

PREDICTIVE MODELS FOR MONITORING AND ANALYSIS OF THE TOTAL ZOOPLANKTON

Milica Obradović^{1*}, Ivana Radojević¹, Aleksandar Ostojić¹ and Nenad Stefanović²

¹*Institute of Biology and Ecology*, ²*Institute of Mathematics and Informatics, Faculty of Science, University of Kragujevac, Radoja Domanovića 12, 34000 Kragujevac, Republic of Serbia.*

*Corresponding author, e-mail: milica-obradovic@live.com

(Received April 1, 2014)

ABSTRACT. In recent years, modeling and prediction of total zooplankton abundance have been performed by various tools and techniques, among which data mining tools have been less frequent. The purpose of this paper is to automatically determine the dependency degree and the influence of physical, chemical and biological parameters on the total zooplankton abundance, through design of the specific data mining models. For this purpose, the analysis of key influencers was used. The analysis is based on the data obtained from the SeLaR information system – specifically, the data from the two reservoirs (Gruža and Grošnica) with different morphometric characteristics and trophic state. The data is transformed into optimal structure for data analysis, upon which, data mining model based on the Naïve Bayes algorithm is constructed. The results of the analysis imply that in both reservoirs, parameters of groups and species of zooplankton have the greatest influence on the total zooplankton abundance. If these inputs (group and zooplankton species) are left out, differences in the impact of physical, chemical and other biological parameters in dependences of reservoirs can be noted. In the Grošnica reservoir, analysis showed that the temporal dimension (months), nitrates, water temperature, chemical oxygen demand, chlorophyll and chlorides, had the key influence with strong relative impact. In the Gruža reservoir, key influence parameters for total zooplankton are: spatial dimension (location), water temperature and physiological groups of bacteria. The results show that the presented data mining model is usable on any kind of aquatic ecosystem and can also serve for the detection of inputs which could be the basis for the future analysis and modeling.

Keywords: abundance of total zooplankton; predictive model; key influencers.

INTRODUCTION

The freshwater zooplankton include representatives from the Protozoa, the Rotifera, and the Crustacea, as well as some less common but still widespread and often important members from such groups as the Insecta. Zooplankton consists of herbivorous, carnivorous, or perhaps most frequently, omnivorous animals. They make up one to several trophic levels in lake ecosystems (LIKENS, 2010; SUTHERS and RISSIK, 2009):

- Their role as herbivores has been particularly well studied (effects of zooplankton grazing on reducing algal abundance);
- They play important role in ‘grazing chain’ and the ‘microbial loop’;

- Zooplankton actively participates in nutrient cycles and simultaneously stimulates algae and microbes by nutrient remineralization, while in the same time zooplankton reduces algal and microbial populations by consuming them directly; and
- Many fish species feed on zooplankton.

In recent years, in aquatic ecosystems, data mining methods have been used for monitoring different communities more frequently. They are also known as knowledge-discovery in databases that implies automatic or semiautomatic research and analysis of great amount of data, in order to discover patterns and relations hidden among the data (HAN *et al.*, 2010).

The Gruža and the Grošnica reservoirs are important sources of water supply for Kragujevac city and its environment. In the previous period, these reservoirs were the subjects of various hydro-biological researches, which include zooplankton (ČOMIĆ and OSTOJIĆ 2005; OSTOJIĆ 2000, 2008; OSTOJIĆ *et al.*, 2005, 2007).

Analysis, modeling and prediction of the total zooplankton abundance were performed by various statistical tools, among which data mining tools were less frequent. Artificial neural networks were used for modeling zooplankton density groups in the Coquerio Lake in the northern Pantanal of Brazil (FANTIN-CRUZ *et al.*, 2010). Artificial neural networks were used in modeling and prediction of zooplankton dynamics (RECKNAGEL *et al.*, 1998) and for prediction of surface zooplankton biomass (WOOD-WALKER *et al.*, 2001). Vertical behavior of zooplankton was modeled as a stimuli-response process where the inputs from the environment (light, food, and predators) are used as decision parameters (EIANE and PAIRSI, 2001). Authors used simple neural networks to control behavior and optimization by genetic algorithms.

The aim of this paper is to automatically determine dependency degree and the influence of physicochemical and biological parameters on abundance and dynamics of total zooplankton, through design of specific data mining models. The analysis is based on the data obtained from the information system of the two reservoirs with different morphometric characteristics and trophic state.

MATERIALS AND METHODS

Study area and water quality data

The Gruža and the Grošnica reservoirs are the main suppliers of fresh water for residents of Kragujevac city (Figure 1). These reservoirs have different morphometric and trophic state parameters (OSTOJIĆ *et al.*, 2007).

The data set used in this study was generated through monitoring water quality of the Gruža and the Grošnica reservoirs. The data set includes the data of the laboratory for water quality inspection of the public service company for water supply and sewerage in Kragujevac. Monthly sampling was being carried out during the two years period (2009-2011). Three permanent sampling sites were selected for qualitative and quantitative sampling for the Grošnica Reservoir and five sampling sites for the Gruža Reservoir (Figure 1). Samples were taken at each 5 m of depth. Qualitative samples of plankton were taken by plankton net (mesh size 25 μm), while quantitative samples were collected by 2-liter Ruttner hydrobiological bottle and then filtered through the plankton net. Samples were preserved with 4% Formalin at the collection site. Analyses were performed by using standard methods (APHA, 1998).

Physicochemical, microbiological and other parameters used for modeling are the same for both reservoirs. They were taken from the information system of Serbian lakes and reservoirs (SeLaR). Database overview and its structure are described in detail in the papers RADOJEVIĆ *et al.* (2008) and STEFANOVIĆ *et al.* (2012).

Data set for the Gruža Reservoir includes 167 samplings of the total zooplankton (ind/dm^3). The input parameters that we've been using for the analysis of key influencers on

modeling and prediction of the total zooplankton are: month (temporal dimension), location and depth (spatial dimension), water temperature ($^{\circ}\text{C}$), turbidity ($^{\circ}\text{SI}$ scale), pH, dissolved oxygen (mg/cm^3), manganese (Mn) (mg/cm^3), iron (Fe) (mg/cm^3), chlorides (mg/cm^3), nitrates (mg/cm^3), nitrites (mg/cm^3), ammonia (mg/cm^3), total phosphates (mg/cm^3), chlorophyll-a (mg/m^3), chemical oxygen demand (COD) (mg/cm^3), 5-day biological oxygen demand (BOD) (mg/cm^3), index of phosphatase activity (IPA) ($\mu\text{mol}/\text{s}/\text{dm}^3$), total bacteria (bact/cm^3), heterotrophic bacteria (psihrophile and mesophile) (cfu/cm^3), total coliforms ($\text{MPN}/100\text{cm}^3$), *Clostridium perfringens* ($\text{N}^{\circ}/\text{dm}^3$), nitrogen fixing bacteria (cfu/cm^3), cellulolytic bacteria (cfu/cm^3), proteolytic bacteria (cfu/cm^3), amilolityc bacteria (cfu/cm^3) and phosphorus minerilizing bacteria (cfu/cm^3). In the analysis we have been using zooplankton group - Protozoa, Rotifera, Cladocera, Copepoda (ind/dm^3), and data of number species that have been found: *Bosmina coregoni*, *B. longirostris cornuta*, *B. longirostris similis*, *Brachionus angularis*, *B. diversicornis diversicornis*, *B. diversicornis homoceros*, *Carchesium polypinum*, *Daphnia cucullata*, *Diaphanosoma brachyurum*, *Eudiaptomus gracilis*, *Filinia longiseta*, *Kellicottia longispina*, *Keratella cochlearis*, *K. cochlearis hispida*, *K. cochlearis macracantha*, *K. cochlearis micracantha*, *K. cochlearis tecta*, *K. quadrata*, *K. quadrata frenzeli*, *Lecane closterocerca*, *Leptodora kindti*, *Polyarthra dolichoptera*, *P. major*, *Synchaeta* sp., *Tintinnidium fluviatile*, *Tintinnopsis lacustris*, *Trichocerca similis* (ind/dm^3).

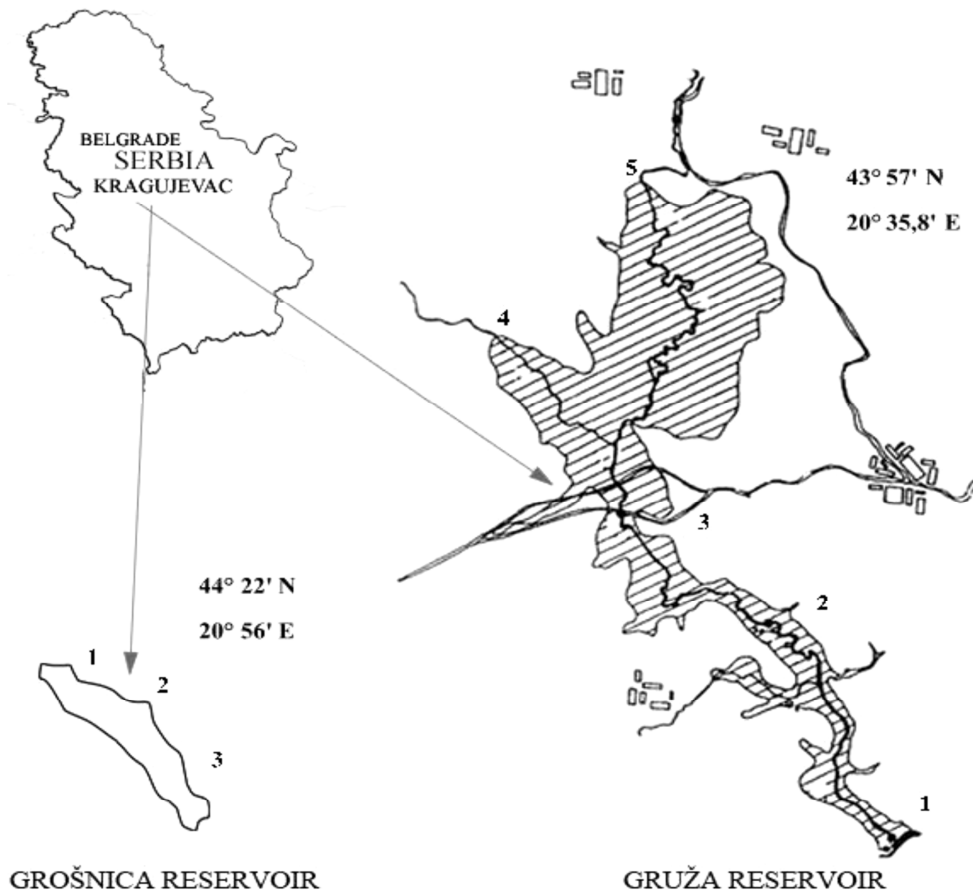


Figure 1. The Gruža Reservoir and the Grošnica Reservoir with sampling points (1 – Dam, 2 – Center, 3 – Bridge, 4 – The mouth of the Borač River, 5 – The mouth of the Gruža River).

Data set for the Grošnica reservoir includes 120 samplings of the total zooplankton (ind/dm^3), with parameters: month (temporal dimension), location and depth (spatial dimension), water temperature ($^{\circ}\text{C}$), turbidity ($^{\circ}\text{SI}$ scale), pH, dissolved oxygen (mg/cm^3), manganese (Mn) (mg/cm^3), iron (Fe) (mg/cm^3), chlorides (mg/cm^3), electrical conductivity ($\mu\text{S}/\text{cm}$), nitrates (mg/cm^3), nitrites (mg/cm^3), ammonia (mg/cm^3), total phosphates (mg/cm^3), chlorophyll-a

(mg/m³), total chlorophyll (mg/dm³), chemical oxygen demand (COD) (mg/cm³), 5-day biological oxygen demand (BOD) (mg/cm³), index of phosphatase activity (IPA) (μmol/s/dm³), total bacteria (bact/cm³), heterotrophic bacteria (psihrophile and mesophile) (cfu/cm³), facultative oligotrophic bacteria (cfu/cm³) and phosphorus minerilizing bacteria (cfu/cm³). The data set includes parameters: Protozoa, Rotifera, Cladocera, Copepoda (ind/dm³) as well as number of detected zooplankton species: *Brachionus diversicornis diversicornis*, *B. diversicornis homoceros*, *Carchesium polypinum*, *Daphnia cucullata*, *Diaphanosoma brachyurum*, *Eudiaptomus gracilis*, *Filinia longiseta*, *Gastropus stylifer*, *Kellicottia longispina*, *Keratella cochlearis*, *K. cochlearis hispida*, *K. cochlearis macracantha*, *K. cochlearis micracantha*, *K. cochlearis tecta*, *K. quadrata*, *Lecane closterocerca*, *Leptodora kindti*, *Polyarthra dolichoptera*, *P. vulgaris*, *Synchaeta* sp., *Tintinnidium fluviatile*, *Tintinnopsis lacustris*, *Trichocerca similis* (ind/dm³).

Data analysis, method and models

During the design and development, we used a multi-phased approach process for data mining as shown in Figure 2 (SHEARER, 2000).

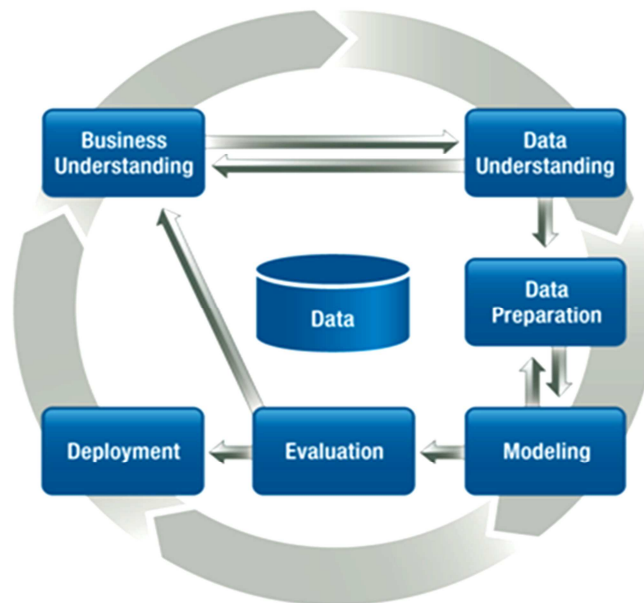


Figure 2. Phases of the CRISP-DM reference model.

The initial phase focuses on understanding the objectives and requirements, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives. The data understanding phase includes: data collection, data quality analysis and discovering first insights into the data, and/or detection of interesting subsets to form hypotheses regarding hidden information. Data preparation phase includes tasks such as table, record, and attribute selection, as well as transformation and cleaning data for modeling tools. During the modeling phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. After building the model, it is important to thoroughly evaluate it and review the steps executed to create it, to be certain the model properly achieves the preset objectives. Finally, the knowledge gained needs to be organized and presented in a way that the end-user can consume and benefit.

Data analysis has the aim to determine hidden or unknown correlation (dependence) between the attributes of entities, common characteristics of the entities' attributes and prediction of their behavior in the future. It enables making conclusions and taking appropriate measures in

accordance with the objectives set. In the database, all relevant changes on the entities are registered by date and location which provides analysis by temporal and spatial dimensions.

The data analysis is accomplished by using Analysis Services component of the SQL Server 2008 (TANG and MACLENNAN, 2008). The data from relational databases are extracted, transformed and loaded (ETL) into appropriate data warehouse through the special ETL package that is designed and executed via SQL Server Integration Services component. This means that the appropriate data structures, which have been transformed from a database, are available to the user without any additional engagement. An approach with data warehouse provides integrated and optimized data source for advanced analytics such as data mining.

Data mining offers a variety of options for data analysis (HART, 2008). Typical analysis has the following steps: modeling, realization of the model and obtaining the report. Modeling involves determining the characteristics included in the model, which depends on the objective of the analysis. Prior to the algorithm execution it is necessary to check and clean the data and determine the parameters of algorithms that have been applied. Realization of the model represents execution of the appropriate algorithm on the model. In this case, Naïve Bayes algorithm was used. For realization of appointed aims the analysis of key influencers was used.

The analysis of key influencers

The analysis of key influencers is used to show how column values in a data set might determine the values of a specified target column. It enables selecting the variable (column, parameter) which contains the desired outcome or target value. Samples within a dataset are analyzed in order to determine which factors have the strongest influence on the outcome. The designed data mining model enable automated analysis through training and testing steps, as well as automatic discovery of hidden relationships. For example, if we have the total number of zooplankton in the column with the values from the past year, we can analyze the table to determine the parameters that have the key impact. There is a choice of several possible outcomes and their comparison, which helps us in determining the potential decision parameters.

The results of key influencers analysis are the new data tables that report on the factors associated with each outcome and graphically show their probabilistic relations. Tables can be filtered out from various factors and outcomes, so that the results are researched at several levels. If the target column contains continual numeric values, the model automatically allots the numeric values into the groups. These groups represent clusters of objects with similar characteristics. However, numeric values are not distributed in typical limits. The analysis of key influencers comprises the following steps:

1. Creating the DM structure (data source) that stores key information about the data;
2. Creating a model in the OLAP (On-Line Analytical Processing) server by using Naïve Bayes algorithm; and
3. Issue a prediction query for each pair of attributes that you specify to identify the factors that strongly distinguish the two target attributes.

The tool sets all parameters automatically after conducting an analysis of the data to determine the optimum settings.

RESULTS AND DISCUSSION

The tool automatically adjusts all the parameters after performing the data analysis to determine optimum settings. The created reports include four columns with the following information: differences factor, assessment of the value that is strongly associated with the objective, favoring of the outcome or target value predicted by the factor and the relative impact which points to the association strength (HAN *et al.*, 2010).

In both reservoirs, values of total zooplankton data were classified into five classes (by using Decision Trees algorithm). In the Grošnica Reservoir: I class: less than 854 ind/dm³ of zooplankton; II class: 854-1483 ind/dm³; III class: 1483 – 2555 ind/dm³; IV class: 2555 – 4403 ind/dm³ and V class: more than 5803 ind/dm³. In the Gruža Reservoir: I class: less than 1116 ind/dm³; II class: 1116 – 2385 ind/dm³; III class: 2385 – 4386 ind/dm³; IV class: 4386 – 5803 ind/dm³ and V class: more than 5803 ind/dm³.

Based on the results of the analysis we can conclude that if we allow selection of all available parameters collected in the database, then analysis with most influence will connect parameters of groups (Protozoa, Rotifera, Cladocera, Copepoda) and species of zooplankton with total zooplankton. In the Grošnica Reservoir (Tables 1 and 2), analysis showed that the temporal dimension (months), water temperature, chemical oxygen demand, chlorophyll, nitrates and chlorides, also had the key influence with strong relative impact, while in the Gruža Reservoir (Table 4) physiological groups of bacteria were influential (amilolityc and proteolytic).

Table 1. Key Influencers Report and their impact over the values of total zooplankton (ind/dm³) with zooplankton groups and species for the Grošnica Reservoir.

Parameter (unit measure)	Value	Favors	Relative Impact
Rotifera (ind/dm ³)	< 733	< 854	100
<i>Keratella cochlearis</i> (ind/dm ³)	< 72	< 854	81
Cladocera (ind/dm ³)	< 40	< 854	71
<i>Keratella cochlearis micracantha</i> (ind/dm ³)	7 – 23	1483 – 2555	88
<i>Keratella cochlearis hispida</i> (ind/dm ³)	337 - 435	1483 – 2555	88
Month	7	1483 – 2555	63
Chlorides (mg/cm ³)	5.6 - 6.3	1483 – 2555	53
<i>Trichocerca similis</i> (ind/dm ³)	< 35	1483 – 2555	52
Copepoda (ind/dm ³)	345 - 689	1483 – 2555	50
Copepoda (ind/dm ³)	188 - 345	2555 – 4403	100
<i>Keratella cochlearis micracantha</i> (ind/dm ³)	>= 23	2555 – 4403	90
<i>Leptodora kindti</i> (ind/dm ³)	1	2555 – 4403	86
<i>Carchesium polypinum</i> (ind/dm ³)	165	2555 – 4403	74
<i>Lecane closterocerca</i> (ind/dm ³)	2	2555 – 4403	74
<i>Gastropus stylifer</i> (ind/dm ³)	48 - 118	2555 – 4403	69
<i>Eudiaptomus gracilis</i> (ind/dm ³)	11 – 39	2555 – 4403	62
<i>Keratella cochlearis hispida</i> (ind/dm ³)	>= 435	2555 – 4403	55
Protozoa (ind/dm ³)	>= 859	2555 – 4403	55
<i>Daphnia cucullata</i> (ind/dm ³)	298 - 409	2555 – 4403	55
<i>Gastropus stylifer</i> (ind/dm ³)	>= 199	2555 – 4403	55
Protozoa (ind/dm ³)	542 - 859	2555 – 4403	50
<i>Tintinnopsis lacustris</i> (ind/dm ³)	72 - 388	>= 4403	100
<i>Keratella cochlearis tecta</i> (ind/dm ³)	32 – 68	>= 4403	80
<i>Keratella cochlearis</i> (ind/dm ³)	>= 554	>= 4403	55
Copepoda (ind/dm ³)	345 - 689	>= 4403	51
<i>Polyarthra vulgaris</i> (ind/dm ³)	9	>= 4403	50
<i>Leptodora kindti</i> (ind/dm ³)	3	>= 4403	50

The analysis also classifies key influencers in several classes. In the tables (1 and 2) it can be seen that the same influential parameter, found in a number of different classes, can have different relative impact on total zooplankton. For example, in the Grošnica Reservoir, group Copepoda in the range of 345 – 689 ind/dm³ is influential parameter on class III of total

zooplankton (1483 – 2555 ind/dm³) with relative impact 50. At the same time, the same group in the range of 188 – 345 ind/dm³ is influential parameter on class IV of total zooplankton (2555 – 4403 ind/dm³) with relative impact 100. Since, in this case, we observe influential parameters on total zooplankton, we did not specify the classification of key influencers.

In the Grošnica Reservoir, we can expect that with increasing abundance of group Rotifera, especially species like *Keratella* spp., abundance of total zooplankton will have a tendency to grow. When the abundance of total zooplankton is high, we expect high values of number of specimens within group Protozoa (especially *Tintinnopsis lacustris*). Low abundance of total zooplankton will most certainly point out low abundance of group Copepoda. Furthermore, similar could be seen in the group Cladocera (species *Bosmina longirostris*), but with significantly lower relative impact. Group Rotifera, especially *Keratella* spp., can predict the presence of average classes (II, III, IV) of total zooplankton, with high values of relative impact, when abundance of *Keratella* species are at maximum level (Table 1).

Table 2. Key Influencers Report and their impact over the values of total zooplankton (ind/dm³) without zooplankton groups and species for the Grošnica Reservoir.

Parameter (unit measure)	Value	Favors	Relative Impact
Month	6	< 854	100
Chemical oxygen demand (mg/cm ³)	< 8.4	< 854	96
Month	3	< 854	85
Nitrates (mg/cm ³)	0.9	< 854	83
Water temperature (°C)	< 8	< 854	77
Chlorophyll-a (mg/m ³)	3.22- 4.39	< 854	65
Conductivity (µS/cm)	>= 458	< 854	62
Total chlorophyll (mg/dm ³)	0.13 - 0.17	< 854	57
Month	2	< 854	52
Month	5	< 854	52
Month	7	854 - 1483	100
Total chlorophyll (mg/dm ³)	< 0.07	854 - 1483	59
Chemical oxygen demand (mg/cm ³)	10.3 - 12.3	854 - 1483	49
Month	7	1483 - 2555	100
Chlorides (mg/cm ³)	5.6- 6.3	1483 - 2555	85
Nitrates (mg/cm ³)	0-0.2	1483 - 2555	67
Dissolved oxygen (mg/cm ³)	8 - 10	1483 - 2555	50
Month	7	2555 - 4403	86
Location	5	2555 - 4403	56
Month	8	>= 4403	100
Water temperature (° C)	>= 21	>= 4403	97

High values of Protozoa abundance do not necessarily indicate high abundance of total zooplankton, but the average. If the abundance of the species *Carchesium polypinum* has extreme high values, it necessarily points to high abundance of total zooplankton (Table 3).

It appears that high abundance of group Protozoa may indicate, with great significance, the average abundance of total zooplankton for the Gruža Reservoir (Table 3), while its species *C. polypinum* in extreme abundance point to high abundance of total zooplankton. Connection between extremely high values of physiological group of bacteria and extremely high values of total zooplankton is shown by the results of the analysis (Table 4).

Table 3. Key Influencers Report and their impact over the values of total zooplankton (ind/dm³) with zooplankton groups and species for the Gruža Reservoir.

Parameter (unit measure)	Value	Favors	Relative Impact
Copepoda (ind/dm ³)	< 234	< 1116	100
<i>Polyarthra dolichoptera</i> (ind/dm ³)	< 132	< 1116	75
<i>Bosmina longirostris similis</i> (ind/dm ³)	< 16	< 1116	57
Cladocera (ind/dm ³)	< 204	< 1116	51
<i>Tintinnopsis lacustris</i> (ind/dm ³)	72 - 156	1116 - 2385	95
Rotifera (ind/dm ³)	375 - 1230	1116 - 2385	90
Copepoda (ind/dm ³)	234 - 575	1116 - 2385	53
Rotifera (ind/dm ³)	>= 1230	2385 - 4386	100
<i>Keratella cochlearis tecta</i> (ind/dm ³)	>= 384	2385 - 4386	72
Protozoa (ind/dm ³)	>= 492	2385 - 4386	64
Copepoda (ind/dm ³)	234 - 575	2385 - 4386	58
Cladocera (ind/dm ³)	899 - 1513	2385 - 4386	53
<i>Bosmina longirostris cornuta</i> (ind/dm ³)	601 - 1295	2385 - 4386	53
<i>Brachionus angularis</i> (ind/dm ³)	>= 93	4386 - 5803	100
<i>Keratella cochlearis tecta</i> (ind/dm ³)	>= 384	4386 - 5803	81
Rotifera (ind/dm ³)	>= 1230	4386 - 5803	77
<i>Bosmina longirostris cornuta</i> (ind/dm ³)	65 - 406	4386 - 5803	58
<i>Lecane closterocerca</i> (ind/dm ³)	6	4386 - 5803	56
<i>Carchesium polypinum</i> (ind/dm ³)	>= 386	>= 5803	100
<i>Brachionus diversicornis diversicornis</i> (ind/dm ³)	342 - 1061	>= 5803	89
<i>Diaphanosoma brachyurum</i> (ind/dm ³)	32 - 84	>= 5803	67
<i>Leptodora kindti</i> (ind/dm ³)	18	>= 5803	67
<i>Brachionus diversicornis homoceros</i> (ind/dm ³)	>= 123	>= 5803	67
<i>Leptodora kindti</i> (ind/dm ³)	24	>= 5803	67
Amilolityc bacteria (cfu/cm ³)	>= 4970	>= 5803	53
<i>Diaphanosoma brachyurum</i> (ind/dm ³)	>= 84	>= 5803	53
Proteolytic bacteria (cfu/cm ³)	>= 4144	>= 5803	53

Table 4. Key Influencers Report and their impact over the values of total zooplankton (ind/dm³) without zooplankton groups and species for the Gruža Reservoir.

Parameter (unit measure)	Value	Favors	Relative Impact
Location	Dam	< 1116	100
Water temperature (° C)	18 - 22	1116 - 2385	100
Water temperature (° C)	>= 22	2385 - 4386	100
Amilolityc bacteria (cfu/cm ³)	3448 - 4970	4386 - 5803	100
Amilolityc bacteria (cfu/cm ³)	502 - 1377	4386 - 5803	71
Proteolytic bacteria (cfu/cm ³)	2274 - 4144	4386 - 5803	65
Proteolytic bacteria (cfu/cm ³)	>= 4144	>= 5803	100
Amilolityc bacteria (cfu/cm ³)	>= 4970	>= 5803	100

Analysis of the key influencers gives us a possibility to perceive relative influence of physical, chemical and other biological parameters, if the inputs are without group and zooplankton species parameters (Tables 2 and 4).

In the Grošnica Reservoir analysis may indicate, with great significance, that lower abundance of total zooplankton could be expected by the summer months. In the warmest period of the year the abundance of total zooplankton are higher. Also, very reliable predicting parameter for low abundance of total zooplankton, is concentration of nitrates where high concentrations cause lower abundance of total zooplankton, and vice versa. In the Gruža Reservoir key influencer parameters for total zooplankton are: spatial dimension (location), water temperature and physiological groups of bacteria. The lowest abundance of total zooplankton was sampled at the dam location, but with increase of water temperature the tendency of increasing its abundance was identified. Abundance of total zooplankton is followed by the abundance for physiological groups of bacteria (amylolytics and proteolytics) (Table 4).

By observing recent researches in the field of modeling and prediction of total zooplankton and the groups of zooplankton, it can be noted that there is a need for necessary choice of selecting inputs in advance, for the most of used models (RECKNAGEL *et al.*, 1998; EIANE and PAIRSI, 2001; WOOD-WALKER *et al.* 2001; FANTIN-CRUZ *et al.*, 2010). For modeling abundance of zooplankton groups (Rotifera, Cladocera, Copepoda) in lakes, some authors use chlorophyll-a, dissolved oxygen, pH, solar radiation, water temperature and secchi depth (RECKNAGEL *et al.*, 1998). For the same purpose, others use dissolved oxygen, pH, water temperature, water level, water transparency, turbidity, electrical conductivity, alkalinity, total nitrogen, total phosphorus, chlorophyll-a in one model, while researches from previous periods, considering outputs, are additionally used in other model (FANTIN-CRUZ *et al.*, 2010). Zooplankton biomass in the Atlantic Ocean has been modeled with two different methods: multiple linear regression and neural networks. Inputs that were taken had been the abundance and size of zooplankton, by using optical plankton counter (WOOD-WALKER *et al.* 2001).

The results of authors mentioned above, like our results, indicate that for modeling and predictions of total zooplankton, there is a need for selecting inputs in advance (eg. previously determined abundance of zooplankton or some other parameter with regard to zooplankton, for example biomass).

It can be noted clearly that a connection of zooplankton groups/species with total zooplankton exists. In the analysis of key influencers, there is a possibility of a singly choice of inputs. It also gives an option of automatic selection of key influencers from the whole database which could be used as a tool for prediction. These cognitions can also serve as the basis for new types of modeling that are based on selection of the most influential parameters.

CONCLUSION

The resulting models, obtained by analysis of key influencers, showed that in both reservoirs, parameters of groups and species of zooplankton have the greatest influence on the total zooplankton abundance. We noted differences in the impact of physical, chemical and other biological parameters which depends on reservoirs. Key influencers in the Grošnica Reservoir are: the temporal dimension (months), nitrates, water temperature, chemical oxygen demand, chlorophyll and chlorides. In the Gruža Reservoir, key influencers are: spatial dimension (location), water temperature and physiological groups of bacteria. The results show that the presented data mining model is usable on any kind of aquatic ecosystem and also can serve for detection of inputs which could be the basis for the future analysis and modeling.

Acknowledgement

This research was supported by the Ministry of Education, Science and Technological Development of Serbia, grants No. III41010 and III41007. The authors are thankful to the Public Service Company for Water Supply and Sewerage which gave access to some data needed for the purpose of this paper.

References:

- [1] APHA (1998). Standard Methods for the Examination of Water and Waste Water, 20th ed. American Public Health Association. Washington, DC.
- [2] ČOMIĆ, LJ., OSTOJIĆ, A. (Eds.) (2005): *The Gruža Reservoir*. Faculty of Science, University of Kragujevac. (in Serbian)
- [3] EIANE, K. and PARISI D. (2001): Towards a robust concept for modelling zooplankton migration. *Sarsia* **86**: 465-475.
- [4] FANTIN-CRUZ, I., PEDROLLO, O., COSTA BONECKER, C., MOTTA-MARQUES D., LOVERDE-OLIVEIRA, S. (2010): Zooplankton Density Prediction in a Flood Lake (Pantanal – Brazil) Using Artificial Neural Networks. *Internat. Rev. Hydrobiol.* **95** (4–5): 330–342.
- [5] HAN, J., KAMBER, M., PEI, J. (2010): *Data mining: concepts and techniques. 3rd ed.*, Elsevier Inc.
- [6] HART, D. (2008): *Microsoft Office 2007*. Business Intelligence, McGraw-Hill.
- [7] LIKENS, G.E. (2010): *Plankton of inland waters*. Elsevier, Academic Press.
- [8] OSTOJIĆ, A. (2000): *Comparative ecological study of zooplankton of the Grošnica Reservoir and Gruža Reservoir*. PhD thesis. Fac. of Biol. Sci., Univ. of Belgrade. (in Serbian)
- [9] OSTOJIĆ, A. (2008): *Zooplankton of the Grošnica and Gruža Reservoir*. Faculty of Science, University of Kragujevac. (in Serbian)
- [10] OSTOJIĆ, A., ČURČIĆ, S., ČOMIĆ, LJ., TOPUZOVIĆ, M. (2005): Estimate of the eutrophication process in the Gruža Reservoir (Serbia and Montenegro). *Acta Hydrochimica et Hydrobiologica* **33**: 605-613.
- [11] OSTOJIĆ, A., ČURČIĆ, S., ČOMIĆ, LJ., TOPUZOVIĆ, M. (2007): Effects of anthropogenic influences on the trophic status of two water supply reservoirs in Serbia. *Lakes and Reservoirs: Research and Management* **12**: 175–185.
- [12] RADOJEVIĆ, I., STEFANOVIĆ, D., ČOMIĆ, LJ., OSTOJIĆ, A. (2008) Information System of Lakes and Reservoirs of the Serbia – Selar. *Kragujevac J. Sci.* **30**: 81-98.
- [13] RECKNAGEL, F., FUKUSHIMA, T., HANAZATO, T., TAKAMURA, N., WILSON H. (1998): Modelling and prediction of phyto- and zooplankton dynamics in Lake Kasumigaura by artificial neural networks. *Lakes and Reservoirs: Research and Management* **3**(2): 123–133.
- [14] SHEARER C. (2000): The CRISP-DM model: the new blueprint for data mining, *J Data Warehousing* **5**: 13-22.
- [15] STEFANOVIĆ, D., RADOJEVIĆ, I., ČOMIĆ, LJ., OSTOJIĆ, A., TOPUZOVIĆ, M., KAPLAREVIĆ-MALIŠIĆ, A. (2012): Management Information System of Lakes and Reservoirs. *Water Resources* **39** (4): 488-495.
- [16] SUTHERS, I. M., RISSIK, D. (2009): *Plankton, a guide to their ecology and monitoring for water quality*. CSIRO Publishing.
- [17] TANG, Z., MACLENNAN, J. (2008): *Data Mining with SQL Server 2008*. Wiley Publishing, Inc.
- [18] WOODD-WALKER, R.S., KINGSTON, K.S., GALLIENNE, C.P. (2001): Using neural networks to predict surface zooplankton biomass along a 50°N to 50°S transect of the Atlantic. *Journal of Plankton Research* **23** (8): 875-888.