)$$

The table of fuzzy control states results after removing the erroneous data and integrating information from various data sources, in particular from the measurements and actions, as well as from the set of fuzzy rules. Particularly, the rule activity and the degree of rule fulfillment are also stored in database. This process constitutes the first step of the overall knowledge discovery process. The second step concerns the data selection and transformation into meaningful representations. In our case, the time stamp may partially eliminated, because we are interesting in the behavior of the system during a period of time rather than a particular time instance. For example, weather conditions for certain periods of time may affect the plant operation. Changes in rainfall, temperature and humidity must be recorded with time stamp and correlated with the rest of mined data.

Table 1: Fragment of records from the database for fuzzy control states

	m_1	m_2	m_3	m_4	m_5	m_6	u_1	u_2	u_3	t
s_1	HI	OK	OK	LO	OK	OK	VH	OK	LO	2
s_2	HI	OK	HI	LO	OK	LO	VH	LO	LO	3
s_3										
…										

4.2 Mining Interesting Patterns

Data mining methods and techniques could be applied into the set of cleaned records, in order to discovering interesting patterns. *Association rule mining*, *clustering*, and *classification* seem to better serve the needs of the particular problem of controlling a plant. Association rule mining aims at finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories (Vazirgiannis et al. 2000). An association rule is a statement of the form $A \Rightarrow B$, where A and B are disjoint subsets of a set of items. The rule is accompanied by two meaningful measures, *confidence* and *support*. Confidence measures the percentage of transactions containing A that also contain B (i.e. *confidence* $(A \Rightarrow B) = P(B \backslash A)$). Similarly, support measures the percentage of transactions that contain A or B (i.e. *support* $(A \Rightarrow B) = P(B \cup A)$) (Han and Kamber 2001).

The application of association rule mining algorithms in the relational database of the control states of the wastewater treatment plant could produce interesting results. If the derived set of association rules contains some rules not already encountered into the input data set, then these new rules could be embedded into the inference mechanism of the expert-fuzzy controller. Techniques of rule induction are also being applied to estimate values of sensors readings based on more easily obtained values, and to determine how reliable the models remain over time. Rules may be generated in forms such as the following,

> *Variable (DO) falls in a particular range of low values if variable D(BOD) falls within a particular range of high values*

accompanied by an indication of the degree of satisfaction of the rule.

Subsequently, the clustering concerns the process of grouping a set of physical or abstract objects into classes of similar objects. So, a cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters (Han and Kamber 2001; Gibert et al. 2005). The application of clustering into a set of data objects requires that the data objects are not class-labeled. In our specific problem a state, consider for example s_1 (see Table 1), constitutes a data object. This data object is labeled by the values of the three output variables, u_1, u_2, and u_3. The unlabeled data object corresponding to s_1 is:

s_1: *(HI, OK, OK, LO, OK, OK)*

The application of clustering algorithms to the set of unlabeled states could conduct to the identification of states of similar behavior and, thus, to the derivation of theoretical and more general rules.

Lastly, classification could be considered as a function that maps (classifies) a data object into one of the several predefined classes (Vazirgiannis et al. 2000). That means (i) a well-defined set of classes, and (ii) a set of pre-classified data objects are required. Consequently, classification is a two-step process: learning and classification. During the learning sub-process, the set of labeled data objects is analyzed using a classification algorithm and a classification model is derived. Next, during the classification sub-process, the model can be applied to the new unlabeled data objects for inferring their classes. To be clearer, in our case, a classification algorithm is firstly applied on the set of cleaned control states in order to find a classification model. Next, whenever a new unlabeled state appears, the model is applied for classifying this new state; in other words for deciding the actions u_1, u_2, and u_3. Classification is, thus, another way for creating new rules, for testing and modifying the existing ones.

4.3 Data Mining Tools

The selection of a commercial data mining tool depends on various similar parameters, such as:

- system issues (like operating system, client-server architecture, etc.)
- support of different types of data sources (ASCII files, relational databases, ODBC connections)
- support of various data mining algorithms
- visualization of the resulted patterns
- price
- ease of learning to use

For the wastewater treatment plant application described above and particularly for a plant of medium size we need a stand-alone PC architecture with windows operating system, to support only flat files with numerical data and various mining algorithms while visualization in not necessary. Some examples of data mining tools are IBM Intelligent Miner, SGI MineSet, Clementine (SPSS) and GESCONDA (Gibert et al. 2005), from which the first one has been selected to implement the current project.

5 EXPERIMENTAL RESULTS

In this work, we used the time series data of four months (April 2006 to July 2006) measured or recorded at the medium-size wastewater treatment plant mentioned above. As one example for all of these data records the measurements in the influent D(BOD), N-NO_3 and DO are shown in Fig.2. Zooming into this plant's monitoring in order to achieve better resolution the corresponding measurements are shown in Fig.3 and Fig.4 in a two-day and one-day section for D(BOD) and N-NO_3 concentrations respectively. The set of acquired data includes also the actions of the fuzzy controller which are recorded in the database by means of rules activity. Figs.5, 6, 7 show the membership functions and the input value appearance frequency for the D(BOD), N-NO_3 and DO respectively. The

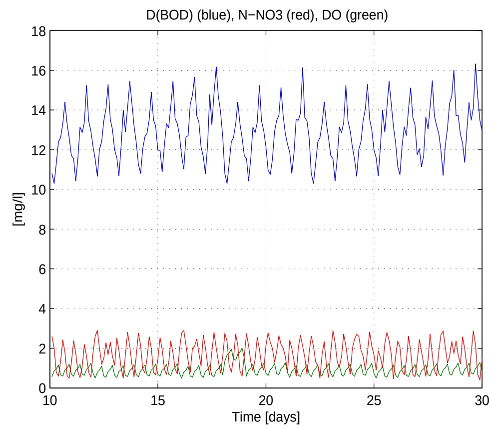

Figure 2: Measured D(BOD), N-NO3 and DO concentrations during the period 10/6 – 30/6 2006.

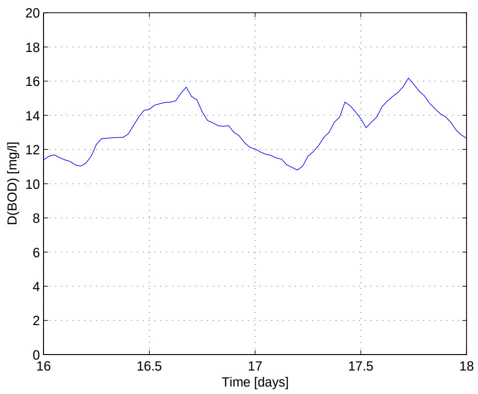

Figure 3: Two-day section of D(BOD) concentration from the four months data (16/6-18/6 2006).

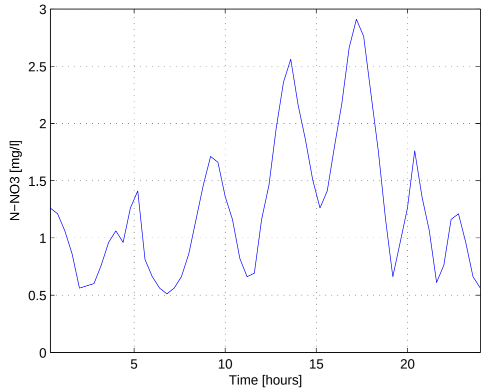

Figure 4: One-day section of N-NO_3 concentration from the four months data.

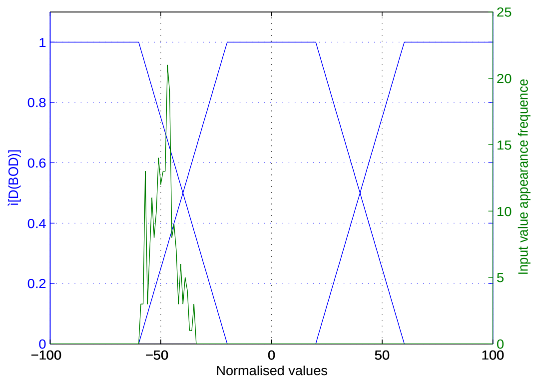

Figure 5: Membership functions and sampled measurements for D(BOD).

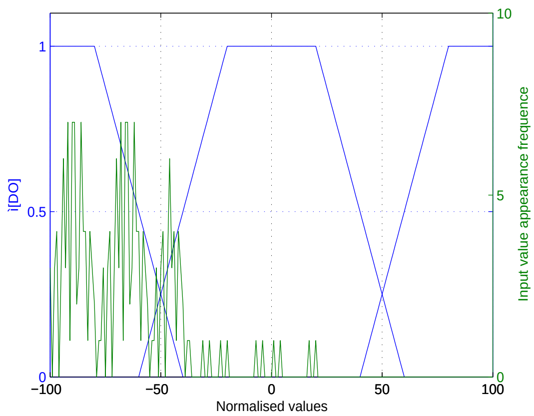

Figure 6: Membership functions and sampled measurements for DO.

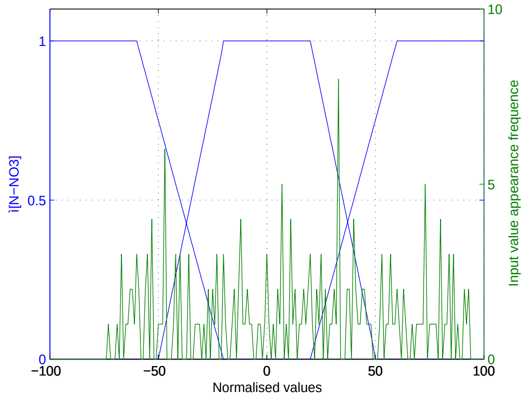

Figure 7: Membership functions and sampled measurements for N-NO_3.

The values in these three figures are normalized into the domain [-100, 100] since the six inputs variables get values in different domains and hence uniformity is needed. By inspection of Fig.5 one can deduce that the most of the reformatted measurements belong partially in both LO and OK membership functions. This means that a reconfiguration of the defined membership functions is required in order to have a set of membership functions which will cover effectively the actual values of D(BOD) in the real process. The narrow distribution of the sampled measurements into the heuristically predefined membership functions indicates furthermore that reconfiguration is necessary. Another conclusion concerns the small variation of D(BOD) measurements in comparison to the corresponding predefined range of D(BOD) values and therefore has to be reconsidered. Fig.6 shows a more uniform distribution of the sampled measurements of DO, while the best distribution of the obtained measurements is depicted in Fig.7 concerning the N-NO_3 variation.

5.1 Rule Activity-Modification

As mentioned above, the second step of the overall knowledge discovery process concerns the data selection based on statistical criteria and the transformation of them into meaningful representations. As a first approach to obtain cleaned data for knowledge creation, the rule activity has been recorded in DMKD off-line module and some results are shown in Fig.8 and Fig.9. Each figure has five subplots for equal number of rules. For each rule, the number of appearances is shown as a function of the degree fulfillment. For example, rule No.5 was satisfied three times with a 20% degree of fulfillment. Fig.8 indicates the least active rules while Fig.9 shows the corresponding most active ones.

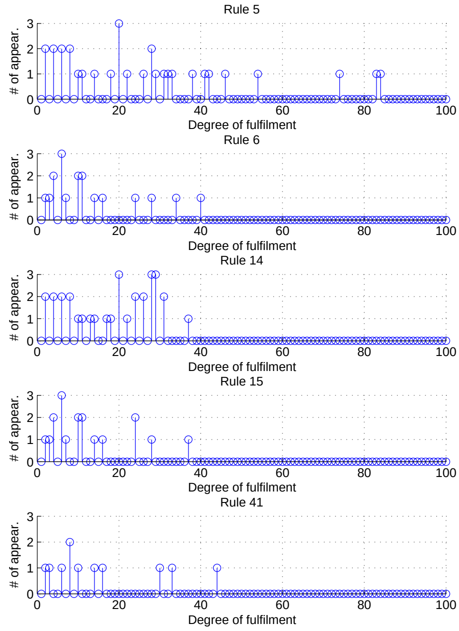

Figure 8: Statistical analysis of the least-active rules behavior.

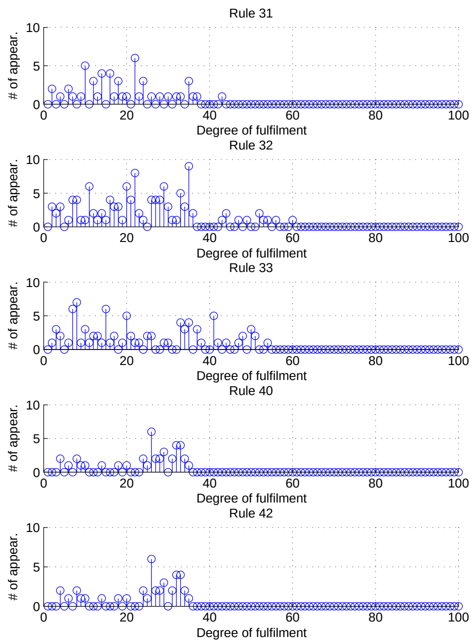

Figure 9: Statistical analysis of the most-active rules behavior.

6 CONCLUSIONS

Although wastewater treatment plants are implemented with properly functioning control loops concerning the biological process, in practice, this type of plant requires a major time investment on the side of the operator, involving many manual operations. These difficulties can be overcome by an intelligent controller which incorporates the human experience. The mined data, characterized as multivariate and interrelated, constitutes a combination of measurements of the process's variables and actions of the controller. The consequences of the mining and knowledge discovery procedure are used to adapt the soft structure of the intelligent controller through a semi-automatic scheme that provides deeper understanding and better operation of the controlled plant. The experimental results give us basic directions to improve the operation of the control system but it is obvious that a longer validation period of data monitoring and processing is needed.

REFERENCES

Boverie, S, Demaya, B and Titli, A, 1991, Fuzzy logic control compared with other automatic control approaches, *Proceedings of 30th CDC Conference*, Brighton.

Comas, J., Dzeroski, S., Gibert, K., Roda, I.R. and Sanchez-Marre M., 2001. Knowledge discovery by means of inductive methods in wastewater treatment plant data", *AI Communications*, Vol.14(1), pp.45-62.

Condras, P, Cook, J and Roehl E, 2002, "Estimation of Tidal Marsh Loading Effects in a Complex Estuary" *American Water Resources Association Annual Conference*, New Orleans.

Condras, P and Roehl, E, 1999, Real-time control for matching wastewater discharges to the assimilative capacity of a complex tidally affected river basin *South Carolina Environmental Conference*, Myrtle Beach.

DeSilva, C W, 1995, *Intelligent Control, Fuzzy Logic Applications*, CRC Press, Boca Raton.

Dixon, M., Gallop, J.R., Lambert, S.C., Lardon L., Steyer P. and Healy J.V., 2004. Data mining to support anaerobic WWT plant monitoring and control, *Proceedings of the IFAC Workshop on Modelling and Control for Participatory Planning and Managing Water Systems*, Venice Italy,.

Dixon, M., Gallop, J.R., Lambert S.C. and Healy, J.V., 2007. Experience with data mining for the anaerobic wastewater treatment process, *Environmental Modelling & Software*, Elsevier, , vl.22, pp.315-322.

Gibert, K., Sanchez-Marre, M. and Flores X., 2005. Cluster discovery in environmental databases using GESCONDA: The added value of comparisons, *AI Communications,* Vol.18(4), pp.319-331.

Han, J and Kamber, M, 2001, *Data Mining: Concepts and Techniques*, San Francisco, Morgan Kaufmann.

Harris, CJ, Moore, CG and Brown, M, 1993, Intelligent control-aspects of fuzzy logic and neural nets, NJ, World Scientific Publishers.

Huang, YC and Wang, XZ, 1999, Application of fuzzy causal networks to wastewater treatment plants, *Chemical Engineering Science* 54:2731-2738.

Jamshidi, M, Vadiee, N and Ross, TJ, 1993, *Fuzzy Logic and Control–Intelligentware and Hardware Applications*, NJ, Prentice Hall.

Katebi, R, Johnson, M and Wilkie, J, 2000, The future of Advanced Control in Wastewater Treatment Plants, *Proceedings of the CIWEM Millennium Conference* Leeds UK.

King, R, 1992, Expert supervision and control of a large-scale plant, *Journal Intelligent and Robotic Systems*, 2(3).

Lambert, S, 2004, (CCLRC-Data mining Resp.) Telemonitoring and advanced tele-control of high yield wastewater treatment plants, TELEMAC, R&D project, No. IST-28156.

Mamdani, EH, 1974, An application of fuzzy algorithms for the control of a dynamic plant, *Proceedings IEE* , 121(12).

Manesis, S, Sapidis DJ and King RE, 1998, Intelligent Control of Wastewater treatment Plants, *Journal Artificial Intelligence in Engineering* (12):275-281.

Patyra, MJ and Mylnak, DM, 1996, *Fuzzy Logic Implementation and Application*, NJ, J. Wiley-Teubner.

Rodriguez-Roda, I, Sanchez-Marre, M, Comas, J, Baeza, J, Colprim, J, Lafuente, J, Cortes, U and Poch M, 2002, A Hybrid Supervisory System to Support WWTP Operation: Implementation and Validation, *Water Science and Technology Journal* (45): 289-297.

Sanguesa, R, Cortés, U and Béjar, J, 1997, Causal dependency discovery with posscause: an application to wastewater treatment plants, *Proceedings of the 1st International Conference on the practical application of knowledge discovery and data mining*, London.

Serra, P, Sànchez, M, Lafuente, J, Cortès, U and Poch, M, 1994, DEPUR: a knowledge-based tool for wastewater treatment plants, *Engineering Applications of Artificial Intelligence*, 7(1):23-30.

Vazirgiannis, M and Halkidi, M, 2000, Data Mining: Concepts and Techniques, Athens University of Economics and Business, TR HELDiNET 10.

Victor Ramos, J, Goncalves, C and Durado, A, 2004, On-line Extraction of Fuzzy Rules in a Wastewater Treatment Plant, *Proceedings of the 1st International Conference on Artificial Intelligence Applications and Innovations*, 1:87-102.