Fuzzy pattern recognition of circadian cycles in ecosystems

S. Marsili-Libelli*

Department of Systems and Computers, University of Florence, Via S. Marta 3, 50139 Firenze, Italy

Abstract

Many ecological variables show a wide range of fluctuations, the most important of which is the diurnal variation. This cycling may contain important information regarding the ecosystem's functioning and, if properly interpreted, can represent a valuable predictive tool in ecosystems management. This paper describes a simple algorithm for extracting meaningful information from daily cycles using fuzzy pattern recognition techniques. The algorithm is organised in three parts: in the first, typical patterns are extracted from experimental data to form the knowledge-base upon which the algorithm operates. The second step is to condense the information contained in the knowledge-base into mathematical objects, referred to in the paper as *fuzzy masks*. The third step is the set-up of an inferential set of fuzzy rules, using the fuzzy masks as antecedents. Depending on how the inference engine is structured, the algorithm output can be viewed as an assessment of the current daily cycle with respect to a set of "typical" system behaviours and decisions can be made accordingly. Two applications are presented to demonstrate the algorithm. In the first a medium-scale wastewater treatment plant with minimum instrumentation is considered. Based on diurnal flow variations of the last 2 days, the algorithm determines the current treatment level in terms of sludge recycling. The other application is the prediction of the eutrophication level in a coastal lagoon. In this case, the algorithm considers the diurnal fluctuations of basic water quality parameters such as dissolved oxygen, oxidation–reduction potential, pH and temperature to detect the possibility of a macroalgae bloom. In both cases, the algorithm provides information about the shift of the system from one trend to another on the basis of the analysis of diurnal cycles. The algorithm performance is assessed against practical situations and the results discussed.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Pattern recognition; Decision support systems; Fuzzy systems; Eutrophication; Phytoplankton; Wastewater treatment; Lagoons

1. Introduction

Ecosystems are subject to fluctuations with a wide range of periods, from the short-term random variations to daily cycles and then on to seasonal changes. Of all these, the diurnal cycle is the most important, being related to the day–night sequence which drives most ecological process and, as a consequence, many human activities. Information contained in the diurnal cycle and its variations is often crucial to detect important changes in the system, and a simple algorithm is presented here to analyse diurnal patterns and extract information about the future state of the system.

The algorithm is based on the classification of daily patterns with respect to a set of typical behaviours (*prototypes*), whose characteristics are embedded in *fuzzy masks*, mathematical objects combining information about the statistical fluctuation and the typical daily trend of the environmental variables. These fuzzy masks are then used as the antecedents in a fuzzy inferential engine, whose output is the required prediction. The derivation of the fuzzy masks and their use in the fuzzy engine is described with the aid of case studies in Sections 3 and 4.4, together with the defuzzification procedure to obtain the required forecast.

* Corresponding author. Tel.: +39-055-47-96-264.
E-mail address: marsili@ingfi1.ing.unifi.it (S. Marsili-Libelli).

Two applications of the algorithm are presented: the first regards a simple decision support system to adjust the treatment level of a medium-sized biological wastewater treatment plant, whereas the second is concerned with the forecast of eutrophic situations in a coastal lagoon.

There is a vast literature regarding predictive tools for ecosystems, a thorough survey of which can be found in Håkanson (1999), whereas the general principles of machine learning applied to ecosystems are well described by Recknagel (2001). The occurrence of algal blooms is a much-studied aspect of shallow water bodies. In many cases neural networks have been used as predictive tools, less frequently, fuzzy algorithms. Freshwater environments appear to be the most widely studied ecosystems, whereas the only neural network studies applied to saltwater eutrophic areas are due to Scardi and Harding (1999) and Scardi (2001), who adopted a very general approach in developing two neural network models with a multilayer perceptron structure for estimating primary production in Chesapeake Bay, an enriched estuary in the eastern United States. A similar approach, but in a freshwater context, was followed by Jeong et al. (2001) and by Yabunaka et al. (1997) who used 10 environmental parameters as network inputs to predict the concentration of five freshwater phytoplankton species. Likewise, Wei et al. (2001) included eight environmental factors to estimate the evolution of four dominant phytoplankton genera. This study allowed species discrimination with respect to nitrogen uptake.

The fuzzy literature is much more limited. Ganoulis (1994) outlined the application of fuzzy systems and probability to the forecasting of ecological variables. Bobbin and Recknagel (2001) considered patterns in the chemical and physical parameters of a lake and the corresponding presence or absence of blue-green algae such as *Microcystis*, *Oscillatoria* and *Phormidium*. Maier et al. (1998, 2001) have applied a mixed neuro-fuzzy technique to the forecasting of *Anabaena* spp. in a river system, while Shen and Chouchoulas (2001) estimated the size of algal populations by a hybrid system involving a combination of Fuzzy Set and Rough Set theories. Finally, Mpimpas et al. (2001) recast a two-dimensional water quality problem using a classical fuzzy approach derived from Ganoulis (1994).

2. A fuzzy pattern recognition tool for diurnal cycles

The basic scheme of the algorithm will now be illustrated in general terms. The starting idea is to extract and classify a set of relevant daily cycles from the observation of a large number of instances. Once an exhaustive collection of basic patterns is established and significant behaviours are extracted, this knowledge-base is embedded in the fuzzy masks (mathematical objects that will be defined in Section 2.3). A fuzzy inferential system is then constructed to compare any new daily pattern with the knowledge-base represented by the fuzzy masks, and to make appropriate decisions. The inferential part is strongly problem-dependent and will be examined on a case-by-case basis. By contrast, the procedure to determine the base patterns, extract the knowledge-base in terms of fuzzy masks and connect this information to the inferential engine is completely general and is now described.

2.1. Analysis of daily patterns

Diurnal variations of environmental variables carry fundamental information, which is often concealed by random short-term variations. The first step of the algorithm is therefore to filter out the non-informative disturbances and classify the circadian information into a set of significant patterns. A fuzzy approach will be used, given the inherent ability of the fuzzy inference mechanism (Yager and Filev, 1994; Bárdossy and Duckstein, 1995) to implement approximate reasoning and produce similarity measures over imprecisely defined patterns (Bezdek, 1981). An example of a repeating daily pattern is the daily cycle of dissolved oxygen in a eutrophic water body, as will be shown in Section 4 (see e.g. Figs. 11 and 12). To extract this information, the data are first grouped into sets of daily patterns with differing significance. This partition is a crucial stage of the algorithm and is largely problem-dependent. Therefore it is discussed in greater details in the sections describing specific applications.

As a general remark, care should be taken to obtain a complete set of patterns, covering all the meaningful behaviours of the ecosystem. These are then translated into numerical objects, hereby termed fuzzy masks, forming the knowledge-base on which the

fuzzy decision system is constructed. Fuzzy masks are the result of two procedures: statistical processing, to account for the day-to-day data variability, and generalisation of the underlying circadian cycle using *fuzzy curves*.

2.2. Statistical analysis of the daily patterns

If the circadian cycle of an ecological variable $y(t)$, $t \in (0, 24)$ is observed over a set of many days, the values at any given hour $k = 0, \ldots, 23$ form a distribution from which a histogram can be obtained. A Gaussian distribution is then fitted to each of the 24 histograms producing a vector of hourly means and variances for each set. The statistical information represented by the estimated variance $\sigma^2(k)$ of the fitted Gaussian indicates the "spread" of the data over each hour of the day and will become part of the knowledge-base.

2.3. Construction of the fuzzy masks

Instead of using the hourly average to represent the circadian cycle $y(t)$, $t \in (0, 24)$, a fuzzy curve is used for its properties of superior smoothing and generality. Fuzzy curves were introduced by Lin and Cunningham (1995) and Lin et al. (1996) as a way to approximate a set of data with a minimum number of fuzzy rules. In this context, fuzzy curves are used to smooth the pattern and to take into account the hourly correlations over the diurnal pattern. They are defined on a fuzzy radial basis function (FRBF), which is a set of Gaussian functions $h_k(t) \to (0, 1)|k = 0, \ldots, 23$:

$$h_k(t) = \exp\left(-\frac{(t-k)^2}{2b^2}\right), \tag{1}$$

where $t \in (0, 23) \in \mathbf{R}$ is the time of day, $k \in (0, 1, \ldots, 23) \in \mathbf{I}$ is the hour set, so that $h_k(k) = 1$, and b is a parameter controlling the spread of the membership function (1). Usually, this parameter is arbitrarily chosen between 10 and 20% of the variable range (Lin and Cunningham, 1995), but in this case the partial correlation function of the residuals between the smoothed curve and the actual data is used to compute b from a simple first-order autoregressive model. Let ρ be the first sample (lag = 1 h) of the partial autocorrelation function. Then from Eq. (1),

the relation holds

$$\rho = \exp\left(-\frac{1}{2b^2}\right). \tag{2}$$

Solving for b yields

$$b = \frac{1}{\sqrt{2\ln(1/\rho)}}. \tag{3}$$

Once the FRBF (1) is established, the fuzzy curve can be obtained as the weighted sum of the FRBF over the 24-h period:

$$c(k) = \frac{\sum_{k=0}^{23} y(k) \times h_k(k)}{\sum_{k=0}^{23} h_k(k)}, \tag{4}$$

where $\{y(k)\, k = 0, \ldots, 23\}$ is a set of hourly samples of the circadian cycle.

In summary, the fuzzy masks $P_{i,j} \equiv \{c_{i,j}(k), \sigma_{i,j}^2(k)\ |k = 0, \ldots, 23\}$, embedding the features of *typical* daily patterns in terms of hourly variances σ_k^2 and the fuzzy curve $c(k)$, are constructed as shown in Fig. 1 from representative data sets.

2.4. From fuzzy masks to fuzzy reasoning

The idea of applying fuzzy pattern recognition techniques to environmental problems is not new (see e.g. Ganoulis, 1994; Bárdossy and Duckstein, 1995; Marsili-Libelli and Müller, 1996), but a new aspect of the present work is the fact that comparisons are made between complex objects such as daily patterns, rather than simple numerical values. This is accomplished by embedding the features of the daily patterns into the fuzzy masks and using them as the antecedents of a rule based fuzzy inferential engine. If the problem involves n environmental variables $\{v_1, v_2, \ldots, v_n\}$ and for each of these, m fuzzy masks $\{P_{i,1}, P_{i,2}, \ldots, P_{i,m}|i = 1, \ldots, n\}$ are defined, m fuzzy inferential rules can be defined, where the antecedents are the degree of truth (DOT) of each variable with respect to its fuzzy masks, namely

$$R_j : \quad \text{IF } v_1 \text{ is } P_{1,j} \text{ AND } v_2 \text{ is } P_{2,j} \text{ AND} \cdots$$
$$\text{AND } v_n \text{ is } P_{n,j} \text{ THEN } y_j = Y_j, \quad j = 1, \ldots, m. \tag{5}$$

The set of Eq. (5) is a classical Sugeno fuzzy inferential system (Takagi and Sugeno, 1985; Yager and
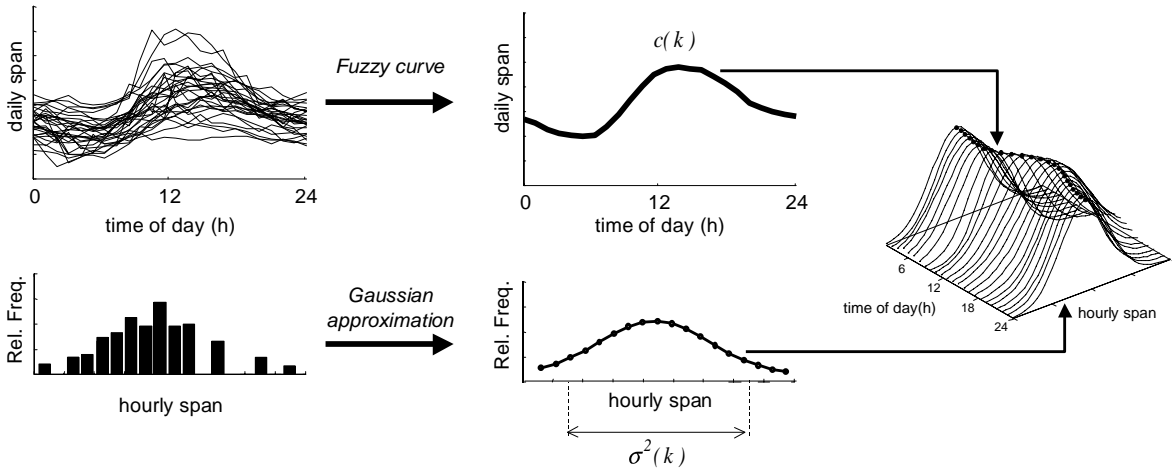
Fig. 1. Construction of a fuzzy mask from the fuzzy curve and the Gaussian approximation.

Filev, 1994) where the consequents Y_j are singletons (i.e. plain numerical values) which can be either specified by the user or estimated from experimental data. The output y is then determined through defuzzification. Practical implementation of Eq. (5) requires that each predicate v_i is $P_{i,j}$ is replaced by its corresponding DOT μ_j^i. This is obtained through the fuzzification of the variable $v_i(k)$ with a Gaussian membership function defined by the fuzzy masks parameters $c_{i,j}(k)$ and $\sigma_{i,j}^2(k)$:

$$\mu_j^i(k) = \exp\left(-\frac{(v_i(k) - c_{i,j}(k))^2}{2 \times \sigma_{i,j}^2(k)}\right),$$
$$j = 1, \ldots, m, \ i = 1, \ldots, n, \ k = 0, \ldots, 23. \quad (6)$$

Since both the fuzzy masks and the variables $v_i(k)$ consist of 24 hourly data, Eq. (6) yields an equal number of DOTs. This is rather redundant information and a single DOT for the daily pattern can be obtained by averaging the hourly DOTs:

$$\bar{\mu}_j^i = \frac{1}{24}\sum_{k=0}^{23}\mu_j^i(k), \quad j = 1, \ldots, m, \ i = 1, \ldots, n \quad (7)$$

if a "prudent" inferential engine is required. A more discriminatory choice might be obtained, for example, by selecting the highest DOT.

These daily DOTs form the antecedents of the inferential equations (5), which can then be written as

$$R_j: \quad \bar{\mu}_j^1 \wedge \bar{\mu}_j^2 \wedge \cdots \wedge \bar{\mu}_j^n \wedge Y_j, \quad j = 1, \ldots, m. \quad (8)$$

The notation $\wedge$ in Eq. (8) applies to both the AND and THEN connectives in Eq. (5) which, though conceptually distinct, represent logical implications and are implemented with the *product*, an admissible *t*-norm operator (Yager and Filev, 1994; Babuska, 1998). The algorithm output y is then obtained by defuzzification as the weighted average of the singletons Y_j through their combined antecedent DOTs $\mu_j = \bar{\mu}_j^1 \times \bar{\mu}_j^2 \times \cdots \times \bar{\mu}_j^n$, namely

$$y = \frac{\sum_{j=1}^m \mu_j Y_j}{\sum_{j=1}^m \mu_j}. \quad (9)$$

Fig. 2 shows the complete fuzzy inference scheme: in summary, the fuzzy masks are used to obtain the degree of truth of the antecedents in a Sugeno inference scheme, whose defuzzification can be used to assess the meaning of the observed circadian cycle and its variations.

To illustrate the features of the algorithm, two examples are now presented: the first considers a medium-size activated sludge wastewater treatment process where the treatment level for the next day is determined based on the flow pattern of the two
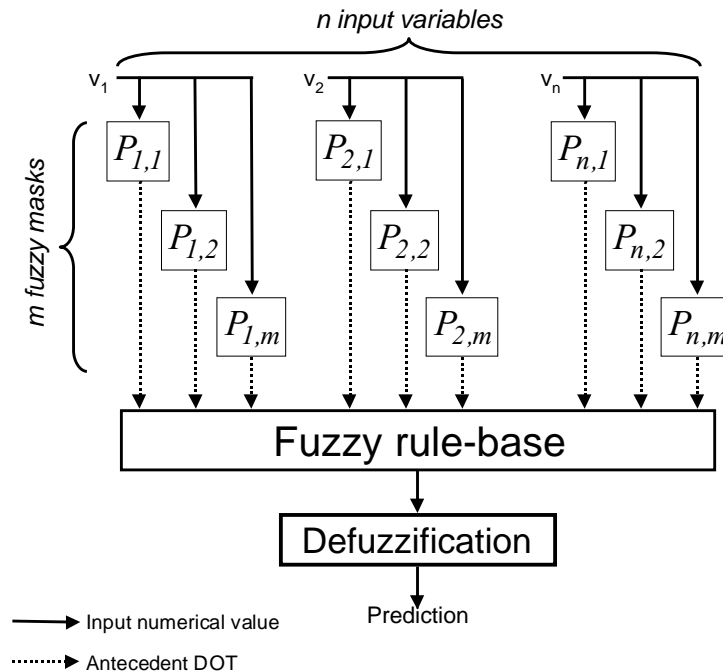
Fig. 2. Complete fuzzy inference scheme. The fuzzy masks are used to obtain the degree of truth of the antecedent variables. Then a Sugeno inference/defuzzification scheme is used to compute the prediction.

previous days. The second application considers the detection of macroalgal bloom conditions in a shallow lagoon subject to eutrophication crises.

3. Application to a medium-size wastewater treatment plant

A municipal wastewater treatment plant located on the outskirts of Florence, Italy, with an oxidation tank of $8000 \, \text{m}^3$ and an average input flow of $400 \, \text{m}^3/\text{h}$ has been considered, with large short-term fluctuations induced by the variable sewage output and the intermittent pumping system. The underlying diurnal flow patterns can be grouped into a limited number of typical behaviours, determined by the level of human activity and the environmental conditions. Fuzzy control has already been successfully applied by the author to wastewater treatment plants (Müller et al., 1997). Since input flow is the only variable for which continuous monitoring was available, it was decided to design a simple control strategy based on its behaviour in the two previous days. A typical daily

flow pattern is shown in Fig. 3. The considerable short-term disturbances is due to the pumping system, which is subject to frequent switching induced by a level control in the front-end storage tank.

Depending on weather conditions and the day of the week, this pattern may change in a fairly predictable way. Processing of an entire year of data, between 2001 and 2002, has led to the conclusion that the meaningful daily patterns reduce to the three profiles shown in Fig. 4, where the short-term noise has been removed and only the significant daily cycles are shown.

These patterns, labelled 'low', 'medium', and 'high', represent the typical flows entering the plant. Medium flow is the most common, being very near the design capacity; high flow represents the typical response to a storm surge, whereas the low pattern is observed during summer dry spells, where most of the population connected to the plant is away and water discharge is reduced accordingly.

After this preliminary pattern selection, an autoregressive model fitted to the residuals provided $\rho = 0.45$, from which the fuzzy curve spread parameter
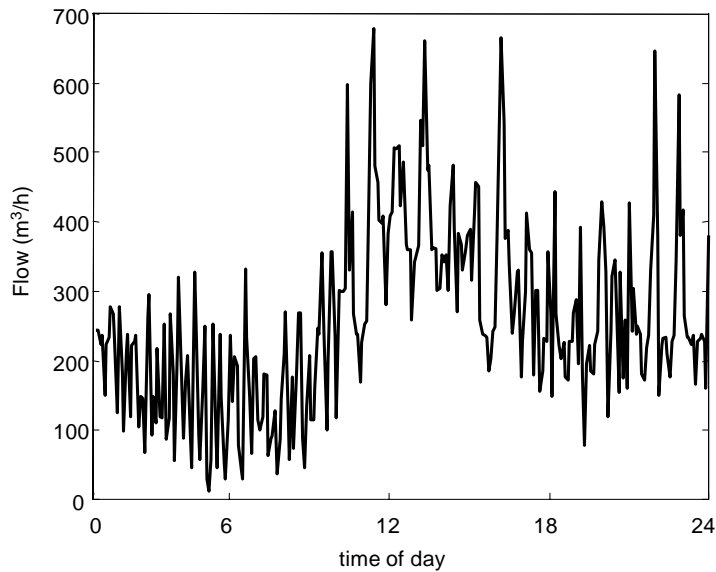
Fig. 3. Typical daily flow pattern for the San Giusto wastewater treatment plant, sampled every 5 min.

was computed through Eq. (3) to yield $b = 0.7913$. Based on the availability of a year of daily records, three fuzzy masks were determined according to the procedures outlined in Sections 2.2 and 2.3. They are shown as shaded plots in Figs. 5–7. The line in the middle of the white region is the fuzzy curve, and the grey scale represents the flow DOT as it departs from

the central fuzzy curve, with white meaning total similarity ($\bar{\mu} = 1$) and black no similarity at all ($\bar{\mu} = 0$).

The next step is the construction of the inferential engine to determine the recycling flow for the next day. Relying on practical experience, it was decided to base this decision on the flow of the previous 2 days and to bound the recycling flow fraction (relative to
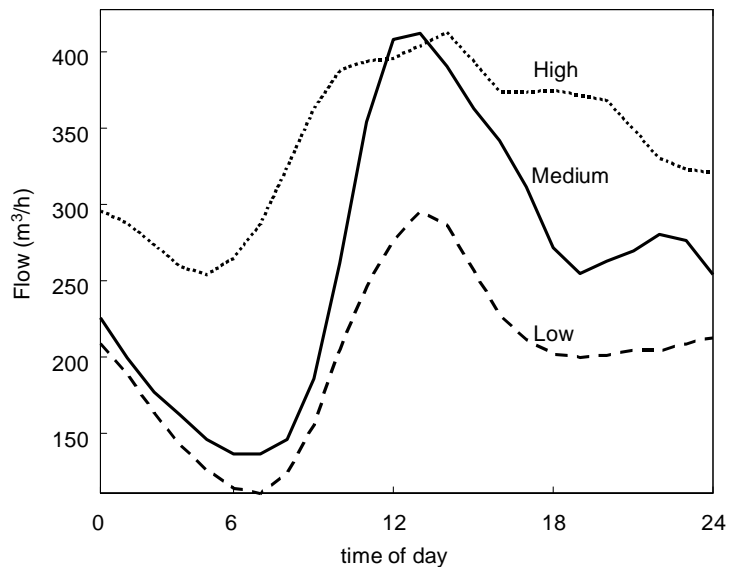

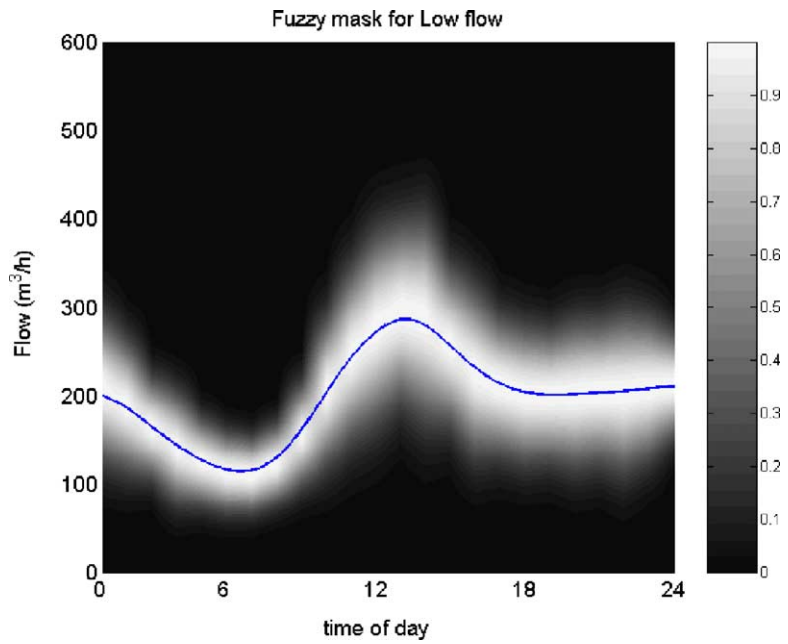Fig. 4. Typical daily flow patterns for the San Giusto plant.

Fig. 5. Fuzzy mask for the 'low' flow condition. The line in the white area represents the fuzzy curve and the grey scale expresses the fuzzy membership to this prototype (white = total similarity; black = no similarity).
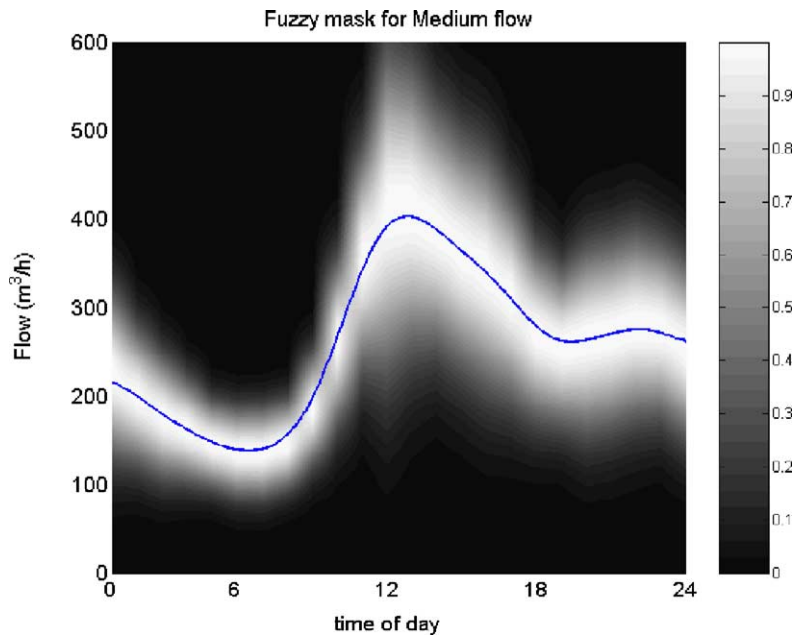

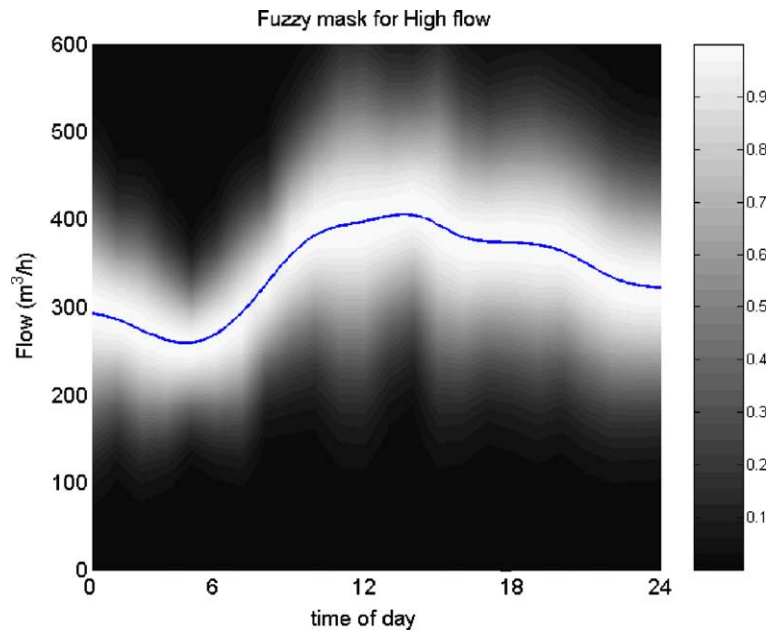Fig. 6. Fuzzy mask for the 'medium' flow condition. Symbols and meanings as in Fig. 5.

Fig. 7. Fuzzy mask for the 'high' flow condition. Symbols and meanings as in Fig. 5.

the process flow) between 0.4 in the low-load situation up to 1.0 in the high-load case, due to the design limitations of the pumping system. The control rules (5) then take the form:

$$R_j: \quad \text{IF } F_t \text{ is } P_{t,j} \text{ AND } F_{t-1} \text{ is } P_{t-1,j}$$
$$\text{THEN } r = r_j, \; j = \text{low, medium, high,}$$
$$(10)$$

where F_t and F_{t-1} represent the flow in the two previous days, and $P_{t,j}$ and $P_{t-1,j}$ are the pertinent fuzzy masks, where the index j refers to the labels low, medium, high. The control rules (10) can be represented as the double-entry control matrix (Yager and Filev, 1994) of Table 1 where each output singleton r_j is activated according to the DOT of each combination of entries (F_t, F_{t-1}). This rule table was defined on the basis of the practical experience gained with that particular plant.

The output singleton from which the treatment levels are obtained were set as $r_{\text{low}} = -1$, $r_{\text{medium}} = 0.5$, $r_{\text{high}} = 2.5$ in order to yield an admissible for the recycling r. It is not surprising that two of the three singletons are outside the operational range,

since the actual r is obtained through defuzzification, i.e. as a weighted average of the three singletons with the DOTs obtained from Table 1 according to Eq. (9).

To test the algorithm, five randomly chosen daily patterns were selected, as shown in Fig. 8. Patterns 1–3 are typical of normal operations, but they differ in the morning lag, which is almost absent in pattern 2. This behaviour is normally observed during a prolonged wet period. Pattern 4 is a typical response to a storm event, whereas pattern 5 is often observed in the summer months when the plant loading conditions are greatly reduced.

Table 1
Fuzzy control matrix for the wastewater treatment control

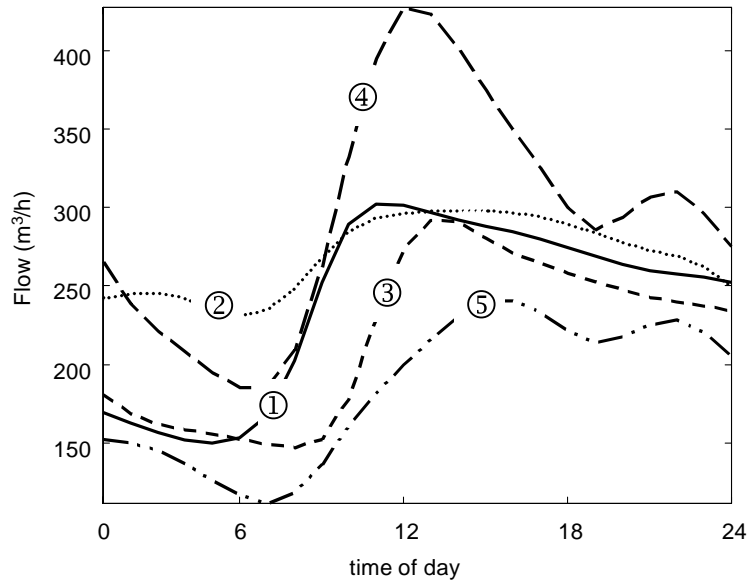|	F_t		
	Low	Medium	Medium
F_{t-1}	Low	Medium	High
	Medium	Medium	High

Fig. 8. Five test flow patterns used to test the algorithm.

Table 2
Results of wastewater treatment control according to the fuzzy pattern estimator

Flow transition	Rule matrix activation			Treatment level
Flow1 to flow5	0.9992	0.8402	0.5101	0.4974
	0.9995	0.8404	0.5102	
	0.7256	0.6101	0.3704	
Flow2 to flow1	0.9991	0.9994	0.7255	0.7206
	0.9988	0.9991	0.7253	
	0.9636	0.9639	0.6997	
Flow1 to flow3	0.9987	0.9992	0.5925	0.5709
	0.9990	0.9995	0.5927	
	0.7252	0.7256	0.4303	
Flow5 to flow4	0.5481	0.9947	0.9994	0.8686
	0.4609	0.8364	0.8404	
	0.2798	0.5078	0.5102	
Flow5 to flow5	0.9991	0.8401	0.5100	0.4054
	0.8401	0.7064	0.4289	
	0.5100	0.4289	0.2604	
Flow2 to flow2	0.9990	0.9987	0.9635	0.8093
	0.9987	0.9984	0.9632	
	0.9635	0.9632	0.9292	
Flow3 to flow4	0.5479	0.9942	0.9990	0.9127
	0.5482	0.9947	0.9995	
	0.3250	0.5898	0.5927	

Several combinations of these five patterns were simulated and the resulting treatment levels were obtained. These results are summarised in Table 2. The first column shows the transition from one flow to another in the last 2 days, F_t and F_{t-1}. They are the inputs to the fuzzy control matrix of Table 1, whose DOTs are shown in the second column. The third column shows the resulting treatment level r. For example, if the flow changes from condition 1 to 5 (first row), which is an extreme variation, the recycling is lowered to 0.4974, close to the absolute minimum value of 0.4. Conversely, a change from condition 3 to 4 (bottom row), possibly due to a storm event, brings the recycling to 0.9127, close to maximum allowed value. The other rows represent other possible flow transitions. The recycling determined by these rules compare favourably with manual operating records, and the algorithm was therefore considered a useful aid by the plant manager.

4. Application to macroalgae bloom forecasting in the Orbetello lagoon

The second application of the algorithm is aimed at predicting macroalgal blooms in the Orbetello lagoon.

In this context, it is used as a decision support tool to determine appropriate harvesting strategies before the full extent of the bloom takes place. For operational reasons, a monthly time-scale is selected.

4.1. The Orbetello lagoon

The Orbetello lagoon, which is situated along Italy's west coast, has an extent of approximately 27 km². It is composed of two shallow coastal ponds of an average depth of 1.2 m with one water inlet at each end of the western pond and one outlet at the south-eastern end of the eastern pond. The two ponds communicate though a narrow passage under a bridge supporting the road connecting the little town of Orbetello with Mount Argentario. Two water quality monitoring stations have been deployed in the middle of each pond and transmit hourly data to the management office in Orbetello. These stations are indicated by the two solid dots in Fig. 9, whereas the hollow circles indicate the manual sampling stations and the solid squares the major sewage discharge points.

4.2. Main characteristics of the Orbetello lagoon ecosystem

Previous studies (Cartei et al., 1998; Innamorati, 1998; Lenzi and Mattei, 1998; Lenzi, personal communication, 2001) have shown that the yearly storage rate of the lagoon, in terms of the total mass, is about 4.8% for nitrogen and 4.3% for phosphorus. This was obtained through a nutrient balance between inputs, consisting of discharge from sewage and fish farms, and output, represented by washout to the sea. These rates are more than enough to sustain the present primary production and represent a considerable potential for further eutrophication. This is because both natural flow, induced by tide and wind, and artificial pumping do not provide the necessary turnover for the removal of a sufficient fraction of the nutrients stored in the sediments. It was estimated that even with a forced through-flow of 4 m³ s⁻¹ the lagoon's hydraulic retention time would exceed 60 days. As a result, in the last few years the frequency of anoxic conditions and algal blooms has increased. This situation is reflected by the repeated occurrence of low values of
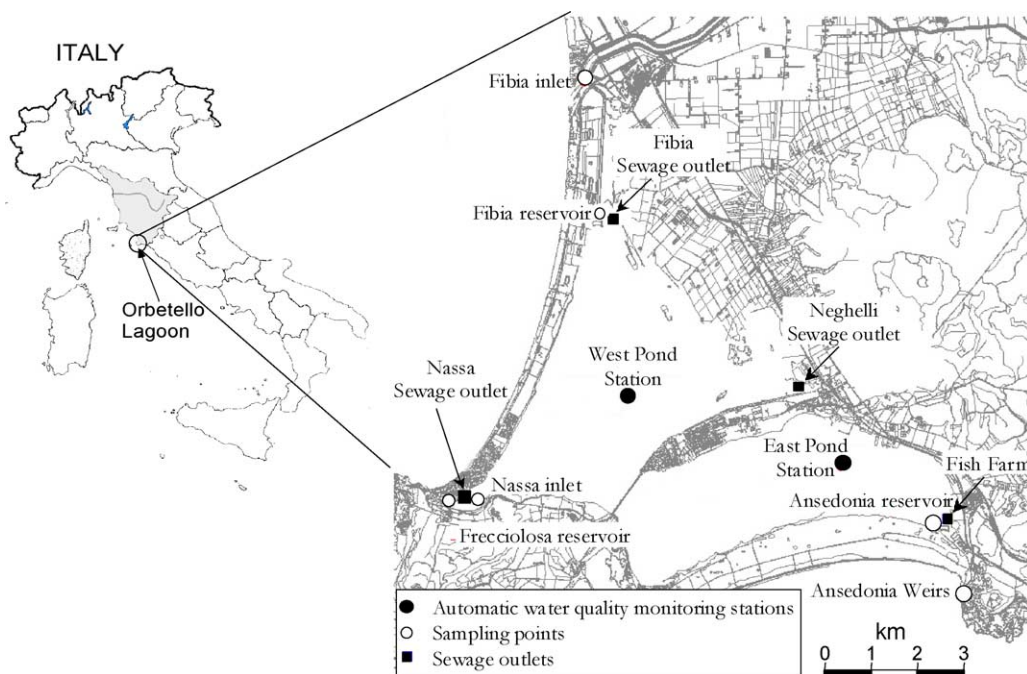

Fig. 9. Map of the Orbetello lagoon showing the two water quality monitoring stations, manual sampling locations and main discharge points.

the dissolved oxygen, below 20% of saturation. This condition prevents the full oxidation of ammonium-N to nitrate-N and favours the entrapment of nutrients in the sediment in organic form.

Macroalgae dominate the lagoon and play a key role in determining the oxidised or reduced state of the sediments, which is the main factor controlling nutrient cycling. This has already been reported by many authors in other lagoons (Castel et al., 1996; Christian et al., 1998; Cioffi and Gallerano, 2000; Viaroli et al., 2001) and by Cioffi et al. (1995) and Cartei et al. (1998) for the Orbetello lagoon in particular. Here the prevailing macroalgae species are *Chaetomorpha linum*, *Cladophora vagabunda*, *Gracilaria verrucosa*, and, most abundant, *Ulva rigida*. Though of epiphytic origin, they are mostly floating during the developed stage and compete for light with seagrass (*Ruppia maritima*). Given the large availability of nutrients and the limited water renewal, when the macroalgae growth ends and decomposition ensues, an oxygen imbalance may occur, causing anoxia. This kind of dystrophic crisis has been described by many authors (Castel et al., 1996; Christian et al., 1998; Viaroli et al., 2001). During the normal growth phase, macroalgae represent a sink for dissolved inorganic nitrogen and oxidising processes prevail: daytime dissolved oxygen (DO) has a well defined afternoon peak, often well above the saturation level. When the growth phase ends, the fast anoxic decomposition enriches the sediment with reduced organic nitrogen (Christian et al., 1998). These reducing conditions can be detected by low, almost constant DO and low daytime oxidation–reduction potential (ORP) in the water, which corresponds to near zero or negative ORP at the sediment surface, as reported by Gomez et al. (1999) and Azzoni et al. (2001). These authors have shown that depending on the season, the gradient between water and bottom ORP may be as high as 300 mV. Hence, a water ORP of 200 mV or less may already indicate anoxic conditions, with negative ORP, in the sediment upper layer. The relationship between macroalgae growth, nutrient cycling and daily patterns of DO is shown in Fig. 10, which was inspired by Christian et al. (1998). It was observed that large daily fluctuations of DO and pH signal the peak of the growth period, with increased oxygen production and CO_2 depletion, after which a dystrophic crisis might be expected if the corresponding daily cycles flatten out. Temperature and ORP

contribute to the detection, since the former acts as a primary factor in controlling growth, whereas the latter decreases whenever reducing processes at the bottom prevail, indicating conditions favourable to nutrient release. Thus, the daily cycles of these four, easily measurable variables can be used as indicators of the macroalgae growth stage, instead of complex nutrient or chlorophyll measurement which require considerable effort in manual sampling and analysis.

4.3. Qualitative analysis of the daily patterns

The basic idea is to detect water conditions favourable to macroalgal bloom by observing the circadian fluctuations of easily measurable physico-chemical water parameters. The inference mechanism is based on their relationship with the macroalgae life cycle, as described in the previous section and Fig. 10 in particular. The algorithm consists of a set of fuzzy rules relating the diurnal pattern of DO, pH, ORP and temperature to the extent of macroalgal development on a monthly time-scale. The predictor is developed following the procedure outlined in Section 2: first a qualitative data analysis is performed to determine the meaningful daily patterns, then the fuzzy masks are derived and the inference fuzzy rules defined. Unlike the previous application, no a priori values for the consequents were assigned, but they were fitted to the observed bloom conditions.

The available data covers the period from October 2000 to January 2002, with an hourly sampling period. In grouping the data and in algorithm design, a monthly time-scale was adopted as a compromise between the timing intervals of the harvests and the changes in the macroalgae phenology. For reasons explained in the previous section, it is assumed that the daily patterns of these parameters indicate the build-up of macroalgal blooms, as they are related to the physico-chemical conditions responsible for the blooms. The extent and variability of the diurnal cycles can be appreciated by observing Figs. 11 and 12, which compare the daily patterns of DO in spring and in winter (each line represents a daily cycle).

It can be seen from Fig. 11 that in spring the DO level is consistently around the saturation level and follows a typical daily pattern, with an early morning sag and a pronounced afternoon peak. This is due to the intense growth activity that may anticipate a macroalgal
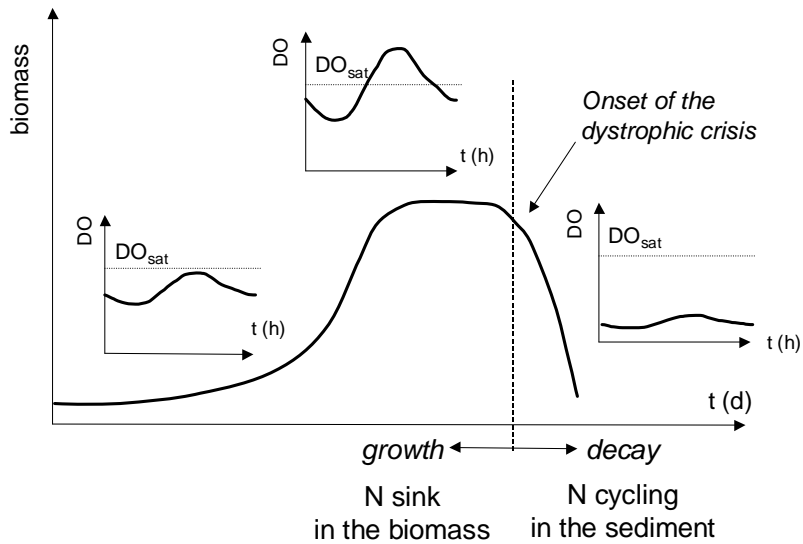
Fig. 10. Relationship between biomass evolution and nitrogen cycling: during growth the circadian oxygen profile shows an ever-increasing daytime peak. In the decay phase, the decomposing biomass indices anoxic processes thereby causing the oxygen to remain at low values.

bloom. Conversely, during the winter months hardly any diurnal cycle is discernible, as Fig. 12 shows. The corresponding pH patterns for the same months, given in Figs. 13 and 14, exhibit a behaviour coherent with the DO cycles. In fact, the April patterns (Fig. 13) show a increase in afternoon pH as a consequence to CO_2 depletion due to algal growth, corresponding to the DO peak, whereas the January values (Fig. 14) are lower and lack the afternoon increase.

The other two parameters (ORP and temperature) also exhibit a circadian pattern with significant sea-

sonal variations. The basic idea in developing the fuzzy predictor is to find a set of fuzzy rules that relate these variations to the likelihood of a macroalgal bloom.

4.4. Data clustering and the extraction of fuzzy masks

To obtain the starting knowledge-base, quantitative information on the extent of macroalgae bloom over the period from October 2000 to January 2002
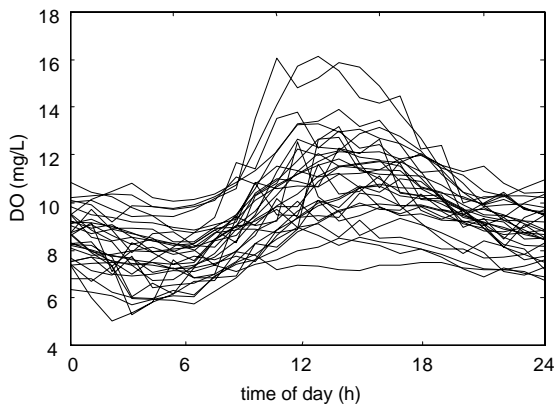

Fig. 11. Diurnal DO variations in April 2001 (each line represents a daily cycle).
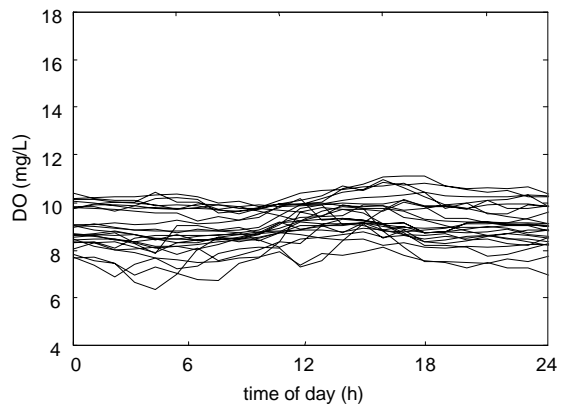

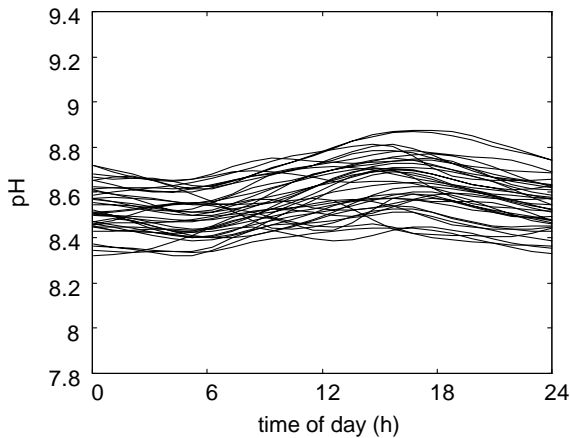Fig. 12. Diurnal DO variations in January 2001 (each line represents a daily cycle).

Fig. 13. Diurnal pH variations in April 2001 (each line represents a daily cycle).

Table 3
Algal bloom experimental classification assumed as the knowledge-base

Month	Bloom	Class
October 2000	0.17	Low
November 2000	0.26	Low
December 2000	0.40	Medium
January 2001	0.50	Medium
February 2001	0.55	High
March 2001	0.67	High
April 2001	1.00	High
May 2001	0.80	High
June 2001	Missing data	Not included
July 2001	0.17	Low
August 2001	0.07	Low
September 2001	0.16	Low
October 2001	0.56	Validation
November 2001	0.50	Validation
December 2001	0.42	Validation
January 2002	0.34	Validation

was obtained. The degree of bloom was based on the observed algal density as reported by the harvesting boats. The scale was determined by assuming that a harvest of $1\,kg\,m^{-2}$ corresponds to 0.4 fuzziness, $3\,kg\,m^{-2}$ to 0.6 and $10\,kg\,m^{-2}$ to 1.0. The data were partitioned into a calibration set, from October 2000 to September 2001, and a validation set from October 2001 to January 2002. Three clusters were determined based on the observed algal density: the months in which a low algal development was observed, i.e. with a degree of bloom below 0.3, were grouped in the 'low' class, those with a degree of bloom between 0.3
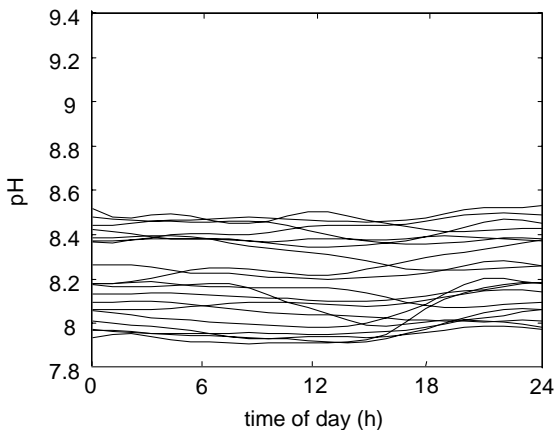
and 0.5 were placed in the 'medium' class and those with a degree of bloom greater than 0.5 were placed in the 'high' class. The results of the bloom clustering to form the training and validation sets are summarised in Table 3. Algorithm training was performed on the first subset (October 2000–September 2001), whereas the last four entries (October 2001–January 2002) were used as the validation set and therefore were not included in the knowledge-base.

Since monthly data were available, a monthly time-scale was adopted and the fuzzy masks were derived using the monthly average values $\{y_{avg}(k)\,k = 0, ..., 23\}$, following the approach described in Sections 2.2 and 2.3. Since there are four variables (DO, pH, ORP, Temp), and for each of them three classes were defined, 12 fuzzy masks were produced. Each of them is characterised by the fuzzy curve $c_{i,j}(k)$ and the variance $\sigma_{i,j}^2(k)$ of the Gaussian approximation, where the index j refers to the class (low, medium, high) and the index i to the variable (DO, pH, ORP, Temp).

4.5. Design of the fuzzy predictor

Now that the fuzzy masks producing the antecedent DOTs have been obtained, fuzzy rules and consequent data are required to complete the fuzzy inferential engine. Based on the observed macroalgae bloom behaviour, summarised in Table 3, on the basis of the

Fig. 14. Diurnal pH variations in January 2001 (each line represents a daily cycle).

Table 4
Rule-base for the fuzzy bloom predictor

Rule #1	Rule #2	Rule #3
IF DO is DO_low	IF DO is DO_medium	IF DO is DO_high
AND pH is pH_low	AND pH is pH_medium	AND pH is pH_high
AND T is T_low	AND T is T_medium	AND T is T_high
THEN Bloom $= B_{\text{low}}$	THEN Bloom $= B_{\text{medium}}$	THEN Bloom $= B_{\text{high}}$

observed phenological evidence (Cartei et al., 1998; Lenzi and Mattei, 1998) and through discussions with marine biologists in charge of lagoon monitoring, the set of inferential rules in Table 4 were defined.

The practical implementation of these rules requires the evaluation of the antecedent DOTs of circadian patterns with respect to the fuzzy masks according to

$$\mu_j^i(k) = \exp\left(-\frac{(\bar{y}_i(k) - c_{i,j}(k))^2}{2 \times \sigma_{i,j}^2(k)}\right),$$

$$j = \text{low, medium, high}, \quad i = \text{DO, ORP, pH, Temp},$$

$$k = 0, \ldots, 23, \tag{11}$$

where $\{\bar{y}_i(k); k = 0, \ldots, 23; i = \text{DO, ORP, pH, Temp}\}$ are the average monthly data. These DOTs are then averaged over the 24-h period to yield the antecedent value for each parameter on a monthly basis:

$$\bar{\mu}_i^j = \frac{1}{24}\sum_{k=0}^{23}\mu_i^j(k), \qquad j = \text{DO, ORP, pH, Temp}, \quad i$$

$$= \text{low, medium, high}. \tag{12}$$

The quantities $\bar{\mu}_i^j$ represent the antecedents in the fuzzy rules of Table 4, which can be written as

$$R_1: \quad \bar{\mu}_{\text{low}}^{\text{DO}} \wedge \bar{\mu}_{\text{low}}^{\text{ORP}} \wedge \bar{\mu}_{\text{low}}^{\text{pH}} \wedge \bar{\mu}_{\text{low}}^{\text{Temp}} \wedge B_{\text{low}},$$

$$R_2: \quad \bar{\mu}_{\text{medium}}^{\text{DO}} \wedge \bar{\mu}_{\text{medium}}^{\text{ORP}} \wedge \bar{\mu}_{\text{medium}}^{\text{pH}} \wedge \bar{\mu}_{\text{medium}}^{\text{Temp}}$$
$$\qquad \wedge B_{\text{medium}},$$

$$R_3: \quad \bar{\mu}_{\text{high}}^{\text{DO}} \wedge \bar{\mu}_{\text{high}}^{\text{ORP}} \wedge \bar{\mu}_{\text{high}}^{\text{pH}} \wedge \bar{\mu}_{\text{high}}^{\text{Temp}} \wedge B_{\text{high}}. \tag{13}$$

As in the previous application, the connective $\wedge$ is implemented with a product operator. The predicted macroalgae bloom for each month is obtained by Sugeno defuzzification, i.e. by a weighted average of

the bloom singletons B_i through their DOTs:

$$\text{Bloom} = \frac{\sum_i \mu_i \times B_i}{\sum_i \mu_i}, \quad i = \text{low, medium, high}. \tag{14}$$

Unlike the previous case, where the output singletons were set in advance, the three macroalgae bloom singletons were estimated in order to adapt the algorithm to the observations. Since only monthly information was available, the daily data were averaged in order to compare with the observed behaviour. Then this composite quantity was used to minimise the squared sum of relative errors between observed and predicted blooms:

$$E(\boldsymbol{B}) = \sum_{m=1}^{11}\left(\frac{\text{Bloom}(m) - \text{Bloom}_{\text{exp}}(m)}{\text{Bloom}_{\text{exp}}(m)}\right)^2 \tag{15}$$

with respect to the singletons $\boldsymbol{B}$ over the 11-month calibration period.

The minimisation of Eq. (15) was obtained through a modified version of the Simplex search method (Marsili-Libelli, 1992) and the result is shown in Table 5 and Fig. 15. The predictor was then validated with a further 4-month period (October 2001–January 2002), during which a secondary bloom occurred. The agreement between observed and predicted bloom retained the same order of accuracy as the calibration period, as shown in Fig. 15. These results were obtained with a monthly time-scale, since the algorithm was based on monthly averages. To test the algorithm on a shorter time-scale, the pattern of each single

Table 5
Estimated values of bloom singletons (B^*)

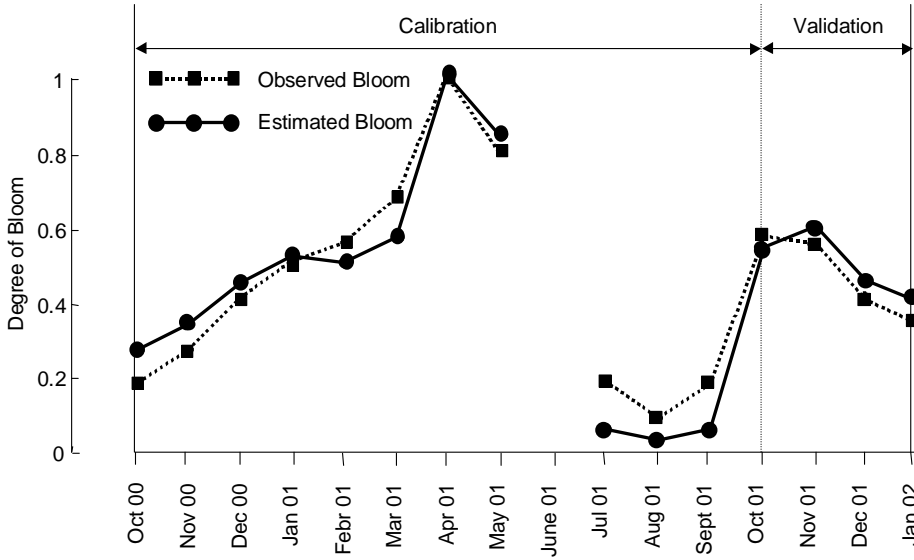|	
B_{low}	−0.213
B_{medium}	0.587
B_{high}	2.134

Fig. 15. Comparison between observed and predicted blooms. The calibration period covers 11 months (October 2000–September 2001) with June 2001 data missing because of sensor failures. The period from October 2001 to January 2002 is used for validation.

day in 2001 was applied as an input to the calibrated predictor and the response of Fig. 16 was obtained without any further tuning. It can be seen that the basic pattern of Fig. 15 is maintained, showing that the predictor can operate on a shorter time-scale. To carry the analysis a step further, the algorithm was applied to the 2002 data (from 1 January to 16 December) and the results of Fig. 17 were obtained, from which some considerations can be drawn. Here the algorithm was applied to both ponds of the lagoon, though the knowledge basis and the calibrated consequents of Table 5

were originally tuned with the east pond data. It can be seen that consistent predictions for both ponds are obtained. Moreover, a malfunctioning of the DO probe during February 2002 in the west pond is not enough to cause a "false positive" since the resulting macroalgae bloom likelihood is only 0.7 which can be considered a "borderline" case, perhaps an alert value, but certainly not a full bloom alarm. Lastly, the observed higher tendency of the east pond to produce macroalgae bloom is confirmed, with two major eutrophication episodes at the beginning of May and of September.
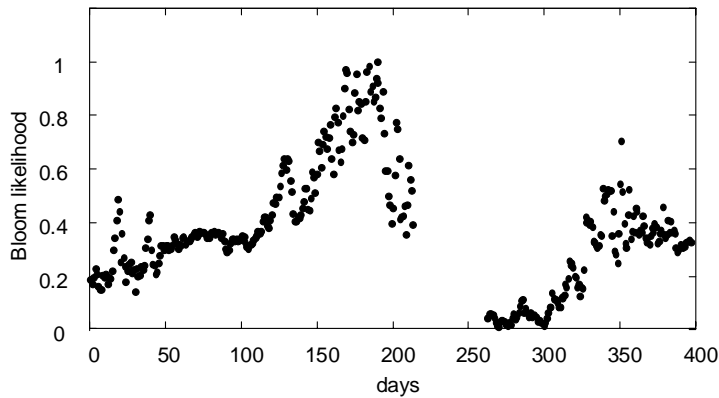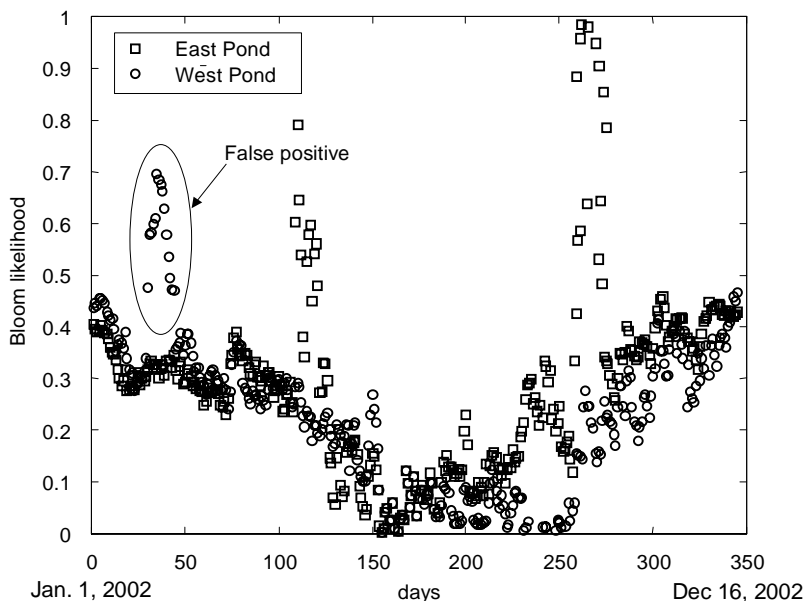

Fig. 16. Daily bloom prediction during 2001.

Fig. 17. Results of the bloom fuzzy prediction during 2002. The nearly false positive detection (indicated by circles) in the west pond in January was due to a faulty DO sensor.

5. Conclusion

The variations of diurnal fluctuations may contain important ecological information and if properly interpreted can represent a valuable predictive tool. This paper has presented a simple algorithm to analyse daily cycles in ecological variables using fuzzy pattern recognition techniques. The algorithm is organised in three parts: in the first, typical patterns are extracted from experimental data to form the knowledge-base upon which the algorithm operates. The second step is to condense the information of the knowledge-base into mathematical objects, termed fuzzy masks. The third step is the set-up of inferential fuzzy rules, which use the fuzzy masks as antecedents to provide an assessment of the current daily cycle. After proper tuning against experimental data, the algorithm can classify the current daily cycle in the context of the typical ecosystem behaviours and infer possible consequences of cycle alterations.

To demonstrate the algorithm, two applications were presented. In the first, recycling control in a medium-scale domestic wastewater treatment plant was determined based on the diurnal flow variations of the previous 2 days. The knowledge-base was extracted from an entire year of flow recordings and the results compare favourably with manual control.

The second, and most elaborate, application was the prediction of macroalgae blooms in the Orbetello lagoon, whose management relies on their early harvesting. A whole-year data set (October 2000–October 2001) of four basic water quality parameters (DO, pH, ORP and temperature) were used as the starting knowledge-base. A consistent cause–effect relationship was observed between circadian variations of these variables and the complex macroalgae–nutrient interactions, justifying the use of these simple measurements as indicators of macroalgal activity. The observed patterns were partitioned into three classes depending on the corresponding observed bloom degree. Twelve fuzzy masks (four variables by three classes) were created, combining daily fluctuation and monthly statistical dispersion. Then a three-rule inferential engine with the fuzzy masks as antecedents was determined to assess the current daily cycle with respect to bloom occurrence. A further step was introduced in this example, since the bloom consequents had to be estimated to tune the algorithm with the observed degree of bloom, related to the harvested macroalgal density. After a 4-month validation period,

more tests were performed. First, the predictor calibrated with the data from one part of the lagoon was used to forecast the bloom in the other part, then the algorithm was also tested on daily data. In all cases, the predictor provided reliable results without the need for recalibration and avoiding "false positive" responses. Finally, the data of the whole 2002 year was processed and more predictions have been produced, from which some important observations, as the higher tendency of the east pond to become eutrophic, are confirmed.

References

Azzoni, R., Giordani, G., Bartoli, M., Welsh, D.T., Viaroli, P., 2001. Iron, sulphur and phosphorus cycling in the rhizosphere sediments of a eutrophic *Ruppia cirrhosa* meadow (Valle Smarlacca, Italy). J. Sea Res. 45, 15–26.

Babuska, R., 1998. Fuzzy Modeling for Control. Kluwer Academic Publishers, Amsterdam, p. 260.

Bárdossy, A., Duckstein, L., 1995. Fuzzy Rule-based Modeling with Applications to Geophysical, Biological and Engineering Systems. CRC Press, Boca Raton, p. 232.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York, p. 256.

Bobbin, J., Recknagel, F., 2001. Knowledge discovery for prediction and explanation of blue-green algal dynamics in lakes by evolutionary algorithms. Ecol. Model. 146, 253–262.

Cartei, P., Innamorati, M., Melillo, C., 1998. Trophic homeostasis sea modulated and lagoon autochtonous hypertrophy. Biol. Mar. Medit. 5 (1), 41–46.

Castel, J., Caumette, P., Herbert, R., 1996. Eutrophication gradients in coastal lagoons as exemplified by the Bassin d'Arcachon and the Etang du Prevost. Hydrobiology 329, ix–xxviii.

Christian, R.R., Naldi, M., Viaroli, P., 1998. Construction and analysis of static, structured models of nitrogen cycling in coastal ecosystems. In: Koch, A.L., Robinson, J.A., Milliken, G.A. (Eds.), Mathematical Modelling in Microbial Ecology. Chapman & Hall, New York.

Cioffi, F., Di Eugenio, A., Gallerano, F., 1995. A new representation of anoxic crises in hypertrophic lagoons. Appl. Math. Model. 19, 685–695.

Cioffi, F., Gallerano, F., 2000. Response of Lake Piediluco to the change of hydrodynamic conditions and nutrient load reductions. Ecol. Model. 135, 199–229.

Ganoulis, J., 1994. Engineering Risk Analysis of Water Pollution Probabilities and Fuzzy Sets. VCH, Weinheim, p. 306.

Gomez, E., Durillon, C., Rofes, G., Picot, B., 1999. Phosphate adsorption and release from sediments of brackish lagoons: pH, O_2 and loading influence. Water Res. 33 (10), 2437–2447.

Håkanson, L., 1999. On the principles and factors determining the predictive success of ecosystem models, with a focus on lake eutrophication models. Ecol. Model. 121, 139–160.

Innamorati, M., 1998. The Orbetello lagoon: restoration and conservative management. Biol. Mar. Medit. 5 (3), 1352–1361.

Jeong, K.S., Joo, G.J., Kim, H.W., Ha, K., Recknagel, F., 2001. Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network. Ecol. Model. 146, 115–129.

Lenzi, M., Mattei, M., 1998. Results of a four-year management of the lagoon ecosystem of Orbetello through algal biomass harvesting. Biol. Ital. 2, 7–12.

Lin, Y., Cunningham III, G.A., 1995. A new approach to fuzzy-neural system modeling. IEEE Trans. Fuzzy Syst. 3 (2), 190–198.

Lin, Y., Cunningham III, G.A., Coggeshall, S.V., 1996. Input variable identification—fuzzy curves and fuzzy surfaces. Fuzzy Sets Syst. 82, 65–71.

Maier, H.R., Dandy, G.C., Burch, M.D., 1998. Use of artificial neural networks for modelling cyanobacteria *Anabaena* spp. in the River Murray. South Aust. Ecol. Model. 105, 257–272.

Maier, H.R., Sayed, T., Lence, B.J., 2001. Forecasting cyanobacterium *Anabaena* spp. in the River Murray, South Australia, using B-spline neurofuzzy models. Ecol. Model. 146, 85–96.

Marsili-Libelli, S., 1992. Parameter estimation of ecological systems. Ecol. Model. 62, 233–258.

Marsili-Libelli, S., Müller, A., 1996. Adaptive fuzzy pattern recognition in the anaerobic digestion process. Pattern Recogn. Lett. 17, 651–659.

Mpimpas, H., Anagnostopoulos, P., Ganoulis, J., 2001. Modelling of water pollution in the Thermaikos Gulf with fuzzy parameters. Ecol. Model. 142, 91–104.

Müller, A., Marsili-Libelli, S., Aivasidis, A., Lloyd, T., Kroner, S., Wandrey, C., 1997. Fuzzy control of disturbances in a wastewater treatment system. Water Res. 31, 3157–3167.

Recknagel, F., 2001. Applications of machine learning to ecological modelling. Ecol. Model. 146, 303–310.

Scardi, M., 2001. Advances in neural network modelling of phytoplankton primary production. Ecol. Model. 146, 33–45.

Scardi, M., Harding, L.W., 1999. Developing an empirical model of phytoplankton primary production: a neural network case study. Ecol. Model. 120, 213–223.

Shen, Q., Chouchoulas, A., 2001. FuREAP: a fuzzy-rough estimator of algae populations. Artif. Intell. Eng. 15, 13–24.

Takagi, T., Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. Syst. Man Cybern. 15 (1), 116–132.

Viaroli, P., Azzoni, R., Bartoli, M., Giordani, G., Taje, L., 2001. Evolution of the trophic conditions and dystrophic outbreaks in the Sacca di Goro lagoon (Northern adriatic sea). In:

Faranda, F.M., Guglielmo, L., Spezie, G. (Eds.), Mediterranean Ecosystems: Structures and Processes. Springer-Verlag, Italy.

Wei, B., Sugiura, N., Maekawa, T., 2001. Use of artificial neural network in the prediction of algal blooms. Water Res. 35 (8), 2022–2028.

Yabunaka, K., Hosomi, M., Murakami, A., 1997. Novel application of a back-propagation artificial neural network model formulated to predict algal bloom. Water Sci. Technol. 36 (5), 89–97.

Yager, R.R., Filev, D.P., 1994. Essentials of Fuzzy Modelling and Control. Wiley, New York, p. 388.