Monitoring Water Quality through a Telematic Sensor Network and a Fuzzy Expert System

Evaggelos V. Hatzikos[1], Nick Bassiliades[*2], Leonidas Asmanis[2], Ioannis Vlahavas[2]

`hatzikos@teithe.gr, nbassili@csd.auth.gr, lasmanis@gmail.com,`
`vlahavas@csd.auth.gr`

[1]Technological Educational Institute of Thessaloniki, Greece

[2]Dept. of Informatics, Aristotle University of Thessaloniki, Greece

Abstract

In this paper we present an expert system that monitors sea water quality and pollution in Northern Greece, through a sensor network called "Andromeda". The expert system monitors sensor data collected by Local Monitoring Stations and reasons about the current level of water suitability for various aquatic uses, such as swimming and piscicultures. The aim of the expert system is to help the authorities in the "decision-making" process in the battle against the pollution of the aquatic environment, which is very vital for the public health and the economy of Northern Greece. The expert system determines, using fuzzy logic, when certain environmental parameters exceed certain "pollution" limits, which are specified either by the authorities or by environmental scientists, and flags out appropriate alerts.

Keywords: Sensor Network, Pollution Monitoring, Aquatic Uses, Expert System, Fuzzy Logic

1. Introduction

The environment includes the atmosphere, the soil and the water. By the term "water", we mean in general the aquatic resources, either surface waters (e.g. seas, lakes, tanks, rivers, torrents) or underground aquatic volumes [2]. In the technologically advanced countries, particularly in the USA and the countries of Western Europe, the need for monitoring the parameters related with qualitative environmental characteristics and especially with the quality of water, has been recognized for a long time. To this aim, several programs of

---

[*] Corresponding author.

automatic measurement of qualitative characteristics and analysis of results have been installed and placed in operation [6].

If the authorities or organizations involved with the management of aquatic resources were able to monitor the quantitative and qualitative parameters related to the aquatic environment, then they would be able to draw conclusions about parameter trends, to predict undesirable situations on time and, therefore, to take counter-measures in a timely fashion. Furthermore, they could enforce longer term actions, including strategic resource planning for regional growth and development.

To this end, the authors have developed two automated water monitoring sensor networks, one for sea waters, called "Andromeda" [7], and one for fresh waters, called "Interisk" [11]. In this paper, we present an expert system that provides added value services on top of the Andromeda network. The purpose for building such an expert system is to help the authorities in the "decision-making" process in the battle against the pollution of the aquatic environment [8], by intelligently monitoring hydrological parameters of sea waters. The expert system is able to determine when the environmental parameters exceed certain "pollution" limits, which are specified either by the authorities or by environmental scientists, and issue appropriate alerts. The system is able to check complex conjunctive situations using fuzzy logic [21]. Furthermore, the system is able to distinguish between different limits (and corresponding alerts) that exist for different aquatic uses (e.g. pisciculture, swimming, etc).

The *Water Monitoring Expert System* (henceforth, WMES) on-line monitors sensor data collected by the Local Monitoring Stations of the Andromeda network and reasons about the current level of water suitability. Activities carried out by the WMES could in theory be successfully performed by a human expert, though with considerable effort. Given the unavailability of such an option in most prefectures of Northern Greece, the alternative solution of using an expert system instead, such as WMES, is considered very important both for the public health and, most importantly, for the economy of Northern Greece, which strongly depends on water.

The WMES processes data collected by the installed network and draws conclusions about the current level of water suitability. If the data are deemed normal, no need for alert arises. On the contrary, if the data do not fall within the preset acceptable ranges for at least one of the possible aquatic uses, an appropriate warning message is displayed on screen. WMES copes with two aquatic uses concerning sea water: shell cultivation and swimming. In addition, it deals with two types of freshwater pisciculture (Cyprinidae and Salmonidae), albeit it is not yet connected on-line to the Interisk network.

This paper is structured as follows. In Section 2 we briefly review relevant systems found in the literature. Section 3 briefly describes the Andromeda network which consists of the Local Monitoring Stations and the Main Station. Section 4 presents the architecture and functionality of the Water Monitoring Expert System, including input data, expert knowledge, the communication protocol between LabView[1] and MATLAB[2], and finally, result presentation. In Section 5, the most important issues regarding the implementation of the WMES within MATLAB are detailed, namely the Fuzzy Inference Systems and the major data structures and algorithms developed in the main module. Finally, in Section 6, we conclude the paper and we sketch the line of future research.

2. Related Work

There exist quite a few systems in the literature that monitor the environment, in general, or sea/fresh water, in specific, and alert the user about possible dangers or water suitability. Usually, such systems are strongly related to the local environment they monitor, since environmental monitoring is a very complex task that is strongly dependent on the geomorphologic features of the monitored area. To the best of our knowledge, our system is the first one for monitoring the aquatic environment of northern Greece. In this section we review few such environmental and/or water monitoring systems.

In [10], an expert system for marine environmental monitoring is presented. This system was the product of the cooperation between the National Oceanic and Atmospheric Administration's Atlantic Oceanographic and Meteorological Laboratory and the Florida Institute of Oceanography (FIO). These two institutes implemented the SEAKEYS (Sustained Ecological Research Related to Management of the Florida Keys Seascape) network, which is situated along 220 miles of the coral reef tract within the Florida Keys National Marine Sanctuary (FKNMS). This network actually extends seven Coastal-Marine Automated Network (C-MAN) stations for long-term monitoring of meteorological parameters (wind speed, wind gusts, air temperature, barometric pressure, relative humidity). SEAKEYS adds oceanographic parameters (sea temperature, photosynthetically active radiation, salinity, fluorometry, optical density) to the station. As a recent enhancement to the SEAKEYS network, an expert system has been employed to provide daily interpretations of near real-time acquired data for the benefit of scientists, fishermen and skin divers. These

---

[1] National Instruments, http://www.ni.com/

[2] The MathWorks, http://www.mathworks.com/

interpretations are designed to be automatically emailed to Sanctuary managers and to the FIO maintainers of the network. The first set of interpretations includes those dealing with environmental conditions conducive to coral bleaching. Other marine environmental interpretations are slated to follow.

Knowledge-based reasoning was employed in [17] to develop an environmental protection system that would deal with air pollution control in urban regions. The main goal of the system is to notify the authorities about the possible locations that air pollution might occur. The estimations are based on current and past measurements of the concentrations of the pollutant substances and to meteorological data of the same time interval. Apart from the warnings, the system is capable of suggesting possible solutions for avoiding possible dangerous situations.

Another knowledge-based approach was in [20] and [19], to build a system whose main goal is to classify lake-water resources in five acid-sensitive regions of the United States. It consists of a network of Decision Support Systems (DSS), one for each region. It is based on a set of rules that is used to represent the knowledge that was derived from the human acid-base chemistry experts. The DSS allows federal land managers to conduct a preliminary assessment of the status of individual lakes prior to consulting an expert. The DSS accurately portrays the decision structure and assessment outcomes of domain experts while capturing interregional differences in acidification sensitivity and historic acid deposition loadings. It is internally consistent and robust with respect to missing water chemistry input data.

The case-based reasoning system, presented in [4] and [5], copes with water pollution. It specializes in forecasting the red tide phenomenon in a complex and dynamic environment in an unsupervised way. Red tides are the name for the sea water discolorations caused by dense concentrations of microscopic sea plants, known as phytoplankton. The system is an autonomous Case-Based Reasoning (CBR) hybrid system that embeds various artificial intelligence tools, such as case-based reasoning, neural networks and fuzzy logic in order to achieve real time forecasting. It predicts the occurrence of red tides caused by the pseudo-nitzschia spp diatom dinoflagellate near the North West coast of the Iberian Peninsula. Its goal is to predict the pseudo-nitzschia spp concentration (cells/liter) one week in advance, based on the recorded measurements over the past two weeks. The developed prototype is able to produce a forecast with an acceptable degree of accuracy. The results obtained may be extrapolated to provide forecasts further ahead using the same technique, and it is believed

that successful results may be obtained. However, the further ahead the forecast is made, the less accurate it may be.

In [3] an expert system called WPC-ES is presented for assisting departments of environmental management in their efforts to improve water quality in a city, applied in the Yellow River Basin of China. The system can analyze relationships between industrial water pollution and economic activities of industrial enterprises of a city. The system includes a decision model at its core, which integrates another four closely related subsystems. The results showed that the system could provide better decision support for environmental management.

Finally, in [12] a fuzzy expert system is presented for the determination of Water Quality Classification for Stream (WQCS) from uncertain and imprecise ecological information. The system employs 30 rules for WQCS determination, generated from a rule matrix of seven water quality grades, toxicity of water and rarity of cases. Results showed that smoothly varying curves of WQCS determination from the fuzzy expert system represented real-world experience more realistically than stepwise curves from a conventional expert system.

3. Description of the Andromeda Water Monitoring Network

The Andromeda network (Figure 1) is a network of sensors plunged into the Thermaikos Gulf that collects aquatic numeric data concerning sea water. After sensor readings are collected, they are transmitted to a main station for processing and storage. The network was developed by a scientific team led by the first author on behalf of the T.E.I. of Thessaloniki[3]. An outline of the system's functions is provided within the next sections. A more thorough description can be found in [7].
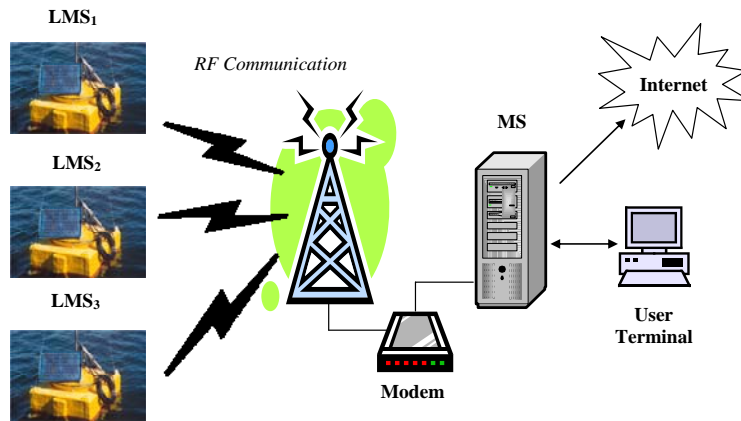
The network consists of:
- *Local Monitoring Stations (LMSs),* which record and transmit aquatic data to the main station.
- *Main Station (MS),* which initiates the communication process with all LMSs[4] and stores the data in the database for future processing.

---

[3] Technological Educational Institute of Thessaloniki

[4] In the current setting of Andromeda network, there are three (3) LMSs.

Figure 1. Architecture of the Andromeda network.

3.1. Local Monitoring Station

The *Local Monitoring Station* (*LMS*) (Figure 2) consists of:

1. A buoy that floats on sea surface;

2. A Programmable Logic Circuit (PLC) by Siemens;

3. Powerful Radio modems;

4. A six meter high pillar for the support of the antenna;

5. Four solar cells;

6. High-capacity rechargeable batteries.
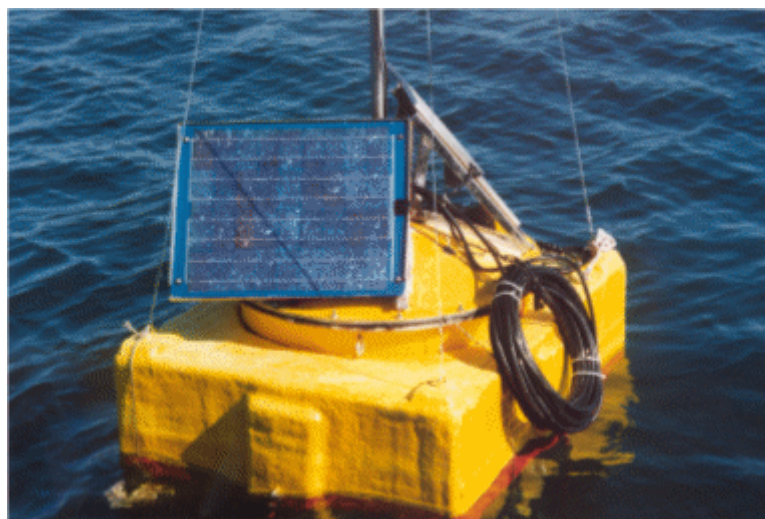
Figure 2. Andromeda's Local Monitoring Station.

The LMS incorporates sensors, batteries, solar cells, electronics and the PLC. The necessary power is provided by the batteries and solar cells to the sensors and the electronics.

The PLC is responsible for the LMS operation and storage of the measurements to the local memory. In predetermined time intervals, it transmits this data to the Main Station (MS), over a wireless network. The sensors measure the following hydrological parameters: *water temperature*, *pH*, *amount of dissolved oxygen (DO)*, *percentage of dissolved oxygen (DO %)*, *conductance*, *turbidity*, *sea currents*, and *salinity*.

Due to its design, the LMS can be easily updated and expanded. It responds well to extreme weather conditions, increased energy requirements, and meets the processing and storage needs for effectively monitoring the sea waters for the purpose of this research.
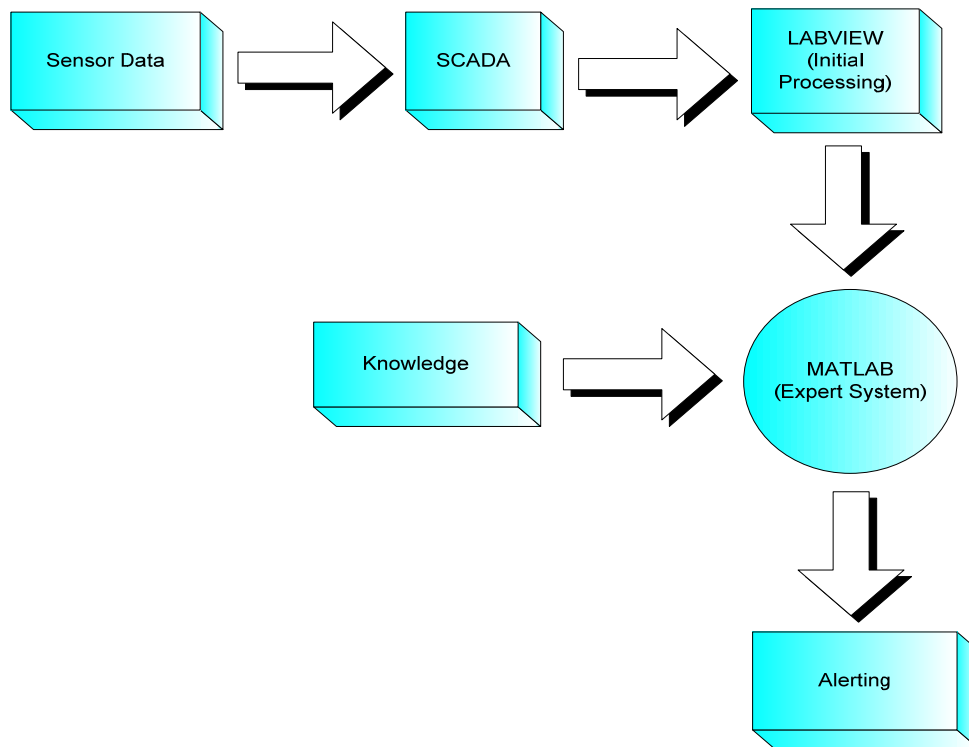
3.2. Main Station

The Main Station (MS) of the Andromeda network is a workstation that collects sensor measurements from all the LMSs and visualizes the results in a SCADA[5] environment [7]. The MS is the "master" of the communication process, i.e. it initiates the communication with each of the LMSs in predetermined time intervals using a hand-shake technique. The MS also adjusts the frequency of measurements depending on the situation at hand, i.e. an emergency in the case of pollution. The LMS operates only during the rendezvous. In this way, less energy is consumed. Furthermore, the on-demand measurement policy achieves a higher level of flexibility. The SCADA software exports the measurements in various formats for further processing and long-term storage in databases. Results are also available dynamically through a website[6]. The description of the system's data flow is depicted in Figure 3.

The data collection from the sensors and their transmission to the MS is performed via the SCADA software. LabView [16] is responsible for the data reception, visualization and storage at the MS. Additionally, the administrator of the MS can set alarms in LabView (to be received by the user) when certain sensor measurements exceed some predefined limits. However, the check is performed in a rigid fashion, not allowing flexibility. The WMES described in the next sections alleviates this by using fuzzy logic to check for violation of environmental parameter limits and alerting the user.

---

[5] Supervisory Control And Data Acquisition

[6] The website was operated internally on the intranet of the Thessaloniki Water Supply and Sewerage Company (details at http://www.eyath.gr/eng/2drastiriotites.htm) and its operation discontinued in 2005, along with the operation of the Andromeda network (see section 6).

Figure 3. Main Station's data flow, including the expert system.

4. Expert System Architecture

After visualizing the sensor measurements, data are processed by the expert system, which is implemented in MATLAB. The sensor values taken into account include water temperature, pH, conductance and other parameters. Depending on the specific values recorded, the system produces results about the water suitability regarding swimming, shell-culture and pisciculture, also alerting the user.

Expert scientific knowledge is required for these issues to be resolved. If a human expert is not available at the corresponding operation centre, then the hydrological parameters values should be monitored automatically in order to draw conclusions about possible hazardous situations for the environment.

The water monitoring expert system we have developed aims to overcome this drawback. More specifically:

1. it embeds the existing scientific knowledge about the *desired* and *allowed* parameter *values,* and according to this
2. it draws conclusions concerning potential danger.

The scientific knowledge required for the expert system was elicited from the Greek environmental legislation for the Region of Central Macedonia [18]. The desired and allowed values for the various measurements are set by the aforementioned law depending on the aquatic use: drinking water, swimming, shell cultivation, etc. This particular expert system deals with shell cultivation, swimming, Cyprinidae cultivation and Salmonidae cultivation.

4.1. *Input Data*

The data that are input to the expert system consist of the hydrological parameter values measured by the sensors and recorded by the LMSs. Recall that these values are transmitted to the MS and stored centrally in a database by LabView, the main software of the monitoring system. Since the expert system is encoded in MATLAB, there is an issue of data communication between LabView and MATLAB. LabView should encode data in such a way that MATLAB can interpret them. The encoding rules and conventions are described below.

The data sent from the LMSs to the MS include the following parameters:

* pH

* Temperature (^{o}C)

* Conductance (mSiemens/cm)

* Salinity (mg/l)

* Dissolved Oxygen (%)

* Dissolved Oxygen (Biochemical Oxygen Demand) (mg/l)

* Turbidity (FNU)

A timestamp including date and time is attached to the above measurements. Moreover, every LMS must send its voltage value as it was recorded during its operation, in order to check its normal operation. For each LMS there is a daily kept log that is updated every time a new tuple in the form {*Timestamp*, *Measurements*} arrives.

In Table 1 the sequence of measurements for each tuple is shown. An indicative example of such a tuple is:

```
25/7/2002 10:50 24.88 8.19 1.0005 3.0008 37.5 0.08 11.7 116.31
```
(*Data from LMS "Andromeda 3" as recorded on July 25, 2002.*)

Assume that *n* sessions per day occur between the LMSs and the MS. At the end of the day there are as many log files as the number of LMSs. Each one contains *n* tuples with 2+8=10 values per tuple. The sessions can vary in frequency and can differ from LMS to

LMS and/or per day. LabView can send out to MATLAB both complete daily log files and single measurement tuples.

Table 1. The columns of each file tuple.

Column	Measurement
1	Date
2	Time
3	Temperature
4	pH
5	Conductance
6	Salinity
7	$\%O_2$
8	Dissolved O_2
9	Voltage
10	Turbidity

4.2.   Expert Knowledge

The expert knowledge consists of the allowed range of values for the various hydrological parameters, as set by the Greek law. Table 2 shows the allowed values for fresh water aquatic uses, namely pisciculture. Table 3 shows the allowed values for sea water aquatic uses, namely shell cultivation and swimming. Of course, there are more hydrological parameters that affect these uses, but the current network sensors can only measure some of them. Therefore, only the conjunction of the set of factors that affect a specific use to the set of available measurements is taken into consideration.

Table 2. Desired and allowed values/ranges concerning pisciculture.

	Pisciculture – Fresh Water			
	Cyprinidae		Salmonidae	
	Desired Value/Range	Allowed Value/Range	Desired Value/Range	Allowed Range
Temperature	25	28	21.5	-
pH	6 .. 8.5	5.5 .. 9	6.5 .. 8.5	6 .. 9
Dissolved O_2	6	-	3	-

Table 3. Desired and allowed ranges concerning shell cultivation and swimming.

	Shell Cultivation		Swimming	
	Desired Range	Allowed Range	Desired Range	Allowed Range
Temperature	Normal	+1.5	-	-
pH	7 .. 8.5	7 .. 9	-	6.6 .. 8.5
Salinity	12 .. 38	≤40	-	-
%O$_2$	>80	70 .. 110	80 .. 120	-

When LabView calls on the WMES to decide whether the values are acceptable or not, it passes the file name as an argument. Then, the WMES scans the input file row-by-row and compares each value recorded to the respective allowed values. Different results are produced for each row and for each aquatic use. For instance, if the system was operating on July 25th, 2002, at 10:50 a.m., it would detect that shell cultivations would be in danger, whereas the rest three aquatic uses water would be safe.

A simplistic solution would be to test whether the values in each line fall exactly into the predetermined range. If it was to be found that one value was out of the desired or allowed range a warning message should appear on screen. Suppose that pH equals 5.52. According to the data for Cyprinidae cultivation, the aforementioned value would be out of the desired range, but within the allowed range. However, the value 5.52 is very close to 5.5 which is the minimum acceptable value; therefore, for such a delicate issue that could possibly affect public health, we believe that range limits should be considered in a fuzzy and not in a crispy manner. In our expert system we have represented the expert knowledge using fuzzy logic rules [21], [1].

4.3. Architecture of the Water Monitoring Expert System (WMES)

The WMES consists of two main components. The first is a set of Fuzzy Inference Systems (FISs) [14] and the second is an m-file (i.e. a MATLAB program [15]) that employs the FISs and combines their results in order to draw the final conclusions. The description of these two components follows.

There is one FIS for each of the following hydrological parameters:

- Temperature

- pH

- Dissolved O$_2$

- Salinity

- $\%O_2$

Each FIS has one input variable and four output variables, one per aquatic use. The use of multiple single-input-multiple-output (SIMO) FISs instead of one multiple-input-multiple-output (MIMO) FIS is justified by the fact that the legislation correlates non-acceptable parameter values with a single water usage, i.e. it regards each variable as independent from the other ones. Therefore, there would be an unnecessary complication to create a single MIMO FIS, since variables are independent from each other. Using multiple SIMO FISs, instead, reduces the complexity of the transition function.

When an aquatic use is not affected by a specific parameter, then the corresponding output variable contains a special membership function named NULL. The rationale for this solution is given in section 5.

The input variable is split in as many membership functions as needed to utilize the desired and allowed ranges for each aquatic use. The output variables affected by this particular measurement contain three membership functions: OK, PROSOXH[7] and KINDYNOS[8] (regarding swimming) and OK, PROSOXH and KATASTROFH[9] (for the rest three aquatic uses).

In each FIS, the following three rules hold:

- If the input falls within the desired range for X uses, the respective output for each one of the X uses is OK.

- If the input is outside the range of desired values, but remains within the allowed range for Y uses, the corresponding output for each one of the Y uses is CAUTION.

- If the input is outside the range of the allowed values for Z uses, then the corresponding output for each one of the Z uses becomes DANGER (or CATASTROPHE).

For each input data tuple, the m-file executes every FIS, passing the corresponding value as an input parameter. For example, for the input tuple of section 4.1 it will execute the FIS for pH, having as input value `8.19`, the FIS for temperature with input value `24.88`, etc. After this stage is completed, the results for each aquatic use are derived and must be communicated to the user. When all aquatic uses are within normal/desired value ranges, then no warnings are issued and the user receives an OK message. If, on the other hand, the

---

[7] "Caution" in Greek.

[8] "Danger" in Greek.

[9] "Catastrophe" in Greek.

conclusion for at least one of the aquatic uses is DANGER or CATASTROPHE, then the user is warned. Different such warnings are issued for each different use.

4.4. Presenting Results to the User

For each aquatic use that the water is found not suitable, MATLAB sends a numerical identifier to LabView which denotes the use, and a string containing the display message. LabView is responsible for alerting the user.

The identifiers are:

1. Shell cultivation
2. Swimming
3. Cyprinidae cultivation
4. Salmonidae cultivation

The possible messages (along with the corresponding numerical identifiers) are the following[10]:

1. '*CAUTION: The desired value ranges have been exceeded!*'
2. '*DANGER: The allowed value ranges have been exceeded!*'

The output is in the form of a two-dimensional array that consists of as many rows as the number of hazardous uses. The first column contains the use identifier and the second one contains the message identifier. An example follows:

```
[1 1; 2 1; 4 2]
```

which is the MATLAB representation of the two-dimensional array:

```
1     1
2     1
4     2
```

Moreover, each use-message pair is accompanied by a vector with five cells. Each cell corresponds to an FIS, with the following order: Temperature, pH, Salinity, Dissolved O_2 and %O_2). If the ith cell is set to zero (0), the corresponding value is normal. If it is one (1), the corresponding value violates the desired range. Finally, the value two (2) means that the allowed range was exceeded. Consequently, the output has the following format:

```
1     1     0     1     0     1     0
2     1     0     1     0     0     0
4     2     0     2     0     1     0
```

---

[10] Only the English translation is shown here.

In the table above, the first row indicates an orange alert for shell cultivation, due to pH and dissolved O_2 measurements being out of desired range.

5. Expert System Implementation

This section presents how the functionality of the WMES, presented in section 4.3, is implemented in MATLAB. The first issue to be dealt with is the transformation of the expert knowledge that is represented through a set of value ranges, into a set of Fuzzy Inference Systems (FIS). Subsequently, the implementation of the main WMES module in MATLAB is presented.

5.1. Knowledge Representation via Fuzzy Inference Systems

In this section we analyze how a MATLAB Fuzzy Inference System is developed out of the allowed value ranges for the hydrological parameter laid down by the Greek legislation. As an example we use the water acidity value (i.e. pH). The rest of the FISs are developed in a similar manner. The pH.FIS has one input variable, the pH sensor measurement, and four output variables (Ostrakoeidh[11], Kyprinidwn[12], Salmonidwn[13], Kolymbhsh[14]), i.e. the decision about the water suitability for four possible aquatic uses: Shell/Cyprinidae/Salmonidae cultivation and Swimming, respectively.

5.1.1. Input Variable

The desired and allowed value ranges for pH values are shown in Table 4. Notice, that when there is an allowed range for a variable but there is no desired range, then we could consider that the desired range is equal to the allowed range (e.g. for the swimming case). However, we did not include cases where the allowed and desired ranges coincide, for efficiency reasons. If the desired and allowed ranges coincide, then the corresponding alert would be at the same time yellow and red. Since red alert is always superior to yellow alert, the latter would never be triggered. Thus, having an extra rule to the system that would never fire, would be an unnecessary burden.

---

[11] "Shell" in Greek.

[12] "Cyprinidae" in Greek.

[13] "Salmonidae" in Greek.

[14] "Swimming" in Greek.

Table 4. Desired and allowed value ranges for pH values.

	Desired Range	Allowed Range
Shell Cultivation	7 .. 8.5	7 .. 9
Swimming	-	6.6 .. 8.5
Cyprinidae	6 .. 8.5	5.5 .. 9
Salmonidae	6.5 .. 8.5	6 .. 9

The next task is to specify the number and form of the membership functions of the input variable. Each one of the ranges shown in Table 4 defines a sub-range of the pH domain, i.e. [0, 14]. The first step is to determine the intervals that these areas overlap and to assign a membership function to each one of these intervals.

A careful study of Table 4 results in the following intervals:

1. *5.5 to 6:* When the pH value varies between 5.5 and 6, it is beyond the acceptable range for Swimming, Shell cultivation and Salmonidae cultivation. However, Cyprinidae cultivation is not affected. We name the corresponding membership function "OK".

2. *6 to 6.5:* This range is not acceptable for Shell cultivation and Swimming. As it concerns pisciculture, Cyprinidae cultivation is not affected. However, when the pH varies between these two marginal values, it is out of the desired range for Salmonidae cultivation, but within the respective allowed range. The corresponding membership function is named "OK1".

3. *6.5 to 7:* This value range is considered normal for both kinds of pisciculture and Swimming (marginally), but it is prohibitive for Shell cultivation (membership function "OK2").

4. *7 to 8.5:* This value range is normal for all four aquatic uses, i.e. within the respective desired ranges (membership function "OK3").

5. *8.5 to 9:* This interval consists of values that violate the desired value range of Shell, Cyprinidae and Salmonidae cultivation, and exceeds the maximum allowed value for Swimming (membership function "OK4").

The five membership functions above do not cover the ranges [0, 5.5] and [9, 14]. Four additional functions are defined in order to fill this gap. When the pH value falls within these ranges, all four aquatic uses are prohibited.

6. *0 to 3.5:* Membership function "Poly-Oksino" (strongly acidulous).

7. *3.5 to 5.5:* Membership function "Oksino" (acidulous).

8. *9 to 11:* Membership function "Basiko" (basic).

9. *11 to 14:* Membership function "Poly-Basiko" (strongly basic).

It should be made clear that all the above ranges are not crispy but fuzzily defined. In other words, the membership functions are overlapping, in order to allow for fluctuations in sensor readings, due to either limited sensor accuracy or random fluctuations of physical conditions. Furthermore, overlapping membership functions can meet ascending or descending parameter trends, in the absence of proper trend analysis within the system. For example, if there is a descending trend for the pH that would turn it from normal to acidulous in a few hours, the overlapping of membership functions would cause an alert before the pH actually outreaches the critical point. This means that authorities will be alerted in time to be prepared for action, when the actual critical situation arises.

A final justification for using fuzzy logic is that the latter can cater for the rigidness imposed by the legislation, which had to be formulated precisely, using crispy values. However, in practice such limits are never rigidly defined, since they are derived statistically. Notice that the overlapping intervals are narrow, since the legislation cannot be highly disputed, meaning that measures against pollution and/or polluters cannot be taken if the actual crispy limits set by the law are not violated. Having large overlaps would cause more pollution alarms that would trigger authorities unnecessarily. In any case, the degree of overlapping was actually established through experimentation and it was tuned for smooth behaviour of the FIS transition function.

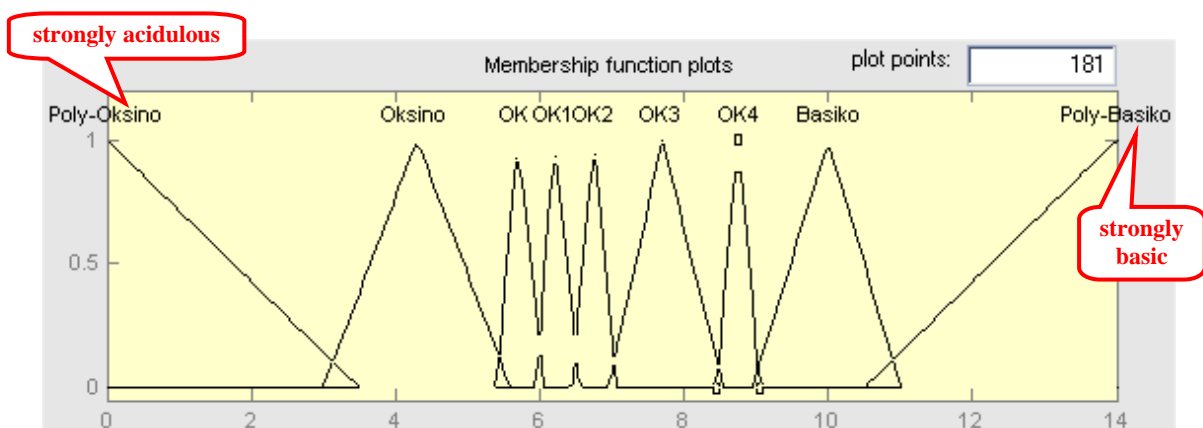**Figure 4.** The membership functions of the input variable "pH".

The membership functions of the pH variable are graphically illustrated in Figure 4. The rationale for choosing triangular membership functions are as follows. First, we needed a membership function with a peak, e.g. a triangular or a Gaussian one, since we deal with value limits, represented as fuzzy numbers. Actually, the shape of the membership function

was established through experimentation and it was tuned for smooth behaviour of the FIS transition function. We have used triangular membership functions instead of Gaussians because the former are simpler, they are calculated faster (efficiency) and they have produced a smoother transition function.

5.1.2. Output Variables

Each of the four output variables has been assigned three membership functions:

- OK, CAUTION and DANGER, to Swimming, and

- OK, CAUTION and CATASTROPHE, to the other three aquatic uses.

The membership functions are defined over three arbitrarily selected intervals. The first interval is [0, 0.33), the second is [0.33, 0.66) and the third is [0.66, 1]. OK is defined over the first interval, CAUTION over the second, and DANGER (or CATASTROPHE) over the third. Some overlapping occurs here, as well (see Figure 5).
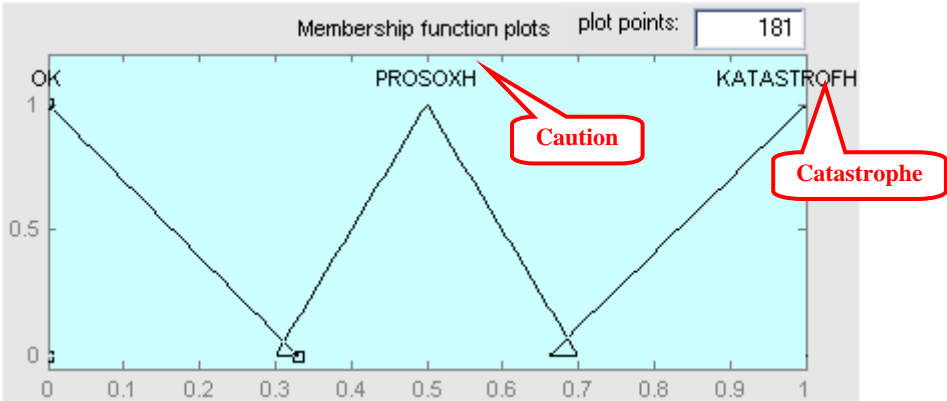

Figure 5. The membership functions of the output variable "Shell".
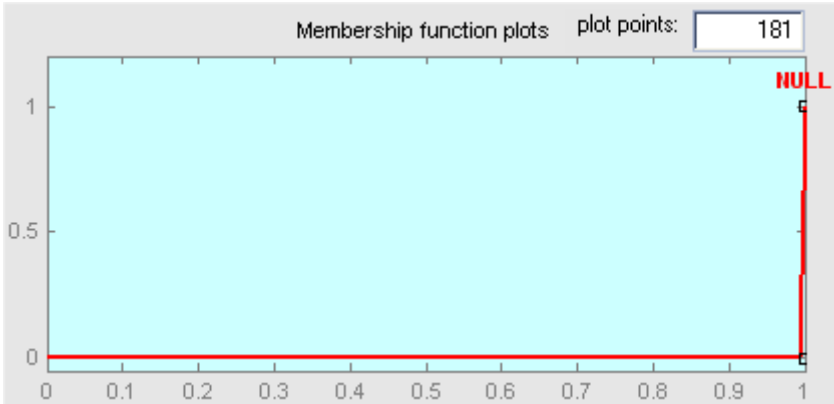
Figure 6. The membership function "NULL" (*Params* = [1 1 1]).

If an aquatic use is not affected by the input variable of a particular FIS, then its corresponding output variable has only one membership function. This function is named NULL and is depicted in Figure 6.

5.1.3. Rule Definition

The rules for this particular FIS were formulated by taking into account: a) the correlation between the input and the output variables, and b) the definitions of the membership functions.

The rules were created via MATLAB's FIS editor and the result is shown in Figure 7. Notice that the Mamdani's min fuzzy inference method [13] and the centroid defuzzification method were used.

Figure 7. The set of rules of pH.fis.

5.2. The Main WMES Module

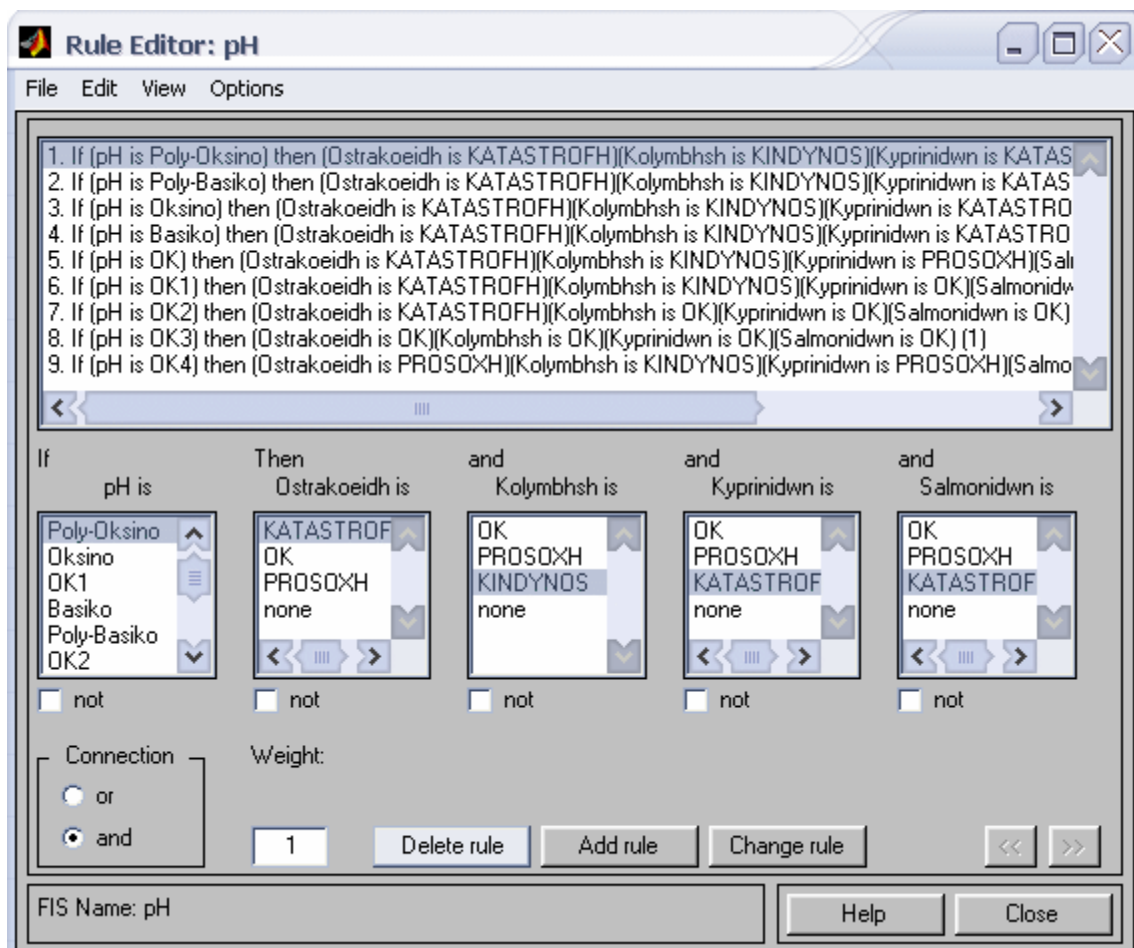The main WMES module was implemented as a MATLAB m-file. The main module performs the following tasks: processing of the input data, inferencing, displaying the results to the user and/or LabView.

5.2.1. Important Data Structures

In this sub-section we describe various important data structures that have been used within the main WMES module.

Initially, the sensor readings (input data) are imported from an Excel file and are stored in the two-dimensional array *Values*. Each row of this array corresponds to a different vector of input data, whereas each column corresponds to a different sensor measurement (Table 1).

The output of each FIS is a one-dimensional array with four cells. Each shell corresponds to a specific aquatic use (e.g. 1 is for shell cultivation, 2 is for swimming, etc). The numerical values for each aquatic use correspond to symbolical constants. A similar approach was followed for the alert types; desired value range violation is alert type 1 (orange alert), while maximum allowed value range violation is alert type 2 (red alert).

A set of arrays containing string constants are also used. Three arrays are defined for the names of the FISs, uses, and measurements. A fourth array contains the alert messages.

The following points are worth mentioning:

- The definition of non-existent FISs is included to cater for future extensions.

- At first, the strings are stored in a `cell` type array (MATLAB data type) for convenience of initialization. Afterwards, this array is converted to a `struct` type array using the built-in function `cell2struct()`. Each `struct` variable contains a field named `str`. The original `cell` array becomes obsolete and it is deleted.

- The `fisName` array is sorted in a predetermined order, derived from the numerical constant definitions. The expression `fisName(nMEASUREMENT).str` returns the name of the corresponding FIS (i.e. `fisName(nTEMP).str` ≡ `'temp.fis'`) due to this sorting.

- An array of `struct` variables is preferred over one consisting of `cell` variables, because in this manner the expression `readfis(fisName(nTEMP).str)` results in `readfis('temp.fis')`. On the other hand, the expression `readfis(cellfisArray(nTEMP))` would be invalid.

The same approach is followed for the definition of the arrays that contain the alert messages, the names of the aquatic uses and the measurement types.

Finally, one of the most important arrays in the system is a 4x9 array named *FisPerUse*, which consists of bits. Each row in this array corresponds to a specific aquatic use, while each column corresponds to the respective column of the input file (sensor readings). During the initialization process, all array elements are set to zero. The (i, j) cell is set to one if, and only if, the ith use is affected by the jth measurement. A sample instance of the *FisPerUse* array is shown in Table 5.

Table 5. Sample instance of the *FisPerUse* array.

	1	2	3	4	5	6	7	8	9
1	0	1	1	0	1	0	1	0	0
2	0	0	1	0	0	0	1	0	0
3	0	1	1	0	0	1	0	0	0
4	0	1	1	0	0	1	0	0	0

In the example above, the levels of pH and the percentage of dissolved oxygen affect swimming. Hence, in order to derive a conclusion about the water suitability for swimming, only the values of these two dimensions need to be examined.

During the execution of the main WMES module, the following three arrays play a significant role:

1. *Results*
2. *PartialConclusions*
3. *FinalConclusions*

Results: When a new vector of sensor readings becomes available it is fed to the FISs. The fuzzy inference mechanisms draw the first results concerning the four aquatic uses. The results are stored in this array in the following fashion. Each row corresponds to a particular aquatic use and each column to a specific FIS (i.e. this is a 4x9 array). The (i, j) cell stores the output of the jth FIS regarding the ith aquatic use. Table 6 shows a sample instance of this array.

Table 6. Sample instance of the *Results* array.

	1	2	3	4	5	6	7	8	9
1	0	0.8396	0.1261	0	0.5	0.8703	0	0	0
2	0	0	0.1261	0	0	0.5	0	0	0
3	0	0.8396	0.1261	0	0	0	0.5	0	0
4	0	0.5	0.1261	0	0	0	0.5	0	0

PartialConclusions: This array follows the same format as the previous one. In fact, it constitutes a mapping of the decimal values in table *Results* to integer values drawn from the domain {0,nCAUTION,nDANGER}. Initially, all the array elements are set to zero. The value nCAUTION (i.e. 1) is stored in the (*i*, *j*) cell, if the respective value of the *Results* array is greater than or equal to 0.33 and less than 0.66. On the contrary, if this value is greater than or equal to 0.66, the value nDANGER (i.e. 2) is stored. Table 7 shows a sample instance of this array.

Table 7. Sample instance of the *PartialConclusions* array.

	1	2	3	4	5	6	7	8	9
1	0	2	0	0	1	2	0	0	0
2	0	0	0	0	0	1	0	0	0
3	0	2	0	0	0	0	1	0	0
4	0	1	0	0	0	0	1	0	0

FinalConclusions: This array is a vector consisting of four cells that correspond to each of the four aquatic uses. The values of this vector are also drawn from the domain {0, nCAUTION,nDANGER}. If all the non-zero values of the corresponding row in the array *PartialConclusions* are equal to nCAUTION, then nCAUTION is stored in the ith cell of the vector *FinalConclusions*. On the other hand, if at least one of the cells in the ith row of the array *PartialConclusions* is assigned the nDANGER value, the same value is stored in the respective cell of the array *FinalConclusions*. Otherwise, the cell retains its initial value which is zero.

To sum up, the 'per use' and 'per measurement' output data of the FISs is stored in the *Results* array. These data are converted into information about the violation, or not, of certain limits, that is stored in the *PartialConclusions* array. Finally, these initial conclusions are reduced to the final 'per use' conclusions and are stored in the *FinalConclusionsvector*.

5.2.2. *Architecture and Functionality*

The main loop of the main module is executed once per row of the *Values* array. If the input file is a daily log with 24 recorded sets of measurements (a new set is added every hour), then 24 iterations are required. However, LabView can also be set to execute the expert system every time it receives data from the LMSs, i.e. once per hour. In any case, for each row, the results, initial conclusions, and final conclusions are elicited, and, finally, appropriate messages appear on the screen, if necessary. These steps are presented in detail in this sub-section.

Initially, each set of measurements must be analyzed to test its suitability for each aquatic use. The algorithm that calculates the results and extracts the initial conclusions is presented in Figure 8. Notice that, a value that is considered normal for the ith use, might be unacceptable for the jth use. Moreover, each use is affected by different measurements, hence only the corresponding FISs should be called on.

```
for use = nSHELL:nSALMONIDAE
    for measurement = nDATE_TIME:nTURB
        if (FisPerUse(use, measurement) == 1) & ...
                (Results(use, measurement) == 0)
            fis = readfis(fisName(measurement).str);
            out = evalfis(VALUES(Row, measurement), fis);
            for x=nSHELL:nSALMONIDAE
                if FisPerUse(x, measurement)
                    Results(x, measurement) = out(x);
                end
                CheckedValue = Results(x, measurement);
                if CheckedValue < 0.33
                elseif CheckedValue < 0.66
                    PartialConclusions(x, measurement) = nCAUTION;
                elseif CheckedValue >=0.66
                    PartialConclusions(x, measurement) = nDANGER;
                end
            end
        end
    end
end
```

Figure 8. Algorithm for result calculation and partial conclusion extraction.

Thus, two nested loops are used (Figure 8). The outer loop is executed once per aquatic use, while the inner loop is executed once per measurement type. The current use is denoted by variable `use`, while the current measurement type by the variable `measurement`. At the beginning of every inner loop, the FIS named `fisName(measurement).str` is executed using the built-in MATLAB functions `readfis` and `evalfis`. However, in order for

this to happen, two conditions must first be met. The first condition ensures that the current use is affected by the current measurement type, i.e. `FisPerUse(use,measurement)==1`. The second condition checks if this is the first time this particular FIS is executed, i.e. `Results(use,measurement)==0`.

The output vector consists of four cells. Each cell is assigned values between 0 and 1. A sample instance of this vector is: [1.0000, 1.0000, 0.1101, 0.5000]. Every output variable that is not affected on the input variable of the current FIS is set to `NULL` (`NULL` becomes 1 during the translation process). The membership function `NULL` was selected for a number of reasons listed below:

- If no membership function was selected for these variables, many annoying warning messages would be displayed on the screen.

- Should these variables be removed altogether, the one-to-one mapping between cell values and aquatic uses would no longer be valid.

- The value 1 does not occur anywhere else.

Upon the execution of the appropriate FIS, each cell of column `measurement` in the array *FisPerUse* is examined. The value `out(x)` is assigned to cell `(x,measurement)` in the array *Results*, if the corresponding cell in *FisPerUse* is set to 1. This is an optimization, since if another aquatic use depends on the results of this particular FIS, these would already be stored in the array *Results* and the respective initial conclusions would be also stored in the *PartialConclusions* array. Hence, each FIS is executed at most once.

Each new entry in the *Results* array is followed by another entry in the *PartialConclusions* array. The reasoning behind this correlation has already been analysed.

```
for use = nSHELL:nSALMONIDWN
    for measurement = nDATE_TIME:nTURB
        if (PartialConclusions(use, measurement) == nCAUTION) & ...
              (FinalConclusions(use) < nCAUTION)
           FinalConclusions(use) = nCAUTION;
        elseif (PartialConclusions(use, measurement) == nDANGER) & ...
                 (FinalConclusions(use) < nDANGER)
           FinalConclusions(use) = nDANGER;
        end
    end
end
```

Figure 9. Algorithm for final conclusion extraction.

The next step involves the drawing of the overall conclusions. Figure 9 shows the algorithm for drawing the appropriate values for the *FinalConclusions* array. The need for 'per use' conclusions came up from a basic user requirement, which insisted that at any given point, the system must be able to provide answers to questions like the following: "*Is the water suitable for swimming?*"

A simple report containing the values violating the predetermined limits could not possibly be considered as a user-friendly answer, even if the nature (desired or acceptable) of the violated limits is included. The solution that was adopted was the assignment of a general alert code to each aquatic use. At the end, what the user reads on his/her screen is a combination of the aquatic use, the alert type, and the precise measurements that were deemed abnormal (Figure 10). The alert type is designated by the array *FinalConclusions* that was explained in the previous sub-section, along with its correlation with the array *PartialConclusions*.

After the previous phase ends, various informative messages are displayed on screen. In particular, if the values are deemed normal for all uses, no messages are displayed. On the other hand, if a violation is detected, the following messages appear on screen, one after the other:

1. Date and time of the violation occurrence.
2. The name of the first aquatic use that the water was found unsuitable for.
3. The alert type elicited from the table *FinalConclusions*.
4. The names of the measurements that exceeded the determined range, as well as the specific values measured (e.g. `PH: 7.9932`). This information is extracted from the arrays `MeasurementName`, *PartialConclusions* and *Values*.
5. The output of the FIS that corresponds to the abnormal value for the current use. This output is stored in the *Results* table.

Steps 2 through 5 are repeated for each aquatic use that was found problematic. An example of the output of the expert system is shown in Figure 10.

In case the alert type for the xth use is `nCAUTION`, only the desired range has been violated. Therefore, just the values that exceeded this range need to be displayed on the screen. On the other hand, if the alert is a red one (`nDANGER`), the allowed range has been breached and, consequently, the values that exceeded this range should be displayed. However, there is a possibility that some other values were out of the desired range, but within the allowed range for the current aquatic use. In that case, those values should be displayed as well. In the example shown above, four values were considered abnormal

concerning shell cultivation. Three of them were out of the acceptable range; hence, the alert type for this use was deemed "DANGER" and one was out of the desired value range, but within the acceptable one. The latter is displayed last.

```
<<<DATE / TIME:        25-Jul-2002 / 06:50>>>

SHELL CULTIVATION:
        DANGER: The allowed value ranges have been exceeded!
                Values that have exceeded the ALLOWED VALUE RANGE are:
                TEMPERATURE:    25.390      Code:   0.8613494
                         PH:    11.180      Code:   0.8485162
                        %O2:    37.100      Code:   0.8691869
                Values that have exceeded the DESIRED VALUE RANGE are:
                   SALINITY:     3.000      Code:   0.5000000

SWIMMING:
        DANGER: The allowed value ranges have been exceeded!
                Values that have exceeded the ALLOWED VALUE RANGE are:
                         PH:    11.180      Code:   0.8485162
                Values that have exceeded the DESIRED VALUE RANGE are:
                        %O2:    37.100      Code:   0.5000000

CYPRINIDAE CULTIVATION:
        DANGER: The allowed value ranges have been exceeded!
                Values that have exceeded the ALLOWED VALUE RANGE are:
                         PH:    11.180      Code:   0.8485162
                Values that have exceeded the DESIRED VALUE RANGE are:
                TEMPERATURE:    25.390      Code:   0.5000000
                DISSOLVED O2:    0.080      Code:   0.5000000

SALMONIDAE CULTIVATION:
        DANGER: The allowed value ranges have been exceeded!
                Values that have exceeded the ALLOWED VALUE RANGE are:
                         PH:    11.180      Code:   0.8485162
                Values that have exceeded the DESIRED VALUE RANGE are:
                TEMPERATURE:    25.390      Code:   0.5000000
                DISSOLVED O2:    0.080      Code:   0.5000000
```

Figure 10. Sample expert system output.

6. Conclusions – Future Work

In this paper we have presented an expert system that monitors sea and/or fresh water quality and pollution in Northern Greece, through the Andromeda sensor network. The expert system

on-line monitors sensor data collected by Local Monitoring Stations and reasons about the current level of water suitability for various aquatic uses, such as swimming and piscicultures. The purpose of the expert system is to help the authorities in the "decision-making" process in the battle against the pollution of the aquatic environment, which is very vital for the public health and the economy of Northern Greece. The expert system determines, using fuzzy logic, when certain environmental parameters exceed certain "pollution" limits, which are specified either by the authorities or by environmental scientists, and issue appropriate alerts.

The main advantage of the system and its architecture is its versatility by means of extensibility and mobility. Concerning the sensor network, new sensors for a variety of environmental readings (e.g. hydrological, meteorological, etc.) can be and have been easily added to the system. Furthermore, existing LMSs can be easily moved to different locations and new LMSs can be easily added, without disturbing the rest of the system. The communication between the LMSs and the MS can be and has been implemented with a variety of technologies, depending on the geomorphologic and socioeconomic features of the installation area.

The WMES expert system is equally flexible and extensible, since its architecture allows for effortless integration of new sensor readings, as each sensor corresponds to a new variable and a new FIS that is independent from the existing ones. Furthermore, new aquatic uses, whose hazardous situations are determined through the new sensors, can be easily added to the system by simply inserting the new use to the existing parameter FISs with a NULL membership function (see sections 4.3 and 5.1.1). Notice that the addition of complex MIMO FISs could again be modelled by having the same output behaviour for all aquatic uses.

The Andromeda network was working productively from 1998 until 2005, when it ceased working due to lack of Governmental funding. The expert system has been working 18 months in conjunction with the pre-existing simple monitoring system of the MS, during 2004-2005. During that period "events" of pollution were regularly recorded, since the LMSs were installed near the port and the industrial area of Thessaloniki, where sea quality is very poor. However, only one really important and severe event was recorded when all sensors indicated strongly that something wrong was happening due to a spill from a ship. Finally, there were several false and/or missed alarms when the maintenance of sensors was inappropriate, thus their indications were not trustful, or when littering was cluttering sensor readings.

The use of fuzzy, instead of classical Boolean, logic is justified by the fact that sensor readings fluctuate constantly, due either to limited accuracy or random fluctuations of the

physical conditions. Therefore, fuzzy logic achieves a more stable behaviour than classical logic, which would constantly change the system's conclusion about water suitability if the sensor measurements would fluctuate around a pollution limit.

Another reason for using fuzzy logic is that the legislation pollution limits are crispy, simply because laws have to be defined precisely. However, in practice such limits are never rigidly defined, since they are derived statistically.

Finally, fuzzy logic can meet ascending or descending hydrological parameter trends, in the absence of trend analysis within the system. The latter is one of the future issues to be implemented within the WMES. We are already working on the prediction of the hydrological parameter values based on the time-sequence analysis of the already observed values [9]. The next step is to integrate these prediction algorithms within the WMES, so that the WMES will be able to issue early warnings based on predicted hydrological parameters values.

Another point for extending the system is to integrate with other water monitoring networks, such as "Interisk" [11], which monitors fresh waters, namely Lake Doerani, Lake Kerkini and the Strymonas River. The Local Monitoring Stations of "Interisk" collect also meteorological data, namely air temperature and humidity, wind direction and velocity, sunlight, evaporation and rainfall. Such data could be combined with hydrological data through more complex FISs (with many input data) developed by more complex knowledge elicited from experts, rather than legislation.

Finally, one of our future aims is the use of model-based reasoning for self-diagnosing the sensors of the Local Monitoring Stations. For example, it is known that pH and conductivity parameters depend linearly on each other. If sensor readings indicate a converse relationship between these parameters, then a possible cause could be the malfunctioning of at least one of the two sensors.

Acknowledgements

7. References

[1]   Abraham Kandel, ed., *Fuzzy Expert Systems*, CRC Press, 1991.

[2] Bing S.L., Bullister J., Murray J., Sonnerup R., "Anthropogenic chlorofluorocarbons in the Black Sea and the Sea of Marmara", *Deep Sea Research Part I: Oceanographic Research Papers*, 49(5), pp. 895-913, 2002.

[3] Cheng H., Yang Z., Chan C.W., "An expert system for decision support of municipal water pollution control", *Engineering Applications of Artificial Intelligence*, Vol. 16, No. 2, Special Issue on "Applications of Artificial Intelligence for Management and Control of Pollution Minimization and Mitigation Processes", pp. 159-166, 2003.

[4] Fdez-Riverola F., Corchado J.M., "CBR based system for forecasting red tides", *Knowledge-Based Systems*, Vol. 16, pp. 321–328, 2003.

[5] Fdez-Riverola F., Corchado J.M., "FSfRT: Forecasting System for Red Tides", *Applied Intelligence*, Vol. 21, pp. 251–264, 2004.

[6] Gross Kopf W.G., Kraus N.C., Militello A., Bocamazo L.M., "Implementation of a regional wave measurement and modeling system, south shore of Long Island, New York", *Proc. Int. Symposium on Ocean Wave Measurement and Analysis*, Vol. 1, pp. 610-619, 2001.

[7] Hatzikos E.V., "A fully automated control network for monitoring polluted water elements", in Greek, *Proceedings of the 4th Conference on Technology and Automation*, October 2002, Thessaloniki, Greece, pp. 443-448.

[8] Hatzikos E.V., "Surface and Underground Water Control for the Region of Central Macedonia", in Greek, *IRIS-INTERISK Project Report*, August 2001, Thessaloniki, Greece.

[9] Hatzikos E.V., Anastasakis L., Bassiliades N., Vlahavas I., "Applying neural networks with active neurons to sea-water quality measurements", *Proc. 2nd Int. Scientific Conf. on Computer Science*, Chalkidiki, Greece, 30th Sep-2nd Oct 2005, IEEE Computer Society, Bulgarian Section, pp. 114-119.

[10] Hendee J.C., "An expert system for marine environmental monitoring in the Florida Keys National Marine Sanctuary and Florida Bay", in the *Proc. 2nd Int. Conf. on Environmental Coastal Regions*, ed. C.A. Brebbia, Computational Mechanics Publications/WIT Press, Southampton, pp. 57-66, 1998.

[11] Interisk project, "Interregional telematic network for the management of Risks in the Balkan and South Mediterranean area", `http://interisk.imbc.gr/`

[12] Lee H.K., Oh K.D., Park D.H., Jung J.H., Yoon S.J., "Fuzzy expert system to determine stream water quality classification from ecological information", *Water Science and Technology*, Vol. 36, No. 12, pp. 199-206, 1997.

[13] Mamdani E.H., Assilian S., "An experiment in linguistic synthesis with a fuzzy logic controller", *International Journal of Man-Machine Studies*, Vol. 7, No. 1, pp. 1-13, 1975.

[14] MathWorks, *Fuzzy Logic Toolbox User's Guide*, 2005. available at: http://www.mathworks.com/access/helpdesk/help/pdf_doc/fuzzy/fuzzy.pdf

[15] MathWorks, *MATLAB Programming*, 2005. available at: http://www.mathworks.com/access/helpdesk/help/pdf_doc/matlab/matlab_prog.pdf

[16] National Instruments, LabVIEW. available at: http://www.ni.com/labview/

[17] Oprea M., "A Knowledge-Based Environmental Protection System", *Proc. 1ˢᵗ Balkan Conf. in Informatics*, Thessaloniki, Greece, 21-23 November 2003, pp. 67-78.

[18] Region of Central Macedonia, Directorate of Environment, *Environmental Legislation*, May 1999.

[19] Saunders M.C., Sullivan T.J., Nash B.L., Tonnessen K.A., Miller B.J., "A knowledge-based approach for classifying lake water chemistry", *Knowledge-Based Systems*, Vol. 18, pp. 47–54, 2005.

[20] Sullivan T.J., Saunders M.C., Tonnessen K.A., Nash B.L., Miller B.J., "Application of a regionalized knowledge-based model for classifying the impacts of nitrogen, sulfur, and organic acids on lake-water chemistry", *Knowledge-Based Systems*, Vol. 18, pp. 55–68, 2005.

[21] Zadeh L.A., "Fuzzy Logic," *Computer*, Vol. 1, No. 4, pp. 83-93, 1988.